

<https://doi.org/10.17488/RMIB.44.4.4>

E-LOCATION ID: 1378

Sistema de Cribado Primario para la Sarcopenia en Personas Adultas Mayores Basado en Inteligencia Artificial

Primary Screening System for Sarcopenia in Elderly People Based on Artificial Intelligence

Santiago Arceo-Díaz¹, Elena Elsa Bricio-Barrios², Xochitl Angélica Rosio Trujillo-Trujillo¹✉,
José Ramón González-Farías¹, Jaime Alberto Bricio-Barrios¹, Mónica Ríos-Silva¹, Miguel Huerta-Viera¹

¹Universidad de Colima, Colima - México

²Tecnológico Nacional de México campus Colima, Colima - México

RESUMEN

Este estudio propone un sistema de cribado primario para diagnosticar la sarcopenia en adultos mayores a través de medidas antropométricas. Esta investigación exploratoria involucró inicialmente a 150 personas de edad avanzada, de las cuales 122 fueron seleccionadas después de un proceso de depuración de datos. Empleando técnicas de aprendizaje automático como el agrupamiento jerárquico y los árboles de decisión, se redujeron las 13 medidas antropométricas originales a cinco características clave. Se crearon tres sistemas de clasificación: el primero basado en parámetros previamente establecidos (masa muscular apendicular, velocidad de marcha y fuerza de agarre); el segundo consideró medidas de las extremidades superiores (masa muscular promedio de ambos brazos, fuerza de agarre, velocidad de marcha y porcentaje de grasa corporal); y el tercero se centró en las medidas de las extremidades inferiores (masa muscular promedio de ambas piernas, fuerza de agarre, velocidad de marcha y porcentaje de grasa corporal). Estos sistemas de clasificación se validaron clínicamente en un grupo de 57 pacientes previamente diagnosticados por especialistas, de los cuales 10 recibieron un diagnóstico positivo de sarcopenia. Los resultados mostraron eficiencias similares en los tres sistemas, con ocho de los diez diagnósticos positivos conocidos clasificados en el mismo grupo. Además, el estudio proporciona puntos de corte específicos para cada sistema, facilitando así el diagnóstico clínico de la sarcopenia por parte de profesionales médicos.

PALABRAS CLAVE: clustering jerárquico, medidas antropométricas, personas adultas mayores, Random Forest, sarcopenia

ABSTRACT

This study proposes a primary screening system for diagnosing sarcopenia in older adults through anthropometric measures. This exploratory research initially involved 150 elderly individuals, of whom 122 were selected after a data purification process. Using machine learning techniques such as hierarchical clustering and decision trees, the original set of 13 anthropometric measures was reduced to five key features. Three classification systems were created: the first based on previously established parameters (appendicular muscle mass, walking speed, and grip strength); the second considered upper limb measures (average muscle mass of both arms, grip strength, walking speed, and body fat percentage); and the third focused on lower limb measures (average muscle mass of both legs, grip strength, walking speed, and body fat percentage). These classification systems were clinically validated in a group of 57 patients previously diagnosed by specialists, of which 10 received a positive sarcopenia diagnosis. The results showed similar efficiencies in all three systems, with eight of the ten known positive diagnoses classified in the same group. Additionally, the study provides specific cut-off points for each system, thus facilitating the clinical diagnosis of sarcopenia by medical professionals.

KEYWORDS: anthropometric measures, elderly people, hierarchical clustering, Random Forest, sarcopenia

Autor de correspondencia

DESTINATARIO: Xochitl Angélica Rosio Trujillo-Trujillo
INSTITUCIÓN: Universidad de Colima
DOMICILIO: Centro Universitario de Investigaciones Biomédicas. Avenida 25 de Julio # 965, Colonia Las Víboras, Colima, Col. México
CORREO ELECTRÓNICO: rosio@ucol.mx

Recibido:

01 Octubre 2023

Aceptado:

22 Noviembre 2023

INTRODUCCIÓN

En la actualidad, las mejoras en la atención de los sistemas de salud, los avances médicos y los programas de planificación familiar han contribuido a incrementar la expectativa de vida en países emergentes como México ^[1]. Sin embargo, al prolongarse la esperanza de vida, se ha vuelto cada vez más frecuente que los sistemas de salud brinden atención a pacientes que sufren de condiciones asociadas con el envejecimiento natural, las cuales pueden tener un impacto negativo importante en sus vidas ^[2].

La sarcopenia es una enfermedad progresiva y generalizada del músculo esquelético, caracterizada por una disminución de la fuerza, de la masa muscular y, finalmente, del rendimiento físico y se asocia al envejecimiento ^[3]. Además, en personas adultas mayores, la sarcopenia está relacionada con la degradación proteica de los músculos, procesos inflamatorios, alteraciones hormonales y comorbilidades ^[4]. La disminución de la masa muscular suele tener como consecuencia una reducción en la capacidad de las personas para desarrollar actividades de la vida cotidiana, aspectos como la velocidad de marcha o la capacidad para sujetar objetos que dependan de la fuerza de agarre ^[5]. De acuerdo con el grado de deterioro, la sarcopenia puede clasificarse como pre-sarcopenia, sarcopenia y sarcopenia severa que, en su estadio más severo, las afectaciones producidas por la sarcopenia pueden afectar la vida de las personas o inducir caídas o fracturas, que pueden llevar al fallecimiento prematuro ^[6].

