



C A P Í T U L O 5

DEL ESTADO DEL ARTE A LA INNOVACIÓN TECNOLÓGICA EN DETECCIÓN DE TRASTORNOS RESPIRATORIOS DEL SUEÑO

Adriana Yessica Luna-Ortiz

Benemérita Universidad Autónoma de Puebla, Doctorado
en Investigación y Educación para la Salud

Luis Enrique Colmenares-Guillén

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación

Ma. Carmen Cortés-Sánchez

Benemérita Universidad Autónoma de Puebla, Instituto de Fisiología

RESUMEN: Los Trastornos Respiratorios del Sueño son alteraciones frecuentes del sueño y constituyen un problema para la salud. Se han relacionado con enfermedades cardiovasculares, metabólicas y neurocognitivas. Estos trastornos del sueño impactan de manera negativa la calidad de vida de quienes los padecen. La presente investigación tiene como propósito revisar los principales conceptos relacionados con los Trastornos Respiratorios del Sueño, se analizan los avances recientes en el diagnóstico y se destaca la importancia de incorporar nuevas tecnologías que faciliten la detección temprana y fomenten la interdisciplinariedad.

PALABRAS CLAVE: Trastornos Respiratorios del Sueño, Tecnología, Ciencia de datos

1. INTRODUCCIÓN

Los Trastornos Respiratorios del Sueño (TRS) son un problema de salud a nivel mundial de gran importancia debido a las diversas enfermedades con las que se han asociado. La presencia de TRS se asocia con un incremento de hasta 1.5 veces en el riesgo de padecer enfermedades cardiovasculares como hipertensión, insuficiencia cardíaca y accidente cerebrovascular (Khokhrina et al., 2022). Se estima que cerca de mil millones de personas en el mundo con edades entre 30 y 69 años padecen Apnea Obstructiva del Sueño que es el tipo más común entre los TRS (Lyons et al., 2020). Estos trastornos impactan de manera individual a quienes los padecen, también generan problemas de pareja e incluso laborales. La comprensión de estos trastornos ayuda a identificarlos de manera oportuna y a la generación de estrategias para la prevención y el tratamiento.

En vista del impacto que tienen estos trastornos, resultó de gran interés elaborar un Estado del Arte (EdA) acerca de este tema con el propósito de identificar los hallazgos en torno a los TRS. La finalidad de esta investigación es proponer una estrategia para elaborar un EdA y destacar los hallazgos más relevantes. Dentro de los que destacan la incorporación y desarrollo de herramientas tecnológicas para el diagnóstico de los TRS. Se identificaron estrategias metodológicas orientadas al desarrollo de dichas herramientas que serán analizadas en el desarrollo de este trabajo.

La estructura de esta investigación está conformada por introducción donde se aborda la justificación del tema. En la sección dos se encuentra el desarrollo del tema en el cual se analizan aspectos teóricos y metodológicos utilizados en la actualidad. Por último, en la sección tres se encuentran las conclusiones que sintetizan los hallazgos de este trabajo.

2. DESARROLLO

2.1 Metodología para la elaboración del estado del arte

En el desarrollo de una investigación para el área de la salud se recolecta información. A través del análisis de la información recopilada se identifican aspectos generales como factores de riesgo, diagnóstico y tratamiento de cualquier enfermedad. Así mismo, se reconocen los esfuerzos realizados en diversas regiones a nivel mundial y se analizan los resultados de otros investigadores.

De acuerdo con el Oxford English Dictionary, el Estado del Arte o State of the Art es el estado más reciente en el desarrollo de un producto, que incorpora las ideas y características más novedosas (Oxford English Dictionary, 2012). El EdA es útil para la identificación de los avances más recientes relacionados con un tema en particular. La elaboración de un EdA va más allá de la simple recopilación de textos. Se trata de un análisis profundo y crítico de las investigaciones previas. Dicho análisis es de utilidad para identificar ideas principales en torno a un tema de investigación, también para detectar las perspectivas de otros investigadores e identificar áreas de oportunidad.

El EdA tiene como propósito identificar las tendencias en la generación de conocimiento y representa una descripción detallada del trabajo de diversos autores, así como alternativas metodológicas que se han empleado (Reyes & Enrique, 2019). De acuerdo con la descripción anterior, el EdA es un documento que exige desarrollar la capacidad crítica personal para construir argumentos derivados de un proceso de reflexión. En un futuro, dichos argumentos se convertirán en las bases que justifiquen la pertinencia de una investigación propia.

El primer paso para elaborar un EdA es la recolección de información. Esta etapa consiste en definir de dónde se obtendrá la información y bajo qué criterios, es decir, delimitar una estrategia de búsqueda. Se sugieren plataformas de acceso a textos científicos como Springer, PubMed y Elsevier, filtrar resultados por año de publicación considerando sólo aquellos cuya antigüedad sea menor a 10 años y definir palabras clave.

Una vez recolectada la información, se realiza una revisión de la literatura. Esta etapa consiste en una revisión general y es la base para la elaboración del Estado del Conocimiento (EdC) cuyo propósito es expresar el grado de desarrollo acerca de un tema y es útil para seleccionar y ordenar elementos debido a la visión panorámica que proporciona (Cerón-Martínez & López-Quiterio, 2015).

