

Clasificación Automatizada de Patrones Respiratorios para la Detección de Enfermedades Pulmonares mediante Inteligencia Artificial

Nava Fabricio^{1, a *}, Juárez Mauricio^{1, b *}, Chirre Luis^{1, c *} and De la Cruz Lewis^{2, d}

¹ Faculty of Sciences and Engineering- Biomedical Engineering, Universidad Peruana Cayetano Heredia UPCH

² LID-UPCH, Laboratorios de Investigación y Desarrollo, Universidad Peruana Cayetano Heredia UPCH

^afabricio.nava@upch.pe, ^bmauricio.juarez@upch.pe, ^cluis.chirre@upch.pe,
^dumbert.de.la.cruz@upch.pe

Keywords: EPOC, enfermedades respiratorias, AI, software, diagnóstico

Abstract. El proyecto propone desarrollar un sistema basado en inteligencia artificial para la detección automatizada y precisa de enfermedades pulmonares, con enfoque en la Enfermedad Pulmonar Obstructiva Crónica (EPOC), a partir de datos obtenidos mediante pruebas de espirometría. Los objetivos se centran en identificar los requerimientos del sistema, diseñar una solución técnica y funcional, y desarrollar una interfaz amigable para la clasificación de EPOC en pacientes. La metodología presenta etapas de identificación de requerimientos, desarrollo del software y validación del mismo. Se emplearon técnicas de preprocesamiento de datos, extracción de características y desarrollo de modelos de inteligencia artificial, como redes neuronales recurrentes, para analizar y clasificar los patrones respiratorios. La validación del modelo se realizó usando métricas de conjuntos de datos separados, evaluando precisión, sensibilidad y especificidad.

Introducción

La Enfermedad Pulmonar Obstructiva Crónica (EPOC) es un problema de salud global debido a su gran prevalencia de 10% en adultos mayores, su incidencia creciente con los años y su significativo costo económico, personal y social [1]. Esta enfermedad es caracterizada como incurable la cual causa una obstrucción de flujo de aire de los pulmones y que puede tener síntomas como dificultad para respirar, tos, producción de moco, sibilancias y fatiga [2] [3]. Las causas son bastante variadas y pueden ser desde la exposición al tabaco (por fumar o exposición pasiva) o al polvo hasta eventos en la vida fetal, asma en la infancia y deficiencia en alfa-1 antitripsina [3]. Esta afección puede generar distintas complicaciones como infecciones respiratorias, problemas cardíacos, cáncer al pulmón, presión arterial alta en arterias pulmonares y en aspectos emocionales como la depresión [2].

Debido a este problema mundial, como lo es la EPOC, la Organización Mundial de la Salud adopta algunas medidas y abarcan esta complicación en su Plan de Acción Mundial para la

Prevención y el Control de las Enfermedades No Transmisibles y la Agenda 2030 para el Desarrollo Sostenible de las Naciones Unidas [3]. Por otra parte, la misma organización ha creado la Global Initiative for Chronic Obstructive Lung Disease (GOLD) la cual está en constante actualización con esta enfermedad, aportando con reportes anuales sobre la definición de la EPOC, patogénesis, diagnósticos, tratamiento, etc. [4].

Las enfermedades pulmonares, como el asma, la enfermedad pulmonar obstructiva crónica (EPOC) y la fibrosis pulmonar, pueden afectar significativamente la calidad de vida de los pacientes. La detección temprana y el monitoreo continuo de los patrones respiratorios son fundamentales para el manejo efectivo de estas enfermedades. Sin embargo, la evaluación manual de los patrones respiratorios es subjetiva y puede ser propensa a errores.

Actualmente la prueba de espirometría es el gold standard en la evaluación para la identificación de obstrucción de flujo aéreo mediante la evaluación de las propiedades mecánicas presentes en el sistema respiratorio [5]. Los parámetros utilizados son la capacidad vital forzada (FVC), el volumen espiratorio forzado en el primer segundo (FEV1) y el cociente FEV1/FVC [5]. Sin embargo, el uso único de esta herramienta para la detección de EPOC no es suficiente pues influyen más variables ajenas a las propiedades mecánicas respiratorias, por ejemplo: peso, edad, género, historial de fumador, condición física, etc.

La espirometría para la evaluación de EPOC tiene buena sensibilidad pero baja especificidad y por ello es difícil diferenciar entre una EPOC y asma [6] [7]. Por ello se utiliza en conjunto con otras pruebas y análisis para tener un mejor diagnóstico [7].

