

Identificación de temas publicados sobre el monitoreo ambulatorio de la presión arterial mediante la minería de texto

Identification of Published Topics on Ambulatory Blood Pressure Monitoring using Text Mining

José Aureliano Betancourt Bethencourt^{1*} <https://orcid.org/0000-0003-0043-9526>

Elizabeth Sellén Sanchén² <http://orcid.org/0000-0001-5570-5428>

Millelys Castro Consuegra¹ <https://orcid.org/0000-0001-8443-3900>

¹Universidad de Ciencias Médicas de Camagüey, Centro de Inmunología y Productos Biológicos. Camagüey, Cuba.

²Universidad de Ciencias Médicas de Camagüey, Hospital Universitario “Manuel Ascunce Domenech”, Departamento de Cardiología. Camagüey, Cuba.

* Autor para la correspondencia: josebetancourt.cmw@infomed.sld.cu

RESUMEN

La hipertensión arterial sistémica es la enfermedad de mayor prevalencia a nivel mundial que incrementa de forma importante el riesgo cardiovascular. El monitoreo ambulatorio de la presión arterial permite registrar las lecturas durante un período de 24 horas, tanto si el paciente se encuentra despierto o dormido; además, detecta la hipertensión oculta y descarta la hipertensión de bata blanca. La investigación tuvo como objetivo identificar los temas publicados sobre el monitoreo ambulatorio de la presión arterial, mediante el empleo de la minería de texto. Con un *script* del R se accedió a las bases de datos de Europa PMC; se solicitó la cantidad de las publicaciones y los temas investigados acerca del monitoreo

ambulatorio de la presión arterial con el descriptor “ABPM”, durante el período de 2010 a 2022. Con el paquete tm se llevó de formato de texto a documento, el cual se inspeccionó; se eliminaron palabras y puntuaciones no necesarias, se preparó la matriz del texto; se buscaron los elementos más frecuentes y se trazó un gráfico de barras con los términos más frecuentes y la nube de palabras con el paquete worcloud2. Se obtuvo el gráfico nube de palabras, el de publicaciones por cada año y el de frecuencias de palabras. Fue posible identificar los temas principales publicados en los últimos 12 años sobre el monitoreo ambulatorio de la presión arterial, así como el interés creciente en el tema.

Palabras clave: inteligencia artificial; presión arterial; minería de textos; monitoreo ambulatorio; ciencia de datos; *software*; Cardiología.

ABSTRACT

Systemic arterial hypertension is the most prevalent disease worldwide that significantly increases cardiovascular risk. Ambulatory blood pressure monitoring allows recording readings over a 24-hour period, whether the patient is awake or asleep; it also detects occult hypertension and rules out white coat hypertension. The research aimed to identify published topics on ambulatory blood pressure monitoring, using text mining. With an R script, the Europe PMC databases were accessed; the number of publications and topics researched about ambulatory blood pressure monitoring with the descriptor "ABPM", during the period from 2010 to 2022, were requested. With the tm package was taken from text format to document, which was inspected; unneeded words and punctuations were removed, the text matrix was prepared; the most frequent elements were searched and a bar chart with the most frequent terms and word cloud was plotted with the worcloud2 package. The word cloud graph, the graph of publications per year and the word frequency graph were obtained. It was possible to identify the main topics published in the last 12 years on ambulatory blood pressure monitoring, as well as the growing interest in the subject.

Keywords: artificial intelligence; blood pressure; text mining; ambulatory monitoring; data science; software; Cardiology.

Recibido: 01/11/2022

Aceptado: 31/01/2023

Introducción

La Ciencia de Datos es considerada un enfoque novedoso y promisorio empleado en la obtención y análisis de información en múltiples disciplinas científicas. A pesar de su amplia difusión y aceptación, el uso de técnicas basadas en la Ciencia de Datos en el campo de la Cardiología está en ciernes.⁽¹⁾

La hipertensión arterial sistémica es la enfermedad de mayor prevalencia a nivel mundial, que incrementa de forma importante el riesgo cardiovascular. Un diagnóstico temprano, junto al logro de metas, disminuye el peligro de complicaciones. Se han actualizado los criterios diagnósticos para la hipertensión, así como la introducción del monitoreo ambulatorio de la presión arterial. Este permite registrar las lecturas durante un período de 24 horas, ya sea que el paciente esté despierto o dormido, además detecta la hipertensión oculta y descarta la hipertensión de bata blanca.⁽²⁾

En este contexto, este artículo tuvo como objetivo identificar los temas publicados en los últimos 12 años acerca del monitoreo ambulatorio de la presión arterial (MAPA).