Posteriormente se realiza una lectura exploratoria útil para generar un primer acercamiento al contenido y seleccionar textos con potencial para ser analizados. Los textos seleccionados conformarán el EdC, se almacenan y se nombran mediante el año de publicación y título. Adicionalmente se organizan en carpetas acorde a la temática.

Una vez que se obtiene el EdC se realiza una lectura selectiva y se eligen los artículos con mayor pertinencia teórica. Dichos artículos son leídos a detalle, se analizan argumentos clave, hallazgos y se identifican relaciones entre conceptos.

A partir de la lectura a detalle se seleccionan los temas principales que se convertirán en los ejes temáticos y se organizan según los objetivos de la investigación.

Finalmente se elabora el EdA que es una revisión crítica de los avances más recientes y es de utilidad para la toma de decisiones en torno a la investigación.

El proceso de análisis de la información para la elaboración del EdA representado en la Fig. 1, es una guía estructurada en nueve pasos de carácter secuencial para el desarrollo del EdA. Esta manera de ordenar las lecturas fomenta un acercamiento a los aspectos teóricos de un tema de manera progresiva.



Fig. 1. Proceso de análisis de la información para la elaboración del Estado del Arte.

Fuente: Elaboración propia.

Derivado de la elaboración del EdA acerca de los TRS, se destacan conceptos generales que se identificaron en el proceso para comprender la relevancia del tema.

2.2 Panorama general de los Trastornos Respiratorios del Sueño

El sueño es un proceso fisiológico esencial, caracterizado por un estado reversible en el que se produce una desconexión de los estímulos externos y una disminución en la capacidad de respuesta al entorno (Mckinney et al., 2015; Y. B. Zheng et al., 2024). La Academia Americana de Medicina del Sueño (AASM) señala que es fundamental para el desarrollo de estilos de vida saludables, ya que un descanso adecuado fomenta la salud y contribuye a una mejor calidad de vida. En efecto, el sueño posibilita la recuperación física, también tiene implicaciones directas en las capacidades cognitivas y la prevención de enfermedades crónicas.

Las alteraciones del ciclo sueño–vigilia son una problemática para estudiar. Los trastornos del sueño son definidos por la AASM como la incapacidad de iniciar, consolidar o mantener el sueño de manera adecuada (American Academy of Sleep Medicine, 2024). Son frecuentes y se han relacionado con el aumento de hábitos que resultan nocivos para la salud, en especial para el sueño, entre ellos las amplias jornadas laborales, el exceso de exposición a dispositivos electrónicos, el estrés y el sedentarismo.

Dentro de la clasificación para los diversos trastornos del sueño, se encuentran los TRS que se caracterizan por la aparición de alteraciones en el patrón respiratorio que ocurren al dormir (Gállego Pérez-Larraya et al., 2007). Dentro de esta categoría

se encuentran la Apnea Obstructiva del Sueño (AOS) que se caracteriza por el bloqueo de la vía respiratoria superior tras el colapso intermitente de la faringe, la Apnea Central del Sueño (ACS) caracterizada por la interrupción de la señal neuronal proveniente de los centros de control respiratorio, lo que provoca episodios de apnea durante el sueño y la Apnea Mixta (AM) que cuando la AOS y la ACS se presentan de manera simultánea (Mohamed et al., 2024; Patel et al., 2025; Winter et al., 2025) .

A nivel mundial, el 45% de la población padece algún tipo de trastorno del sueño, siendo la AOS la que se presenta con mayor frecuencia (Álvarez-García & Jiménez-Correa, 2020). Los TRS afectan la calidad de vida de quienes los padecen ya que producen consecuencias físicas, psicológicas, sociales, de pareja e incluso económicas (Yathish & Manjula, 2024). Aunado a lo anterior, se han relacionado con desórdenes metabólicos como resistencia a la insulina, hipertensión arterial, diabetes y obesidad (Ott et al., 2017; Yathish & Manjula, 2024). En relación con el impacto que tienen estos trastornos en la salud individual y colectiva, resulta cuestionable hasta qué punto las herramientas de diagnóstico que existen actualmente responden a esta problemática.

2.3 Panorama general de ciencia de datos

El término tecnología se refiere al conjunto de herramientas, técnicas y conocimientos aplicados para resolver problemas o realizar tareas (Hund et al., 2021) . La creación y uso de cualquier herramienta tecnológica favorece el mejoramiento de habilidades. La tecnología se clasifica en tecnología tradicional cuyo propósito principal es la transformación física de materiales o energía en un bien o servicio como los molinos de viento, la imprenta o la maquinaria industrial. En contraste, la tecnología digital incluye el procesamiento automatizado, la creación de algoritmos y la conectividad de tal manera que la tecnología digital gestiona los datos.

La ciencia de datos surgió a partir de la generación masiva de información y la imposibilidad de gestionarla con medios tradicionales. Se encarga de la gestión y análisis de datos con el propósito de identificar y comprender patrones para obtener decisiones informadas y mejorar los servicios de un sistema, por ejemplo, el sistema de salud (Dash et al., 2019). De acuerdo con la definición que se menciona en Vidal-Ledo et al., 2023, la ciencia de datos representa un enfoque multidisciplinario que integra principios matemáticos, estadística, computación, aprendizaje automático, minería de datos, entre otros. En la actualidad, la ciencia de datos se ha convertido en un eje transversal para la optimización del área de la salud.