En años recientes, diversos avances tecnológicos han facilitado el diagnóstico de la sarcopenia. Esta capacidad de diagnóstico se fundamenta en la detección de cambios en la composición corporal mediante técnicas como la medición de la bioimpedancia bioeléctrica, la resonancia magnética o las tomografías computarizadas ^[7]. Sin embargo, en muchos casos, el costo relativo de los equipos de medición es demasiado alto para la mayoría de los hospitales o clínicas.

Esta problemática ha sido abordada por diversos grupos de trabajo que han identificado características y valores de referencia para el diagnóstico primario de sarcopenia. La Tabla 1 muestra los principales estudios que recopiló Aquiles *et al.* en ^[8].

En la segunda, tercera y cuarta columna reporta las medidas antropométricas y puntos de corte como: FPP: Fuerza de presión palmar en kg; IMMA: Índice de masa muscular apendicular en $\text{kg}\cdot\text{m}^{-2}$; MA/IMC: corresponde a Masa apendicular sobre Índice de masa corporal, adimensional, MMA: Masa muscular apendicular en $\text{kg}\cdot\text{m}^{-2}$;

TABLA 1. Diagnóstico primario de sarcopenia.

| Grupo de trabajo | Masa muscular | Fuerza muscular | Rendimiento físico |
|--|--|---|-------------------------------|
| Grupo Europeo de Trabajo sobre Sarcopenia en Personas Mayores | MMA: < 20 en hombres y < 15 en mujeres | FPP: ≤ 27 en hombres y < 16 en mujeres | VC < 0.8 en hombres y mujeres |
| Grupo Asiático de Trabajo sobre Sarcopenia en Personas Mayores | IMMA: ≤ 7 en hombres y ≤ 5.4 en mujeres | FPP: < 26 en hombres y < 18 en mujeres | VC ≤ 0.8 en hombres y mujeres |
| Fundación para el Instituto Nacional de Salud | MA/IMC: <0.789 en hombres y <0.512 en mujeres | FPP: < 26 en hombres y <16 en mujeres | VC ≤ 0.8 en hombres y mujeres |
| Grupo con interés especial | ≤ 2 desviaciones estándar del promedio del porcentaje de masa muscular | No se considera | VC ≤ 0.8 en hombres y mujeres |
| Grupo internacional de sarcopenia | IMME < 7.23 en hombres y <5.67 en mujeres | No se considera | VC < 1.0 |
| Sociedad de Sarcopenia y Caquexia | ≤ 2 desviaciones estándar del promedio de MA/altura evaluada | No se considera | TC6M < 400 |

TC6M: Test de caminata de 6 minutos; VC: velocidad de caminata en $m*s^{-1}$.

Si bien, la medición de parámetros antropométricos involucra un costo inferior a las técnicas computacionales mencionadas, los puntos de corte de los parámetros antropométricos (los valores críticos que definen los umbrales en los que puede considerarse que una medición es anómala) difieren entre estudios debido a las características de las muestras evaluadas [9]. Esta variabilidad limita el uso de estos puntos de corte a muestras en las que las etnias de los participantes coincidan con la de las personas que se usaron para obtenerlos. Además, dado que la sarcopenia es una enfermedad multifactorial, su diagnóstico puede confundirse con síndromes metabólicos, neuromusculares y reumáticos [10]. En México, Espinel en [11] emplearon una base de datos conformada por 5046 registros de adultos mayores en el 2012 e identificaron, sin puntos de corte, que las mujeres son más propensas de adquirir sarcopenia y este riesgo incrementa con la edad, caídas previas, deterioro cognitivo, obesidad abdominal y alta marginación.

En este contexto, la Inteligencia Artificial (IA) ha emergido como una herramienta para el diseño de sistemas que emulan la inteligencia humana que ayude en la toma de decisiones a los especialistas de la salud, así como la creación de modelos predictivos de diagnósticos clínicos [12]. Dentro de las aplicaciones en el área de ingeniería médica y biomédica, los árboles de decisión permiten crear modelos predictivos a partir de registros históricos de pacientes que previamente fueron diagnosticados con padecimientos o enfermedades [13] como apendicitis aguda, cáncer de mama, hepatopatía crónica [14]. Sin embargo, esta técnica es susceptible a la cantidad de variables evaluada y la variabilidad de cada registro.

Esta problemática fue abordada por Breiman en [15], que propuso la creación y combinación de los mejores árboles de decisión que permiten incrementar la preci-

sión del modelo predictivo, esta técnica es conocida como *Random Forest*.

En el área de geriatría, *Random Forest* ha sido aplicado en beneficio de las personas adultas mayores para identificar la sintomatología y puntos de cohorte de personas con sarcopenia a través de biomarcadores moleculares [16], parámetros bioquímicos [17], daño muscular [18].