Métodos

Se accedió a las bases de datos de Europa PMC⁽³⁾ mediante el programa R⁽⁴⁾ en el que se solicitó la cantidad de las publicaciones y los temas investigados en torno al monitoreo ambulatorio de la presión arterial con el descriptor “ABPM”. Se consideró el período 2010-2011 y se limitaron a 10000 los artículos a trabajar.

Con el paquete tm,⁽⁵⁾ que tiene métodos para la importación de datos, el manejo de corpus, preprocesamiento, gestión de metadatos y creación de matrices término-documento, se preparó el documento para su análisis. El programa tm convierte de formato de texto a documento, el cual se inspecciona, se eliminan palabras y puntuaciones no necesarias, se prepara la matriz del texto, se buscan los elementos más frecuentes y se traza un gráfico de barras con los términos más frecuentes y la nube de palabras con el paquete *wordcloud2*.⁽⁶⁾

Resultados y discusión

El análisis de grandes cantidades de datos textuales requiere un trabajo complejo para leer todo el texto y organizar el contenido. La minería de textos permite extraer información y conocimientos interesantes y no triviales de un texto no estructurado, así como detectar patrones significativos implícitamente presentes en los datos.⁽⁷⁾

En la figura 1 se pueden observar los artículos publicados por año. Se evidencia que las publicaciones van en ascenso, lo que significa que el interés en esta temática se ha incrementado, pues los investigadores necesitan disminuir el riesgo global de las enfermedades cerebrocardiovasculares.