La industria de la salud ha generado grandes cantidades de datos de manera histórica debido a la necesidad de mantener registros, cumplir requisitos regulatorios, así como la atención al paciente. La tendencia actual es digitalizar la gran cantidad

de datos que se generan a partir de diversas fuentes que se clasifican de la siguiente manera: a) datos clínicos y de registro del paciente que son los que se generan durante la atención directa como las notas escritas por médicos y enfermeras, datos de admisión y alta hospitalaria, informes de patología, radiografías, resonancias magnéticas, datos de laboratorio y farmacia, b) datos generados por máquinas y sensores que provienen de dispositivos de monitorización de signos vitales, provenientes de monitoreo en tiempo real o de dispositivos portátiles así como datos biométricos o imágenes en 3D, c) datos administrativos que incluyen información de seguros, datos financieros y operacionales y d) datos generales generados por humanos y fuentes externas que frecuentemente son datos no estructurados como notas médicas y correos electrónicos, entre otros (Raghupathi & Raghupathi, 2014). Derivado de la amplia diversidad de los datos y la necesidad de que los sistemas de información de salud sean capaces de intercambiar datos, surge el concepto de interoperabilidad (Rosa & Frutos, 2022). Dicho concepto representa un pilar indispensable para el acceso y uso de datos actualizados; sin embargo, para lograr esto es necesario contar con infraestructura adecuada y un marco normativo que proteja la privacidad de los datos.

La interoperabilidad es un requisito para la gestión de datos y forma parte de los principios guía FAIR por sus siglas en inglés (Findable, Accesible, Interoperable, Reusable). El acrónimo FAIR se refiere a las características necesarias para los datos: a) Findable o localizable, lo que significa que deben registrarse o indizarse específicamente en un recurso, b) Accesible, que hace referencia a que se garantice el acceso de los datos para la comunidad científica, c) Interoperable, que es el uso de lenguaje formal para la estandarización de los datos, d) Reusable, que implica el desarrollo de herramientas reproducibles. La interoperabilidad pone de manifiesto dos desafíos. El primero de ellos es la estandarización, debido a que los datos provienen de diferentes fuentes, es indispensable armonizar la información y mantener un lenguaje universal para hacer eficiente la colaboración. El segundo es la vinculación ya que la ausencia de un identificador único por paciente es el principal obstáculo (Shilo et al., 2020).

La ciencia de datos utiliza el aprendizaje automático para identificar patrones. Un modelo de aprendizaje automático aprende patrones a partir de los datos de un gran número de pacientes. Como resultado, esta herramienta tiene la capacidad de anticipar eventos basándose en información que va más allá de la opinión de expertos. Adicionalmente, a nivel poblacional identifica pacientes que desarrollarán condiciones de alto riesgo (Rajkomar et al., 2019). Por lo tanto, el aprendizaje automático proporciona información para la identificación de enfermedades y para el desarrollo de estrategias preventivas.

Hoy en día, la ciencia de datos se ha convertido en un eje transversal para optimizar diversas áreas, entre ellas, el área de la salud. La integración de la ciencia de datos y el aprendizaje automático en el estudio de los TRS conlleva a un avance en el diagnóstico y tratamiento de estos trastornos a través del análisis de datos fisiológicos, antropométricos, demográficos, síntomas, comorbilidades, entre otros (Korkalainen et al., 2024). En la Fig.2 se observa la tecnología como el origen de la ciencia de datos

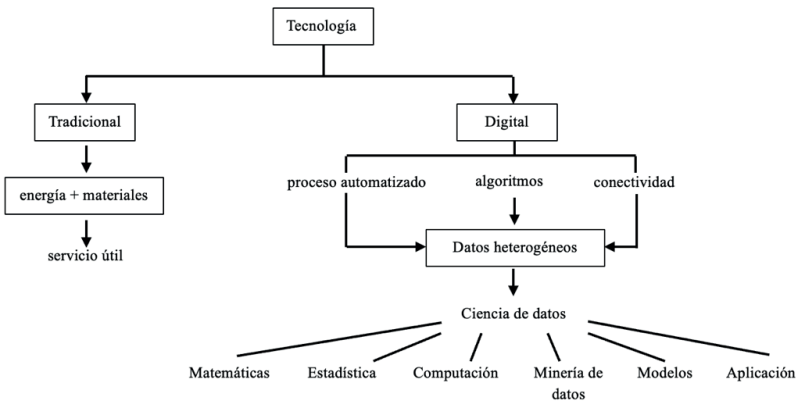


Fig. 2. Panorama general de ciencia de datos

2.4 Herramientas tecnológicas en el diagnóstico de los Trastornos Respiratorios del Sueño

La incorporación de herramientas tecnológicas para el diagnóstico de los TRS y las estrategias metodológicas que orientan su desarrollo son temas que destacan por su relevancia en la actualidad. El problema principal en torno al diagnóstico de los TRS tiene que ver con el sub diagnóstico y la falta de acceso a pruebas especializadas. Hoy en día la polisomnografía sigue siendo el estándar de oro. A pesar de ello, su costo elevado y la poca disponibilidad de centros especializados dificultan que la población tenga acceso a una evaluación oportuna. Como resultado, una gran parte de personas que padecen TRS ni siquiera lo saben y en ocasiones tienen acceso al diagnóstico posterior a la aparición de complicaciones sistémicas que probablemente se pudieron prevenir. Esta situación genera la necesidad de desarrollar estrategias más accesibles para diagnóstico que permitan la detección temprana de los TRS.