Si bien la IA es capaz procesar grandes volúmenes de información y desempeña un papel fundamental en la investigación médica, a menudo nos enfrentamos a un dilema ético y práctico. En muchos casos, los participantes pueden sentirse utilizados o preocupados por la invasión de su privacidad al compartir un gran número de datos personales o médicos y, en consecuencia, se niegan a participar en estos estudios [19].

Dada la importancia de ofrecer calidad de vida a las personas adultas mayores, este trabajo reporta una metodología basada en la depuración de una base de datos, la clasificación y la construcción de un modelo predictivo basado en registros históricos, para el diagnóstico primario de sarcopenia empleando mediciones antropométricas, así como su validación respecto a estrategias reportadas en la literatura abierta. A diferencia de otras propuestas, este trabajo evalúa variables ajenas a la medición en personas adultas mayores que carecen de alguna o algunas extremidades o se encuentran postradas. Asimismo, se proponen los puntos de corte en una muestra de la población de personas adultas mayores del estado de Colima, México.

MATERIALES Y MÉTODOS

Este trabajo se realizó en tres etapas: i) la identificación de los participantes y la recolección de variables antropométricas, ii) la depuración de los datos a partir de herramientas estadísticas y iii) la categorización de los participantes en función de las variables antropométricas. Los detalles de cada etapa se mencionan a continuación:

Identificación y recolección de datos

Durante el 2021, como parte de un estudio longitudinal, profesorado del área de ciencias de la salud del estado de Colima capacitó a un grupo de estudiantes de ambos géneros adscritos a la Facultad de Medicina de la Universidad de Colima, quienes acudieron al Centro de Asistencia Social de Día para Personas Adultas Mayores Zona Oriente del DIF Estatal Colima ubicado en la zona metropolitana Colima-Villa de Álvarez, donde invitaron a personas adultas mayores a participar en el estudio. Dicho estudio tuvo como objetivo comprender el envejecimiento patológico en la población adulta mayor, así como el impacto que tuvo el cese de actividades recreativas tras el periodo de contingencia por la COVID-19; dicha propuesta contó con la aprobación del Comité de Ética en Investigación de una institución de salud que cuenta con registro ante la Comisión Nacional de Bioética (CEICANCL08012021-MARCOGP-01). Los datos generados de dicha investigación se encuentran en un repositorio, el cual puede ser compartido a través de un convenio de colaboración con el grupo de investigación de la Universidad de Colima.

Para ofrecer una propuesta que cumpla con los principios éticos y evitar sesgos en los resultados, se realizó un filtro primario de personas adultas mayores que cumplieran características como edad igual o superior a 60 años cumplidos. En cambio, no se tomaron como participantes aquellos que se desplazaban en sillas de ruedas o tenían prótesis en brazos o piernas.

Con base en lo anterior, 150 participantes cumplieron con los requisitos y firmaron un acuerdo de consentimiento informado que indicaba el objetivo del proyecto, las actividades que se comprometían a realizar, así como los riesgos de su participación. Asimismo, para asegurar la ética del proyecto, la propuesta se sometió bajo los principios éticos de la Declaración de Helsinki y ante la Comisión Nacional de Bioética de México [20], en la que los autores se comprometieron a mantener la confidencialidad y privacidad de los datos,

así como la equidad e imparcialidad y la transparencia en la selección de los participantes para evitar sesgos en el análisis y presentación de resultados. Cada persona adulta mayor que cumplió con los requisitos y manifestó su interés de manera voluntaria fue incluida en el estudio.

TABLA 2. Mediciones registradas para la base de datos.

| Variable | Unidad de medida | Equipo de medición |
|--|---------------------|---------------------------------|
| Sexo | variable categórica | Entrevista |
| Edad | años | Entrevista |
| Estatura | m | Estadímetro |
| Peso medido | kg | Balanza de bioimpedancia |
| Grasa corporal total | % | Monitor de composición corporal |
| Masa muscular total (MUS) | kg | Monitor de composición corporal |
| Masa muscular en pierna izquierda y derecha (MPI y MPD) | kg | Monitor de composición corporal |
| Masa muscular en brazo izquierdo y derecho (MBI y MBD) | kg | Monitor de composición corporal |
| Masa muscular en zona central (MC) | kg | Monitor de composición corporal |
| Circunferencia a media de brazo (CMB) | cm | Cinta métrica |
| Circunferencia a media de pantorrilla (CMP) | cm | Cinta métrica |
| Fuerza de agarre en brazo izquierdo y brazo derecho (FIZ y FDER) | kg fuerza | Dinamómetro |
| Velocidad de Marcha (Marcha) | m*s ⁻¹ | Cinta métrica Cronómetro |

A cada persona adulta mayor que cumplió con los requerimientos se le asignó un número de folio para garantizar la confidencialidad de sus datos y como primera actividad, el personal de apoyo solicitó a los participantes que proporcionaron sus datos demográficos, medidas antropométricas y pruebas de función muscular. La Tabla 2 muestra las variables registradas, unidades de medida y equipo empleado en este estudio.