Una inclusión adecuada es la aplicación de modelos computacionales de Inteligencia Artificial (IA), esta ha sido fundamental en el avance de la automatización en diversas áreas. En la salud, la IA se emplea para analizar datos médicos, diagnosticar enfermedades y personalizar tratamientos. La capacidad de la IA para analizar grandes cantidades de datos de manera rápida y precisa ha revolucionado la automatización en múltiples industrias, mejorando la eficiencia y la calidad del trabajo realizado. [8]

Metodología

Base de datos. Para el desarrollo de nuestro sistema de detección de Enfermedad Pulmonar Obstructiva Crónica (EPOC), se utilizó una base de datos compuesta por datos clínicos y parámetros fisiológicos obtenidos por bases de datos libres, de acceso no restringido. Esta base de datos contiene información de pacientes que han sido sometidos a pruebas de espirometría y que han sido diagnosticados con EPOC u otras enfermedades pulmonares.

Los datos clínicos incluyen información relevante sobre el historial médico de los pacientes, como antecedentes de tabaquismo, exposición a factores de riesgo ambientales y comorbilidades asociadas. Además, se recopilaron mediciones específicas de la función pulmonar, como la

capacidad vital forzada (FVC), el volumen espiratorio forzado en el primer segundo (FEV1) y el cociente FEV1/FVC.

Esta base de datos fue fundamental para definir los criterios de detección de la EPOC y para entrenar y validar nuestros modelos de inteligencia artificial. Se utilizaron técnicas de preprocesamiento de datos para eliminar artefactos y ruido, y se realizó una cuidadosa normalización de las señales respiratorias para garantizar la calidad de los datos utilizados en el desarrollo del software.

Diseño de software. El diseño del software para la detección automatizada de EPOC se llevó a cabo siguiendo un enfoque basado en inteligencia artificial y aprendizaje automático. Durante la etapa de desarrollo del producto, se implementaron varios módulos para procesar y analizar los datos respiratorios obtenidos de las pruebas de espirometría.

El proceso de desarrollo del software incluyó las siguientes etapas:

Preprocesamiento de Datos: Se aplicaron técnicas de preprocesamiento para limpiar y normalizar los datos respiratorios, eliminando artefactos y ruido para mejorar la calidad de las señales.

Extracción de Características: Se identificaron y extrajeron características relevantes de los patrones respiratorios, como la duración de la inspiración y espiración, la frecuencia respiratoria y la variabilidad.

Desarrollo de Modelos de Inteligencia Artificial: Se implementaron modelos de aprendizaje automático, incluyendo redes neuronales recurrentes (RNN) y modelos de clasificación, para analizar las características extraídas y clasificar los patrones respiratorios en categorías normales y anormales.

Entrenamiento del Modelo: Se entrenaron los modelos utilizando conjuntos de datos etiquetados que contenían ejemplos de patrones respiratorios asociados con diferentes enfermedades pulmonares, incluida la EPOC.

Validación y Evaluación del Modelo: Se realizaron pruebas de validación utilizando conjuntos de datos separados para evaluar la precisión, sensibilidad y especificidad del modelo en la detección de patrones respiratorios anormales.

Estado del arte:

Rosaly Moreno, et al. utilizaron variables como edad, sexo, número de cigarrillos fumados diariamente, años de fumador, FVC, FEV1 y su cociente (FEV1/FVC) de un dataset con 1190 pacientes luego del EDA (Exploratory Data Analysis) para entrenar al modelo de machine learning. Se estimó las variables usando un gradient tree boosting (GTB) y un decision tree. Dicho modelo obtuvo métricas favorables para el diagnóstico de EPOC (sensitivity: 93%, specificity: 97%, accuracy: 95%, precision: 94%) [1]

X. Wang et al. usaron 38 variables en total como la edad, sexo, presencia de algún problema respiratorio, tos, presión, etc. para evaluar el riesgo de contraer EPOC en la población fumadora. En este estudio se elaboraron 7 modelos diferentes como Support Vector Machine, regresión

logística, Random Forest, XGBoost, NGBoost, LGBM y CatBoost y luego se interpretaron los resultados de dichos modelos con Shapley additive explanations (SHAP) y Partial Dependence Plot (PDP). Luego de balancear la data el modelo de SVM demostró un sensitivity: 0.608, accuracy: 0.736, specificity: 0.8, F1: 0.372, G-mean values: 0.646. [2]

[1] “Artificial Intelligence Applied to Forced Spirometry in Primary Care | Open Respiratory Archives,” *Elsevier.es*, 2022.
<https://www.elsevier.es/en-revista-open-respiratory-archives-11-resumen-artificial-intelligence-applied-forced-spirometry-S265966362400016X> (accessed May 29, 2024).

[2] X. Wang *et al.*, “An explainable artificial intelligence framework for risk prediction of COPD in smokers,” *BMC public health*, vol. 23, no. 1, Nov. 2023, doi: <https://doi.org/10.1186/s12889-023-17011-w>.