En una publicación se examinaron las limitaciones de los métodos de diagnóstico actuales para la detección de AOS, principalmente de la polisomnografía que, como se mencionó previamente, es costosa y poco accesible. Los autores de dicha

publicación proponen el uso de las herramientas tecnológicas como una alternativa para mejorar tanto el diagnóstico como el tratamiento de los TRS. Sin embargo, se menciona que se han identificado seis barreras para la adopción de las herramientas tecnológicas, entre ellas la tecnología disponible, los recursos humanos, el nivel de educación y cultura acerca del uso de estas herramientas, la regulación de los datos usados para el desarrollo de modelos y su calidad (Brennan & Kirby, 2023). Para que un conjunto de datos sea de calidad tiene que cumplir con características como validez, precisión, integridad y ser adecuado para el propósito que se utiliza (IBM, 2025). La identificación de estas seis barreras hace evidente que la implementación de las herramientas tecnológicas en el área de la salud será un proceso paulatino. Cada una de estas barreras representa un desafío. Por ejemplo, la falta de datos de calidad para el entrenamiento de los modelos genera fallas en el resultado y está relacionada con a) factores económicos que afectan la recolección y almacenamiento de datos, b) falta de capacitación y c) la resistencia cultural al uso de nuevas tecnologías por parte del personal de salud.

2.5 Perspectiva general para el desarrollo de herramientas tecnológicas

Recientemente, se publicó un proceso de nueve pasos para el desarrollo de herramientas tecnológicas en el área de la salud (Chew & Ngiam, 2025). Véase Fig. 3

Los autores hacen énfasis en la necesidad de que el personal de salud comprenda y participe en el proceso para diseñar estas herramientas y señalan que la información que los clínicos proporcionan en el diseño es invaluable. La participación del personal de salud garantiza que estas herramientas sean integradas con facilidad en la atención médica y, por si fuera poco, su participación fomenta la aceptación entre otros proveedores de salud.

El primer paso para el desarrollo de herramientas tecnológicas es la identificación de problemas de salud adecuados para soluciones con el uso de la tecnología, es decir, problemas que se resuelvan con una solución automatizada. Una vez identificado el problema, se forma un equipo de investigación que incluya expertos en el manejo de datos y expertos en el área de la salud.

El tercer paso se relaciona con la gestión de datos. Para la generación de modelos es indispensable tener datos de calidad. Si existen faltantes en los conjuntos de datos, se utilizan diversas estrategias metodológicas. El establecimiento de una infraestructura y arquitectura adecuada que se menciona en el cuarto paso hace referencia a la capacidad de las computadoras y los servidores para operar grandes volúmenes de datos a una velocidad óptima.

El quinto paso indica que el modelo seleccionado es adecuado para resolver la problemática acorde al tipo de datos. El sexto paso consiste en la validación de los modelos para lo cual se evalúa su rendimiento. También se requieren estudios de investigación clínica para evaluar su utilidad en el mundo real.

Como siguiente paso, se realiza el registro de los modelos con las autoridades reguladoras para que estos modelos sean implementados de manera exitosa en la atención médica. Los criterios de evaluación varían según la regulación de cada país. No obstante, existen principios éticos para el desarrollo de cualquier herramienta tecnológica como el principio de no maleficencia, autonomía, justicia, gobernanza, privacidad de los datos, diversidad, inclusión, transparencia, fiabilidad, equidad, bienestar social y la rendición de cuentas.

La implementación clínica y el monitoreo continuo del rendimiento del modelo constituyen el octavo paso. La implementación es el método mediante el cual las herramientas tecnológicas se integran en un flujo de trabajo clínico, mientras que el monitoreo continuo es indispensable para identificar efectos indeseados y garantizar la integridad de los datos.

El paso número nueve se enfoca en asegurar la adaptabilidad de las herramientas a las necesidades clínicas, cuya principal característica es el cambio constante.

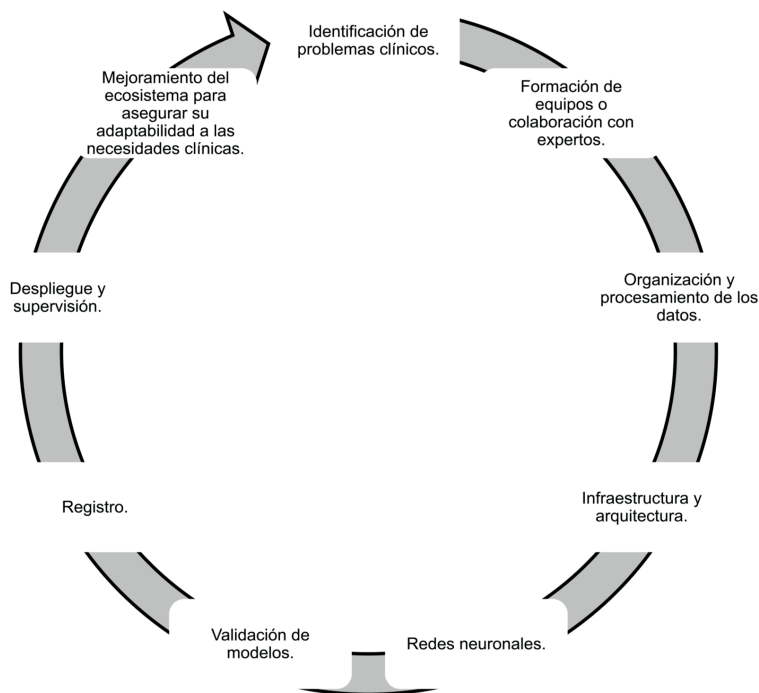


Fig 3. Proceso para el desarrollo de herramientas tecnológicas.
Adaptado de: (Chew & Ngiam, 2025)