Las variables de la Tabla 2 se registraron en una hoja de cálculo de Excel. Además de estas variables se calcularon también los índices de masa corporal (a partir del peso total y la altura), de masa muscular en la zona central (definido como la masa muscular del tronco dividido entre la altura al cuadrado) y de masa muscular apendicular (a partir de la suma de masa muscular brazos y piernas dividida entre la altura al cuadrado), todos en unidades de $\text{kg}\cdot\text{m}^{-2}$.

Depuración de datos

En esta etapa, se implementó la metodología propuesta por Osborne en [21] para la depuración de datos. Esta metodología destaca la importancia de seguir ciertos pasos para asegurar la calidad de los datos antes de su análisis. El procesamiento de la información se llevó a cabo en *Google Colab*, una plataforma gratuita en línea que permite ejecutar código en la nube.

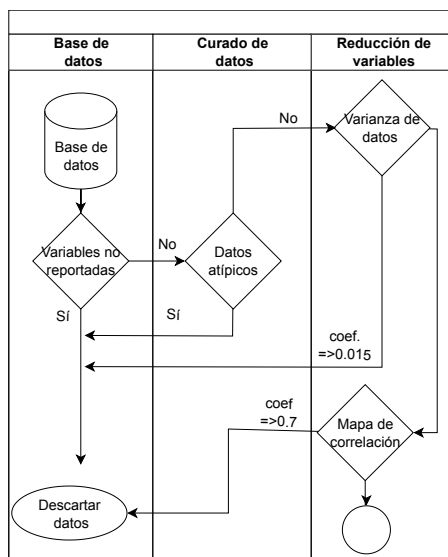


FIGURA 1. Diagrama de flujo de la depuración de datos.

La Figura 1 esclarece el proceso de depuración de datos, que se inició con la identificación y exclusión de participantes cuyos registros carecían de información completa de las variables detalladas en la Tabla 2. Inicialmente, la base de datos original se redujo a 144 participantes una vez que se eliminaron los registros incompletos. Posteriormente, se llevó a cabo la identificación de datos atípicos empleando diagramas de caja y bigotes, herramientas que facilitan la visualización de la distribución de los datos y permiten la detección de valores que se hallan fuera del rango esperado [22]. Aquellos registros que presentaron valores atípicos fueron excluidos, lo que resultó en una base de datos depurada con 122 pacientes (con una edad promedio de 71.8 años y una desviación estándar de 6 años). El personal de salud en atención primaria evaluó los resultados de las pruebas clínicas en un total de 64 pacientes, aunque no fue posible completar el diagnóstico para toda la muestra debido a limitaciones logísticas. La mayoría de los pacientes, que formaban un grupo de 52 individuos de ambos géneros, recibieron un diagnóstico negativo, en contraste con un grupo más reducido de 12 pacientes, en su mayoría mujeres, que obtuvieron un diagnóstico positivo. Para asegurar la integridad de los análisis, se excluyeron a los pacientes con datos faltantes o valores atípicos. Esto resultó en dos grupos finales: 47 pacientes con diagnóstico negativo (compuestos por 33 mujeres y 14 hombres) y 10 pacientes con diagnóstico positivo confirmado (9 mujeres y 1 hombre).

Una vez depurada la base de datos, se procedió a reducir las mediciones registradas para el diagnóstico de sarcopenia, siguiendo la metodología de “reducción de dimensionalidad” [23]. Este proceso, se llevó a cabo en dos fases. En la primera fase se analizó la varianza de cada una de las variables cuantitativas. De acuerdo con la reducción de dimensionalidad, se considera que las variables con las varianzas normalizadas más pequeñas son menos eficientes en la clasificación de pacientes a través de sus diferencias. En el caso de la muestra, se observó que la variable cuantitativa de “Edad” tiene

una varianza normalizada mucho menor al resto, por lo que fue eliminada de la lista original de variables. Este resultado es consistente por el reportado por Lera en ^[24] quienes identificaron mayor prevalencia de sarcopenia en personas adultas mayores a 80 años que radican en Chile.

En la segunda fase, se evaluó la correlación entre las variables de la Tabla 2 utilizando el coeficiente de correlación de Spearman. A partir del mapa de correlación resultante, se eliminaron aquellas variables que tuvieran un índice de correlación mayor 0.7, siguiendo la propuesta de la eficiencia de un proceso de clasificación no se ve comprometido si se eliminan variables que estén altamente correlacionadas con otras ^[25]. El resto de las variables fueron consideradas como la lista de variables con las que se continua en el proceso de clasificación de la base de datos depurada.

Clasificación de datos y modelo predictivo

Una vez concluido el proceso de depuración de datos, se procedió a la clasificación de estos y a la construcción del modelo predictivo. En particular, los modelos predictivos construidos por IA, han reportado alto desempeño en el análisis de imágenes en el diagnóstico de diferentes tipos de cáncer y pronosis ^[26], reducen sesgos y subjetividad al evaluar al paciente ^[22], capacidad de procesamiento de información para la construcción de modelos predictivos cualitativos y cuantitativos basados en registros históricos de pacientes que comparten la misma sintomatología de una enfermedad ^[23], entre otros.