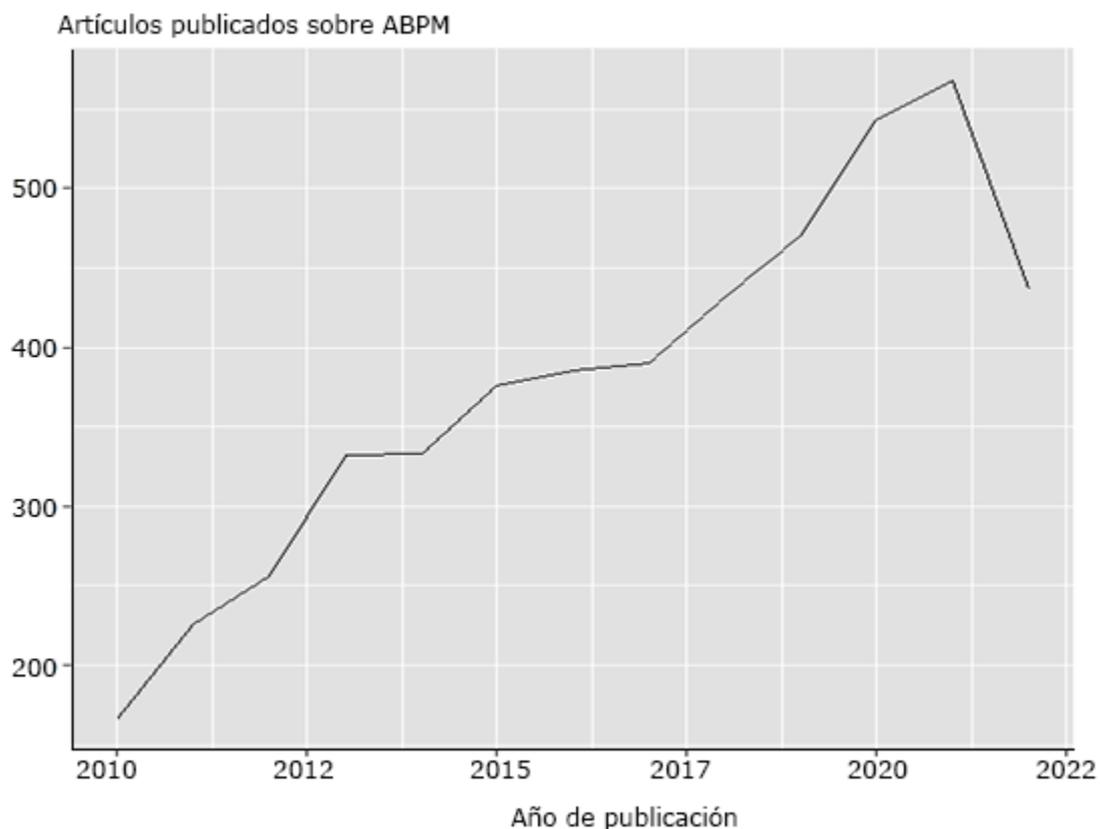


Fig. 1 – Publicaciones sobre MAPA por año.

Lo anterior se evidencia mediante resultados como los de *Hermida* y otros quienes documentan importantes limitaciones de la estratificación del riesgo de enfermedades cardiovasculares (ECV), basada en las mediciones clínicas, incluidas las del modelo de

Framingham y corroboran el valor del monitoreo ambulatorio para diagnosticar la hipertensión y estratificar el riesgo de la ECV.⁽⁸⁾

En la figura 2 se expresan las frecuencias mayores de las palabras utilizadas en las publicaciones sobre MAPA: hipertensión arterial, monitoreo ambulatorio, trastornos pediátricos, renales y cardiovasculares.

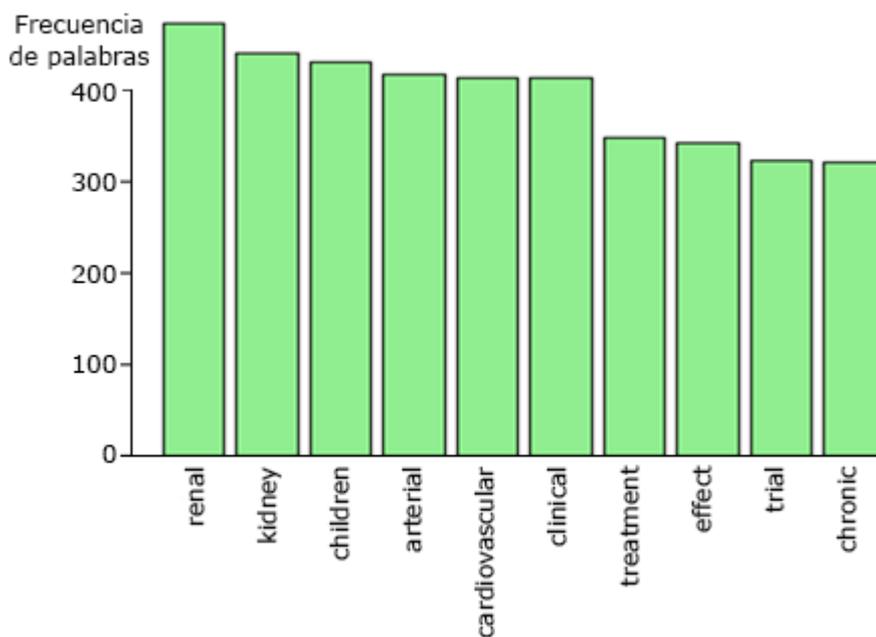


Fig. 2 – Palabras más frecuentes.

En la figura 3 se muestra la nube de palabras (también conocida como nube de etiquetas o de texto) que constituye una representación visual de la frecuencia y el valor de las palabras. Se emplea para resaltar la frecuencia con la que un término específico o categoría aparece en una fuente de datos. Las palabras que manifiestan más frecuencia en un texto se muestran en un tamaño mayor.