Esta serie de pasos propuestos por (Chew & Ngiam, 2025) son iterativos de tal forma que se adaptan a los cambios. La generación de un entorno propicio para la implementación de herramientas tecnológicas en el área de la salud asegura que las herramientas sean adaptables, confiables y brinden beneficios a largo plazo. El modelo de los nueve pasos para el desarrollo de herramientas tecnológicas descrito previamente proporciona lineamientos con fundamento ético; sin embargo, su éxito dependerá de la superación de algunas limitaciones, entre ellas el nivel del compromiso de los profesionales involucrados, el desarrollo de políticas para la regulación de estas herramientas y las limitaciones con respecto al uso y disponibilidad de datos.

Aunado a lo anterior, en un estudio publicado en 2022, los autores mencionan que, si bien la tecnología se presenta como un medio para fortalecer los sistemas de salud en países poco favorecidos, aún existen diversos desafíos entre los que destaca la escasez de grandes cantidades de conjuntos de datos completos y actualizados (Ciecierski-Holmes et al., 2022).

2.6 Estrategias metodológicas para la gestión de datos

La disponibilidad y acceso a los datos, así como la cantidad y características de los datos generan limitaciones para el entrenamiento de modelos. Una de las estrategias metodológicas propuestas para superar esta limitación es la armonización de datos.

La armonización de datos se define como la práctica de combinar diferentes conjuntos de datos para maximizar su compatibilidad; al aumentar el volumen y la diversidad de los datos es posible hacer generalizaciones (Cheng et al., 2024). En 2024, un grupo de investigadores propuso tres pasos básicos para la armonización de datos. Los pasos son adaptables y tienen la capacidad de complementarse según las necesidades de cada caso. Véase Fig. 4.

Durante la etapa de preparación es necesario considerar las habilidades deseadas y el número de sujetos que integrará el equipo de investigación encargado de la armonización de datos; los recursos materiales que se requerirán acorde al número de conjuntos de datos que se analizará; los objetivos de la armonización y el acceso a los datos. Se decide si la armonización será de tipo retrospectiva o prospectiva. Esta elección dependerá de la disponibilidad de los datos, así como del tiempo y los recursos disponibles.

En la etapa de implementación, el equipo de investigación define qué variables se desean armonizar con base en el objetivo general planteado, la disponibilidad de los datos y el nivel de armonización que se planteó previamente. Para la armonización retrospectiva es necesario que sean comparables los datos existentes con los previos, por lo que, es indispensable resolver diferencias en los formatos, en la sintaxis, en la

estructura y en el significado o semántica de los datos. Posterior a la armonización de estas dimensiones, es posible combinar los diferentes conjuntos de datos. En contraste, para la armonización prospectiva se requiere homogenizar la metodología de recolección de datos y considerar la sintaxis, estructura y semántica antes de la recolección, asegurando que se sigan los protocolos en lugar de hacer correcciones posteriores. Una vez recolectados los datos, se combinan.

En la etapa post armonización se validan los datos y se asegura que el resultado sea un conjunto de datos en el que las variables se han hecho compatibles y comparables para el análisis. Los procedimientos para validar dicha situación pueden incluir la replicación y verificación del procedimiento de armonización en una submuestra de datos o el empleo de aprendizaje automático. También es necesario documentar realizando manuales de manera que transparenten las decisiones metodológicas que se realizaron para la armonización con la intención de que futuros usuarios comprendan cómo se utiliza el conjunto de datos armonizados, así como sus limitaciones.

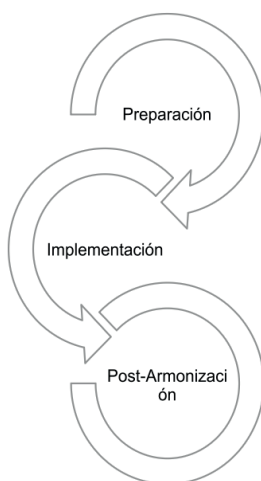


Fig. 4. Pasos básicos para la armonización de datos. Adaptado de: (Cheng et al., 2024).

La armonización de datos es una estrategia metodológica que busca la integración de múltiples conjuntos de datos. A pesar de ser un proceso complejo, resulta un paso indispensable en investigaciones que buscan integrar datos de diversas fuentes. La propuesta planteada previamente es de gran utilidad puesto que surgió a partir de una investigación en el área de la salud, aunque sería de utilidad incorporar aspectos éticos y legales para el uso y protección de los datos e incorporar estrategias para el procesamiento de datos médicos.