Como actividad previa a la implementación de la IA como herramienta para apoyar al especialista de la salud en la toma de decisiones se recomienda desarrollar la metodología propuesta por Azar y *El-Metwally* en ^[27]. La descripción general se reporta en la Figura 2 y los detalles de cada elemento se mencionan a continuación:

La Figura 2, en su primera columna titulada

“Agrupamiento de Pacientes”, muestra cómo se agruparon a los participantes con características similares. Usamos una técnica llamada clustering jerárquico aglomerativo ^[28]. El clustering jerárquico aglomerativo es utilizado para agrupar objetos o puntos de datos de manera jerárquica. Inicialmente, cada punto se considera un clúster independiente y, mediante el cálculo de la distancia entre este punto y el resto, todos aquellos puntos que estén a una distancia menor se incluyen como parte del mismo clúster en la siguiente iteración. El clustering jerárquico aglomerativo es utilizado para agrupar objetos o puntos de datos de manera jerárquica. Inicialmente, cada punto se considera un clúster independiente y, mediante el cálculo de la distancia entre este punto y el resto, todos aquellos puntos que estén a una distancia menor se incluyen como parte del mismo clúster en la siguiente iteración. Dos parámetros que deben definirse para aplicar este algoritmo son la distancia y el tipo de enlace. En este contexto se optó por usar la “distancia euclidiana” y el “enlace de Ward”. Esta combinación de distancia y métrica es efectiva para formar clústeres densos y cohesivos, lo que facilita la identificación de grupos bien definidos en aplicaciones como el estudio de sarcopenia. Esta técnica se ha empleado en el ámbito médico para diagnosticar enfermedades y comorbilidades en adultos mayores ^[29]. Tanto el algoritmo de clustering jerárquico como el tipo de distancia y enlace para su creación están contenidos dentro de la librería de Python “*Scikit-learn*” ^[30]. Esta librería es una herramienta gratuita y abierta para el aprendizaje automático que incluye funciones para tareas como clasificación, regresión, clustering y reducción de dimensionalidad ^[31].

Para determinar el número ideal de clústeres, se emplearon las técnicas del método del codo y el análisis de inercia, tal como se detallan en el trabajo de Castañeda en ^[32]. Además, se llevó a cabo un análisis de hipótesis utilizando pruebas *t-Student* ^[33] para asegurar que no hubiera clústeres redundantes, es decir, aquellos que no pudieran distinguirse significativamente del resto en función de los valores promedio de los

parámetros utilizados en la clasificación de los pacientes. Además, se utilizaron las métricas de “Silhouette Score”, “Índice Davies-Bouldin” y “Calinski-Harabasz” para evaluar el desempeño de los algoritmos de agrupamiento. Respectivamente, se encontraron los valores 0.23, 1.12 y 58.4. Estos valores sugieren que los grupos están relativamente bien separados, en relación con la dispersión interna de los datos.

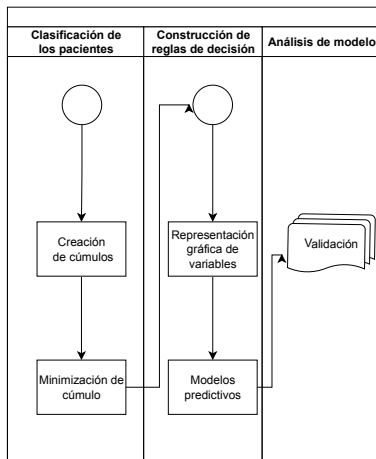


FIGURA 2. Diagrama de flujo de la clasificación de datos.

En la segunda columna del diagrama de flujo de la Figura 2, se muestra el proceso de construcción de las reglas de decisión. Para abordar esta problemática, se utilizó la técnica del *Random Forest*, que permite la creación de múltiples árboles de decisión con el propósito de mejorar la precisión del modelo predictivo. En este estudio, se optó por generar 100 árboles de decisión y seleccionar aquellos con un mejor desempeño para la creación del árbol de decisión final. Este procedimiento conlleva la identificación del árbol individual del *Random Forest* que se considera el óptimo en función de la importancia de las características. Dicha selección se fundamenta en la relevancia de las características del modelo y permite utilizar un árbol específico del conjunto con el fin de comprender con mayor profundidad el proceso de toma de decisiones o llevar a cabo predicciones basadas en ese árbol particular. Este árbol final proporciona un conjunto de reglas, un modelo predictivo, que permiten el diagnóstico de la sarcopenia a partir de la clasificación de los pacientes de acuerdo con los puntos de corte calculados para los

parámetros considerados pertinentes [34].

Para ello, se empleó el método de análisis de componentes principales, que permite reducir las variables a sólo dos dimensiones y visualizar gráficamente las agrupaciones de datos que comparten características semejantes. Este método fue propuesto por Bakator y Radosav en [35], pero tiene la limitación de estar restringido a magnitudes cualitativas difusas, como “alto” o “bajo”, lo que puede tornar subjetivo el diagnóstico de un paciente.