Jerárquicamente, representan los términos más frecuentes, lo que permite agilizar el análisis de datos tipo texto, puesto que, a través de ellas, se puede identificar e interpretar de manera rápida y sencilla las palabras con mayor relevancia en el texto analizado; contribuyen a trabajar la capacidad de síntesis, fomentar la creatividad y establecer el índice de los temas que se van a tratar o, a modo de resumen, de cada apartado que se aborde. Los autores del

presente estudio han utilizado esta herramienta en la docencia en tormentas de ideas para fortalecer las habilidades cognitivas de los estudiantes al propiciar el debate.

Las nubes de palabras son una herramienta de visualización de contenidos que permite al investigador realizar análisis descriptivos de un texto. Este tipo de visualización permite resaltar entre otras cosas las principales ideas del contenido. La frecuencia de las palabras en los textos denota la importancia de ellas en cada contexto, por lo que cada palabra puede llegar a capturar el propósito e ideas de los autores.⁽⁹⁾



Fig. 3 – Nube de palabras.

Se destaca en la nube la palabra el término *nocturnal*. La hipertensión arterial (HTA) nocturna y la enmascarada implican un peor pronóstico de la enfermedad cardiovascular; sin embargo, la información de la que se dispone en relación con el monitoreo ambulatorio de pacientes con insuficiencia cardíaca es escasa.⁽¹⁰⁾

Aparece en la nube la palabra *circadian*. El monitoreo ambulatorio de la presión arterial demostró ser un valioso instrumento para facilitar información precisa del perfil de la presión arterial en 24 horas; permitió individualizar el tratamiento según el ritmo circadiano y determinar el daño vascular.⁽¹¹⁾ Un ritmo circadiano alterado se relaciona con un riesgo cardiovascular más elevado y un peor control de los factores que conducen a este.⁽¹²⁾

De acuerdo con un grupo de investigadores,⁽¹³⁾ la aplicación de algoritmos de minería de texto a los informes de procedimientos cardiovasculares puede ser un método superior para identificar a pacientes con afecciones cardiovasculares de interés, como la estenosis aórtica y la enfermedad de las arterias coronarias.

En el presente trabajo la minería de texto aportó información útil. *Jonnagaddala* y otros argumentan que la minería de textos clínicos puede ser utilizada para extraer datos relacionados con factores de riesgo de notas clínicas no estructuradas. La salida del sistema de minería de texto fue confiable, además ha demostrado cómo se puede utilizar para evaluar riesgos de la enfermedad cardiovascular.⁽¹⁴⁾ Por otra parte, se demostró en un estudio la relevancia clínica de la integración de las características del texto y los predictores clásicos en la canalización de la minería de textos para la predicción del riesgo cardiovascular.⁽¹⁵⁾

Otros investigadores desarrollaron un sistema basado en el aprendizaje automático, diagnóstico y sistema de minería de texto para la predicción de ataques cardíacos, mediante el uso de un conjunto de datos de enfermedades cardíacas. Un grupo propuso un sistema que puede predecir y clasificar fácilmente a las personas sanas con posibilidades de ataque cardíaco.⁽¹⁶⁾

Conclusiones

Fue posible identificar los temas principales publicados en los últimos 12 años sobre el monitoreo ambulatorio de la presión arterial. De manera sencilla se explica una de las aplicaciones de la minería de texto en las publicaciones relacionadas con el MAPA. Se demuestra el creciente interés en el monitoreo ambulatorio de la tensión arterial y su mayor precisión para prevenir enfermedades cardiovasculares.