Las estrategias metodológicas que se implementan en el procesamiento de datos se eligen una vez que realizó un análisis exploratorio, que representa un paso crítico para comprender la distribución, estructura y particularidades de los conjuntos de datos (Konopka et al., 2018). En la Fig. 5 se propone un enfoque estructurado de aprendizaje automático para la construcción de un modelo predictivo en un entorno clínico.

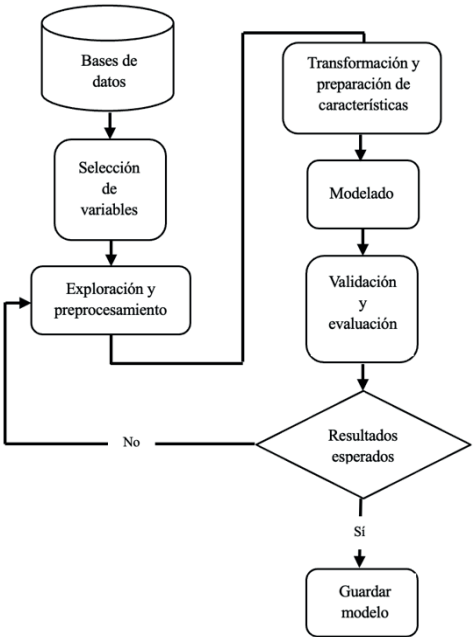


Fig. 5. Metodología para la construcción de un modelo predictivo

Una vez realizado el análisis exploratorio de datos se realiza el preprocesamiento. Durante el análisis exploratorio de datos se identifica la proporción de valores faltantes y en el preprocesamiento se elige la técnica para tratarlos. Si la proporción de datos faltantes representa menos del 5% del conjunto total de datos, se utiliza el método K-Nearest Neighbors (KNN) en lugar de eliminar de manera directa los valores faltantes (Park et al., 2025). El método KNN es un clasificador de aprendizaje supervisado no paramétrico que considera las relaciones entre variables y completa los datos faltantes en función de pacientes con características similares. Este método mantiene la consistencia interna del conjunto de datos y reduce la pérdida de información (Nagarajan & Dhinesh Babu, 2022). También es indispensable hacer el tratamiento de valores atípicos. Dada la naturaleza de los datos médicos, el

tratamiento de valores atípicos requiere un enfoque permisivo a través de un rango intercuartílico ampliado ($IQR \times 3$) con la finalidad de identificar valores extremos que podrían representar condiciones clínicas reales (Rajkomar et al., 2019). Adicionalmente se utilizan técnicas de ingeniería de características que consiste en la selección y formulación de características considerando el tipo de datos y el modelo de aprendizaje que se utilizará para mejorar su desempeño y eficiencia (A. Zheng & Casari, 2018).

Para garantizar que el modelo sea entrenado con un conjunto de datos representativo y balanceado, es de utilidad ajustar la distribución mediante muestreo estratificado utilizando variables demográficas como criterios de estratificación. Esta estrategia mantiene la representatividad demográfica, mejora la capacidad del modelo para generalizar y añade la integridad del análisis clínico al asegurar una cobertura adecuada para cada subgrupo poblacional (Patiño-Pérez et al., 2023). También es elemental balancear la distribución de clases en un conjunto de datos ya que los modelos son sensibles a las distribuciones desbalanceadas, lo que compromete su rendimiento. Para ello se implementan estrategias como SMOTE-Tomek, una técnica híbrida de balance de datos que consiste en generar sobre muestreo de clases minoritarias mediante la técnica SMOTE que significa Synthetic Minority Oversampling Technique. A continuación, en el conjunto de datos sobre muestreado se identifican enlaces Tomek que consisten en el solapamiento de datos. Los datos solapados se eliminan como un método de limpieza. El resultado es un conjunto de datos equilibrado (Batista et al., 2004; He & Garcia, 2009).

Una vez que los datos están listos, se divide el conjunto de datos. El conjunto de entrenamiento corresponde a la porción más grande de los datos ya que a través de este conjunto el modelo identificará patrones. El conjunto de prueba se utiliza para evaluar el rendimiento del modelo entrenado. La proporción 80:20 se reconoce como una de las formas más habituales de dividir los datos entre entrenamiento y prueba y garantiza que el modelo aprenda bien y, al mismo tiempo, que sea probado (Sivakumar et al., 2024). Véase Fig. 6

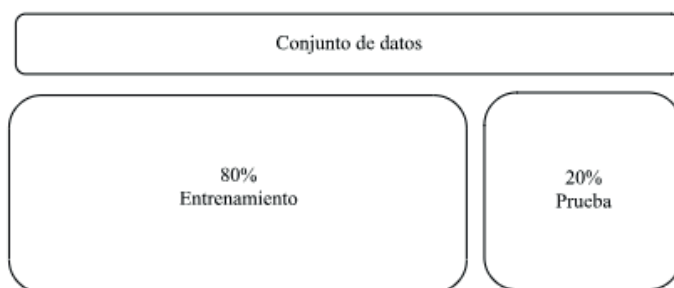


Fig. 6. División del conjunto de datos

3. CONCLUSIONES

El EdA es un documento que exige desarrollar la capacidad crítica para construir argumentos derivados de un proceso de reflexión que se convertirán en las bases que justifiquen la pertinencia de una investigación.

Los TRS se presentan con frecuencia a nivel mundial y tienen implicaciones médicas, sociales y económicas. Dichas implicaciones convierten a los TRS en un problema de salud que requiere estrategias integrales para la búsqueda de una solución. La elaboración de un EdA acerca de este tema fue de utilidad para identificar los hallazgos más recientes.