Luego de obtener un modelo de clasificación basado en *Random Forest*, los datos de pacientes con diagnóstico confirmado por personal médico se utilizaron para validar las reglas de decisión generadas por el árbol final. Para evaluar la coherencia tanto de los agrupamientos como de las reglas de decisión obtenidas, se estableció como criterio esencial que un modelo adecuado debía ser capaz de agrupar a los pacientes con diagnósticos confirmados en grupos específicos sin superposición. En otras palabras, las reglas de agrupamiento propuestas debían permitir la formación de un grupo para los diagnósticos positivos y otro para los negativos sin que se produjeran solapamientos entre ellos. Además, se consideró deseable que se pudieran agrupar todos los pacientes con un diagnóstico compartido en un grupo específico.

Dado que cada árbol final involucra la consideración de múltiples variables para representar a los pacientes, se aplicaron técnicas secuenciales de reducción de dimensionalidad con el propósito de facilitar la identificación de posibles patrones. Se optó por la aplicación de dos técnicas disponibles dentro de la librería *Sci-kit-learn* (t-SNE y PCA).

Inicialmente, se implementó t-SNE (*t-Distributed Stochastic Neighbor Embedding*) para disminuir la complejidad del conjunto de datos y mejorar la identificación de patrones. Esta técnica de reducción de dimensionalidad es utilizada principalmente para visualizar

datos complejos de alta dimensión de una manera más sencilla y comprensible. Su propósito principal es ayudar a identificar patrones, estructuras y agrupaciones en datos que son difíciles de apreciar en su forma original. t-SNE asigna a cada punto de datos en un espacio de alta dimensión un punto en un espacio de baja dimensión de tal manera que los puntos similares en el espacio de alta dimensión se representan como puntos cercanos en el espacio de baja dimensión.

A continuación, se llevó a cabo el análisis de componentes principales (PCA) para lograr una reducción adicional de la dimensionalidad [36]. El Análisis de Componentes Principales es una técnica que, en esencia, busca transformar un conjunto de variables originales en un nuevo conjunto de variables llamadas "componentes principales". Cada componente principal es una combinación lineal de las variables originales y están ordenadas en función de su capacidad para explicar la variabilidad en los datos. Esta técnica simplifica los datos al resaltar las direcciones en las que los datos varían más y facilita la identificación de las variables originales que contribuyen significativamente a esa variabilidad.

Finalmente, se procedió a la aplicación de t-SNE una vez más en el espacio de menor dimensión resultante con el propósito de perfeccionar la visualización y destacar con precisión los patrones y agrupamientos de interés, así como de identificar clusters específicos o agrupamientos relevantes en los datos. Este enfoque se centra en preservar la estructura local de los datos, lo que facilita la identificación de clusters y patrones específicos.

La última etapa del proceso se muestra en la tercera columna del diagrama de flujo en la Figura 2. En esta etapa, se realizó el análisis del modelo predictivo utilizando la base de datos depurada. Se identificaron a los pacientes con diagnóstico confirmado de sarcopenia y se compararon sus registros con las reglas de decisión creadas por el modelo predictivo. De esta manera, se

pudo evaluar si el modelo es capaz de describir el diagnóstico clínico realizado por bioimpedancia eléctrica utilizando medidas antropométricas.

A partir de la lista depurada de parámetros de interés, se formularon tres sistemas de diagnóstico, cada uno basado en combinaciones específicas de parámetros para los cuales se determinaron puntos de corte que se utilizaron para generar un diagnóstico de sarcopenia. En el primer sistema, se incluyó la combinación del Índice de Masa Muscular Apendicular, la velocidad de marcha y la fuerza de agarre. Esta combinación se propuso con el propósito de permitir la comparación de puntos de corte y reglas previamente reportados en investigaciones anteriores [36]. El segundo sistema propone una combinación que incluye la masa muscular promedio de ambos brazos, la fuerza de agarre, la velocidad de marcha y el porcentaje de grasa corporal. Del mismo modo, el tercer sistema se basa en la combinación de la masa muscular promedio de ambas extremidades inferiores, la fuerza de agarre, la velocidad de marcha y el porcentaje de grasa corporal. La lógica subyacente en la elección de estas variables radica en la evaluación de la viabilidad de un sistema de diagnóstico que requiera únicamente la medición de las extremidades superiores o inferiores, en lugar de abordar las cuatro extremidades. Además, se incorporó el porcentaje de grasa corporal con el fin de determinar la capacidad de distinguir casos de sarcopenia primaria de aquellos que puedan estar relacionados con la obesidad sarcopénica. Cada sistema de diagnóstico propuesto fue evaluado en términos de la consistencia de los puntos de corte que sugieren y del grado de coincidencia con los diagnósticos hechos por el personal médico de atención primaria. Además, los tres sistemas fueron integrados en una plataforma de libre acceso que permite el análisis y diagnóstico desde un teléfono celular.

RESULTADOS Y DISCUSIÓN

Sistema: ASMI + FA + Marcha

El primer sistema de clasificación se basa en la evalua-

ción de la fuerza de agarre (FA), la velocidad de marcha y el índice de masa muscular apendicular (ASMI), una combinación de parámetros propuesta por Yuki *et al.* en [36] para detectar sarcopenia en personas adultas mayores que radican en la zona metropolitana de Japón. Es importante mencionar que, debido a posibles diferencias étnicas, los puntos de corte sugeridos por dicho estudio no son directamente aplicables a la población mexicana.