Referencias bibliográficas

1. Lemus-Delgado D, Pérez Navarro R. Ciencia de datos y estudios globales: aportaciones y desafíos metodológicos. Colombia Internacional. 2020 [acceso 06/06/2021];(102):41-62. DOI: <https://doi.org/10.7440/colombiaint102.2020.03>
2. Sauza-Sosa JC, Cuéllar-Álvarez J, Villegas-Herrera KM, Sierra-Galán LM. Aspectos clínicos actuales del monitoreo ambulatorio de presión arterial. Archivos de Cardiología de México. 2016 [acceso 06/06/2021];86(3):255-9. Disponible en: https://www.scielo.org.mx/scielo.php?pid=S1405-99402016000300255&script=sci_abstract&tlng=pt
3. Consortium EP. Europe PMC: a full-text literature database for the life sciences and platform for innovation. Nucleic acids research. 2015 [acceso 06/06/2021];43(D1):D1042-D8. Disponible en: <https://academic.oup.com/nar/article/43/D1/D1042/2437114>
4. R_Core_Team. A language and environment for statistical computing. R Foundation for Statistical Computing. 2021 [acceso 06/06/2021]. Disponible en: <https://www.r-project.org/>
5. Feinerer I. Introduction to the tm Package Text Mining in R. 2022 [acceso 06/06/2021]. Disponible en: <https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf>
6. Lang D CG. Wordcloud2: Create Word Cloud by 'htmlwidget'_. R package version 0.2.12018. Disponible en: <https://CRAN.R-project.org/package=wordcloud2>
7. Hassani H, Beneki C, Unger S, Mazinani MT, Yeganegi MR. Text mining in big data analytics. Big Data and Cognitive Computing. 2020;4(1):1. Disponible en: <https://www.mdpi.com/2504-2289/4/1/1/pdf?version=1584533390>
8. Hermida RC, Ayala DE, Mojón A, Smolensky MH, Crespo JJ, Otero A, *et al.* La presión arterial ambulatoria, en comparación con la medida clínica, mejora notablemente la estratificación del riesgo cardiovascular de Framingham. Revista Española de Cardiología. 2021;74(11):953-61. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0300893220304760>
9. Lee W-J. A study on word cloud techniques for analysis of unstructured text data. The Journal of the Convergence on Culture Technology. 2020;6(4):715-20. Disponible en: <https://koreascience.kr/article/JAKO202034965718309.page>

10. Babkowski MC. Monitorización ambulatoria de la presión arterial e insuficiencia cardíaca. Hipertensión y riesgo vascular. 2011 [acceso 06/06/2021];28(3):102-7. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S1889183711000493>
11. Sellén-Sanchén E, Ferrer-Herrera IM, Coll-Bujardon D. Monitoreo ambulatorio de presión arterial y respuesta a la cronoterapia en pacientes diabéticos hipertensos. Revista Archivo Médico de Camagüey. 2019 [acceso 06/06/2021];23(6):697-708. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1025-02552019000600697
12. García Serrano C, Micol Bachiller M, Betrán Biurrun D, Aran Solé L, Pujol Salud J. El ritmo circadiano de la presión arterial y su relación con los factores de riesgo cardiovascular. Enfermería nefrológica. 2019;22(2):151-8. Disponible: https://scielo.isciii.es/scielo.php?pid=S2254-28842019000200151&script=sci_arttext&tlng=pt
13. Small AM, Kiss DH, Zlatsin Y, Birtwell DL, Williams H, Guerraty MA, *et al.* Text mining applied to electronic cardiovascular procedure reports to identify patients with trileaflet aortic stenosis and coronary artery disease. Journal of biomedical informatics. 2017 [acceso 06/06/2021];72:77-84. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/28624641/>
14. Jonnagaddala J, Liaw ST, Ray P, Kumar M, Chang NW, Dai HJ. Coronary artery disease risk assessment from unstructured electronic health records using text mining. Journal of biomedical informatics. 2015;58:S203-S10. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1532046415001707>
15. Bagheri A, Groenhof TKJ, Asselbergs FW, Haitjema S, Bots ML, Veldhuis WB, *et al.* Automatic Prediction of Recurrence of Major Cardiovascular Events: A Text Mining Study Using Chest X-Ray Reports. Journal of Healthcare Engineering; 2021. Disponible en: <https://www.hindawi.com/journals/jhe/2021/6663884/>
16. Yadav D, Pandi G. A Survey Techniques Used for Prediction of Heart Attack with Machine Learning and Medical Text Mining. 2019. Disponible en: <https://www.ijert.org/research/a-survey-on-prediction-techniques-of-heart-disease-using-machine-learning-IJERTV9IS060298.pdf>

Conflicto de intereses

Los autores declaran que no tienen conflicto de intereses.