El conocimiento del procedimiento para el desarrollo de herramientas tecnológicas y de estrategias metodológicas para la gestión de datos amplía las competencias del personal de salud con respecto al desarrollo de la tecnología y su implementación en la práctica clínica.

La comprensión de las dificultades que existen en torno al uso de datos es fundamental para comprender las limitaciones de algunos modelos para el diagnóstico de los TRS y promover la participación del personal de salud en la mejora de estos.

La aplicación de estrategias metodológicas en el procesamiento de datos como KNN, ingeniería de datos y SMOTE-Tomek garantiza la validez de los modelos predictivos. La división de datos asegura que el modelo disponga de datos suficientes para entrenamiento y validación.

Finalmente, la colaboración entre especialistas en la gestión de datos y el personal de salud es esencial para avanzar hacia un futuro donde el desarrollo de herramientas tecnológicas sea implementado en la práctica clínica de manera exitosa.

REFERENCIAS

Álvarez-García, H. B., & Jiménez-Correa, U. (2020). Intervención psicológica en trastornos del sueño: una revisión actualizada. *Revista Clínica Contemporánea*, 11(2). <https://doi.org/10.5093/cc2020a9>

American Academy of Sleep Medicine. (2024, September 12). *Sueño saludable: educación sobre el sueño por parte de la AASM*. <https://sleepeducation.org/healthy-sleep/>

Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, 6(1), 20–29. <https://doi.org/10.1145/1007730.1007735>

Brennan, H. L., & Kirby, S. D. (2023). The role of artificial intelligence in the treatment of obstructive sleep apnea. In *Journal of Otolaryngology - Head and Neck Surgery* (Vol. 52, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s40463-023-00621-0>

Cerón-Martínez, A. U., & López-Quiterio, A. E. (2015). El uso del estado del conocimiento en investigación educativa y su relación con las tradiciones de investigación como ejes orientadores. *Revista Latinoamericana de Metodología de La Investigación Social*.

Cheng, C., Messerschmidt, L., Bravo, I., Waldbauer, M., Bhavikatti, R., Schenk, C., Grujic, V., Model, T., Kubinec, R., & Barceló, J. (2024). A General Primer for Data Harmonization. In *Scientific Data* (Vol. 11, Issue 1). Nature Research. <https://doi.org/10.1038/s41597-024-02956-3>

Chew, B. H., & Ngiam, K. Y. (2025). Artificial intelligence tool development: what clinicians need to know? In *BMC Medicine* (Vol. 23, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s12916-025-04076-0>

Ciecierski-Holmes, T., Singh, R., Axt, M., Brenner, S., & Barteit, S. (2022). Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. In *npj Digital Medicine* (Vol. 5, Issue 1). Nature Research. <https://doi.org/10.1038/s41746-022-00700-y>

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0217-0>

Gállego Pérez-Larraya, J., Toledo, J. B., Urrestarazu, E., & Iriarte, J. (2007). Clasificación de los trastornos del sueño Classification of sleep disorders. In *An. Sist. Sanit. Navar* (Vol. 30, Issue 1).

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>

Hund, A., Wagner, H. T., Beimborn, D., & Weitzel, T. (2021). Digital innovation: Review and novel perspective. In *Journal of Strategic Information Systems* (Vol. 30, Issue 4). Elsevier B.V. <https://doi.org/10.1016/j.jsis.2021.101695>

IBM. (2025). ¿Qué es la calidad de los datos? IBM. <https://www.ibm.com/mx-es/think/topics/data-quality>

Khokhrina, A., Andreeva, E., & Degryse, J. M. (2022). A systematic review on the association of sleep-disordered breathing with cardiovascular pathology in adults. In *npj Primary Care Respiratory Medicine* (Vol. 32, Issue 1). Nature Research. <https://doi.org/10.1038/s41533-022-00307-6>

Konopka, B. M., Lwow, F., Owczarz, M., & Łaczmanski, Ł. (2018). Exploratory data analysis of a clinical study group: Development of a procedure for exploring multidimensional data. *PLoS ONE*, 13(8). <https://doi.org/10.1371/journal.pone.0201950>

Korkalainen, H., Kainulainen, S., Islind, A. S., Óskarsdóttir, M., Strassberger, C., Nikkonen, S., Töyräs, J., Kulkas, A., Grote, L., Hedner, J., Sund, R., Hrubos-Strom, H., Saavedra, J. M., Ólafsdóttir, K. A., Ágústsson, J. S., Terrill, P. I., McNicholas, W. T., Arnardóttir, E. S., & Leppänen, T. (2024). Review and perspective on sleep-disordered breathing research and translation to clinics. In *Sleep Medicine Reviews* (Vol. 73). W.B. Saunders Ltd. <https://doi.org/10.1016/j.smr.2023.101874>

Lyons, M. M., Bhatt, N. Y., Pack, A. I., & Magalang, U. J. (2020). Global burden of sleep-disordered breathing and its implications. In *Respirology* (Vol. 25, Issue 7, pp. 690–702). Blackwell Publishing. <https://doi.org/10.1111/resp.13838>