Los diagramas de cajas se utilizan para la comparación de los clústeres generados en base a variables relevantes en el proceso de clasificación. Las Figuras 3, 4 y 5 comparan directamente la distribución de valores para cada parámetro de interés en cada una de las agrupaciones obtenidas por el algoritmo de clustering. Las cinturas de las cajas sirven como una guía visual que permite evaluar si existen diferencias significativas entre las medianas de cada clúster con relación al parámetro de interés. Cuando las cinturas están separadas y no se superponen, indican una clara diferencia entre los clústeres.

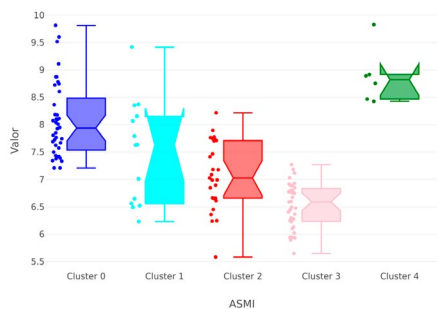


FIGURA 3. Diagrama de cajas para los valores de ASMI en los clústeres del sistema ASMI + FA + Marcha.

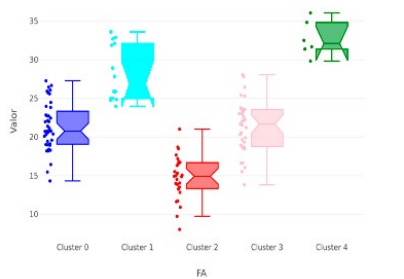


FIGURA 4. Diagrama de cajas para los valores de FA en los clústeres del sistema ASMI + FA + Marcha.



FIGURA 5. Diagrama de cajas para los valores de Marcha en los clústeres del sistema ASMI + FA + Marcha.

De acuerdo con los diagramas de caja, los clústeres 2 y 3 exhiben los valores más bajos en los tres parámetros. El clúster 2 se caracteriza por presentar los valores más bajos en los tres parámetros, lo que lo convierte en un grupo de pacientes con un alto riesgo de sarcopenia. Mientras que, en el clúster 3 también se encuentran pacientes con valores bajos en el índice de masa muscular apendicular (ASMI). Dado que el ASMI depende específicamente de la cantidad de músculo en las extremidades, estos pacientes también son susceptibles a presentar síntomas característicos de la sarcopenia.

En la Figura 6, se muestra el diagrama de árbol generado por el algoritmo *Random Forest*. Este árbol proporciona un conjunto de condiciones que explican la clasificación de pacientes en los clústeres obtenidos mediante el algoritmo de clustering jerárquico aglomerativo. Para simplificar la interpretación del diagrama de árbol, se especificó un número mínimo de 15 muestras por nodo terminal. Esto significa que los modelos de árbol no describen a todos los pacientes de la muestra, pero esta simplificación apenas afecta la precisión de la clasificación [34].

En la Figura 6 se observa que la primera condición se relaciona con la medición de la fuerza promedio de agarre. Si esta medida es menor que 17.9, los participantes de cada clúster siguen un conjunto de pruebas a la izquierda; de lo contrario, se someten a las pruebas indicadas en el lado derecho para este árbol de decisión.

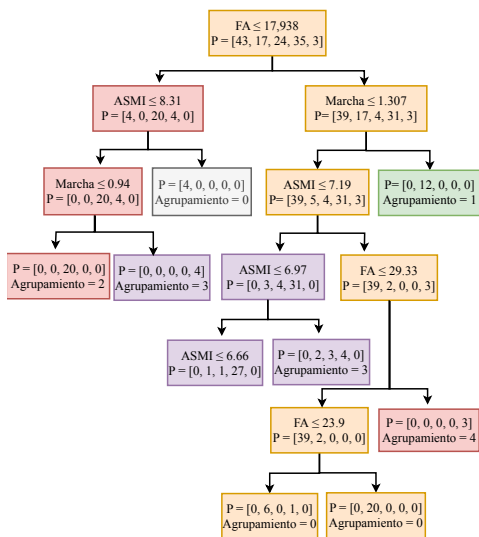


FIGURA 6. Árbol de decisión para clasificar pacientes de acuerdo con el sistema ASMI + FA + Marcha.

Con base a los diagramas de cajas (Figuras 3 a 5), los clústeres 2 y 3 contienen a los pacientes que podrían recibir un diagnóstico de sarcopenia, ya que muestran los valores más bajos para ASMI, FA y marcha. El clúster 1 también presenta valores bajos en la marcha, similares a los del clúster 2, pero se caracteriza por valores significativamente más altos en FA. El clúster 2 consta de 24 pacientes, de los cuales 20 presentaron una FA menor a 17.94, ASMI < 8.31 y marcha < 0.94. Estas condiciones permiten describir a 20 de los 24 pacientes en este clúster, considerándose como la rama principal. Los otros 4 pacientes cumplen con las condiciones FA > 17.94, marcha < 1.3 y ASMI < 7.20. Respecto al tercer clúster, de los 35 participantes, 31 cumplen con las condiciones FA > 17.93, marcha < 1.30 y ASMI < 6.97. Estas secuencias de condiciones representan la base de la clasificación de los pacientes en los clústeres de interés.