Mckinney, J., Ortiz-Young, D., & Jefferson, F. (2015). Gender differences in obstructive sleep apnea and the associated public health burden. In *Sleep and Biological Rhythms* (Vol. 13, Issue 3, pp. 196–209). Blackwell Publishing. <https://doi.org/10.1111/sbr.12107>

Mohamed, B., Yarlagadda, K., Self, Z., Simon, A., Rigueiro, F., Sohooli, M., Eisenschenk, S., & Doré, S. (2024). Obstructive Sleep Apnea and Stroke: Determining the Mechanisms Behind their Association and Treatment Options. In *Translational Stroke Research* (Vol. 15, Issue 2, pp. 239–332). Springer. <https://doi.org/10.1007/s12975-023-01123-x>

Nagarajan, G., & Dhinesh Babu, L. D. (2022). Missing data imputation on biomedical data using deeply learned clustering and L2 regularized regression based on symmetric uncertainty. *Artificial Intelligence in Medicine*, 123. <https://doi.org/10.1016/j.artmed.2021.102214>

Ott, S. R., Korostovtseva, L., Schmidt, M., Horvath, T., Brill, A. K., & Bassetti, C. L. (2017). Sleep-disordered breathing: Clinical features, pathophysiology and diagnosis. *Swiss Medical Weekly*, 147. <https://doi.org/10.4414/smw.2017.14436>

Oxford English Dictionary. (2012). *state-of-the-art*, *adj.* & *n.* meanings, etymology and more | *Oxford English Dictionary*. https://www.oed.com/dictionary/state-of-the-art_adj?tab=meaning_and_use-paywall#20895388

Park, C., Byun, J. I., Choi, S. H., & Shin, W. C. (2025). Machine learning classifier solving the problem of sleep stage imbalance between overnight sleep. *Biomedical Engineering Letters*. <https://doi.org/10.1007/s13534-025-00466-8>

Patel, S. R., Genuardi, M. V., DeSensi, R., Fudim, M., Lund, L. H., Mehra, R., Germany, R., & McKane, S. (2025). Real world characteristics of central sleep apnea: a review of 29,000 consecutive patients undergoing polysomnography. *Sleep and Breathing*, 29(5), 296. <https://doi.org/10.1007/s11325-025-03446-5>

Patiño-Pérez, D., Iñiguez-Muñoz, F., Ochoa-Flores, Á., Córdova-Aragundi, J., Castro-Carrasco, J., Luque-Letechi, A., & Munive-Mora, C. (2023). Stratification for the Improvement of the Performance of an ANN in Diabetes Detection. *Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology, 2023-July*. <https://doi.org/10.18687/laccei2023.1.1.1308>

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/nejmra1814259>

Reyes, G., & Enrique, C. (2019). Estrategia metodológica para elaborar el estado del arte como un producto de investigación educativa. *Praxis Educativa (Arg)*, 23(3). <https://doi.org/10.19137/praxiseducativa-2019-230307>

Rosa, J. M., & Frutos, E. L. (2022). Health data science: Challenges and opportunities in Latin America. *Revista Médica Clínica Las Condes*, 33(6), 591–597. <https://doi.org/10.1016/j.rmcl.2022.09.007>

Shilo, S., Rossman, H., & Segal, E. (2020). Axes of a revolution: challenges and promises of big data in healthcare. In *Nature Medicine* (Vol. 26, Issue 1, pp. 29–38). Nature Research. <https://doi.org/10.1038/s41591-019-0727-5>

Sivakumar, M., Parthasarathy, S., & Padmapriya, T. (2024). Trade-off between training and testing ratio in machine learning for medical image processing. *PeerJ Computer Science*, 10. <https://doi.org/10.7717/PEERJ-CS.2245>

Vidal-Ledo, M. J., Delgado-Ramos, A., Gutiérrez-Vera, D., & Rodríguez-Díaz, A. (2023, March 1). Ciencia de Datos en Salud. *Educación Médica Superior*. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412023000100009

Winter, Y., Mayer, G., Benes, H., Burghaus, L., Floam, S., Parks, G. S., & Kallweit, U. (2025). Patients with obstructive sleep apnea in Germany. *Sleep and Breathing*, 29(1). <https://doi.org/10.1007/s11325-025-03275-6>

Yathish, T., & Manjula, C. (2024). Risk Assessment of Obstructive Sleep Apnoea Symptoms and Its Correlation with Oral Manifestations: A Cross-sectional Study. *Journal of Clinical and Diagnostic Research*, 9–12. <https://doi.org/10.7860/jcdr/2024/68767.19226>

Zheng, A., & Casari, A. (2018). *Feature Engineering for Machine Learning PRINCIPLES AND TECHNIQUES FOR DATA SCIENTISTS* (R. Roumeliotis & J. Bleiel, Eds.; 1st ed.). O'Reilly Media, Inc.

Zheng, Y. B., Huang, Y. T., Gong, Y. M., Li, M. Z., Zeng, N., Wu, S. L., Zhang, Z. B., Tian, S. S., Yuan, K., Liu, X. X., Vitiello, M. V., Wang, Y. M., Wang, Y. X., Zhang, X. J., Shi, J., Shi, L., Yan, W., Lu, L., & Bao, Y. P. (2024). Association of lifestyle with sleep health in general population in China: a cross-sectional study. *Translational Psychiatry*, 14(1). <https://doi.org/10.1038/s41398-024-03002-x>