La siguiente Figura es la representación simplificada de los clústeres proporcionando una visión visual de la distribución de las personas adultas mayores en función de sus características. La Figura 5, se presenta un diagrama bidimensional generado mediante la aplicación secuencial de técnicas de reducción de dimensionalidad, como t-SNE (*t-distributed Stochastic Neighbor*

Embedding), PCA (*Principal Component Analysis*) y t-SNE nuevamente.

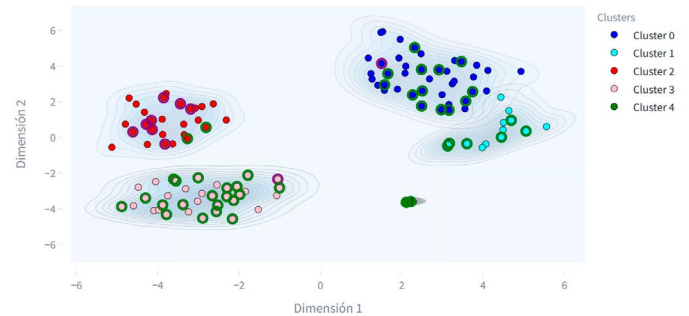


FIGURA 7. Gráfico bidimensional que muestra la densidad de puntos del sistema ASMI + FA + Marcha.

La Figura 7 muestra los registros con diagnóstico confirmado de sarcopenia, estos son representados por puntos encerrados en círculos. De los 11 pacientes con diagnóstico confirmado, 10 se localizan dentro del clúster 2, mientras que 1 se encuentra en el tercer clúster. Esta observación sugiere una alta concordancia entre la clasificación de los pacientes con diagnóstico confirmado y su agrupación en los clústeres identificados.

Además, es importante señalar que los pacientes con diagnóstico confirmado que se encuentran dentro de cada clúster cumplen con las condiciones descritas por las ramas principales correspondientes en el diagrama de árbol previamente mencionado. Esta concordancia entre las condiciones establecidas por el algoritmo de árbol y la distribución de los pacientes con diagnóstico confirmado refuerza la utilidad de las condiciones propuestas para la identificación de la sarcopenia en la población estudiada.

Sistema: MP + FA + Marcha + Grasa

Las Figuras 8 a 11 muestran los diagramas de cajas correspondientes al segundo sistema de clasificación, el cual se basa en la fuerza promedio de agarre (FA), la velocidad de marcha, la masa muscular promedio de

piernas (MP) y el porcentaje corporal de grasa.

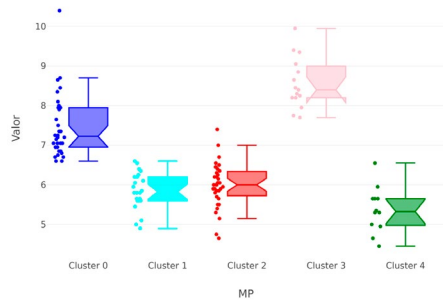


FIGURA 8. Diagrama de cajas para los valores de MP en los clústeres del sistema MP + FA + Marcha + Grasa.

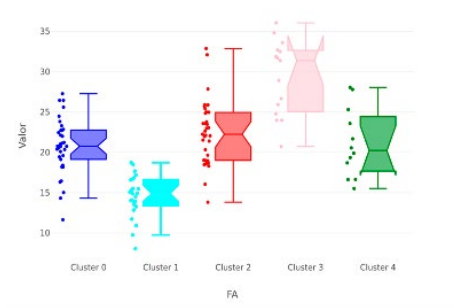


FIGURA 9. Diagrama de cajas para los valores de FA en los clústeres del MP + FA + Marcha + Grasa.

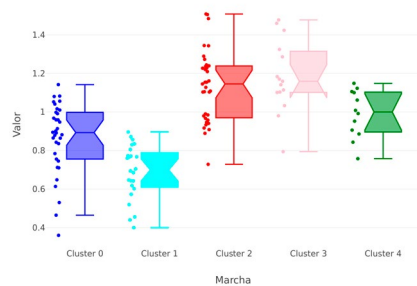


FIGURA 10. Diagrama de cajas para los valores de Marcha en los clústeres del sistema MP + FA + Marcha + Grasa.

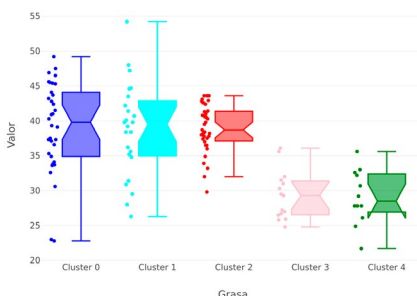


FIGURA 11. Diagrama de cajas para los valores de Grasa en los clústeres del sistema MP + FA + Marcha + Grasa.

De la Figura 8 a la 11, los clústeres de interés son el 0 y el 1. En el caso del clúster 0, se destacan por mostrar valores bajos en velocidad de marcha y fuerza promedio de agarre, así como valores altos en el porcentaje de grasa corporal, lo que sugiere una proporción muscular reducida, siendo esta una medida complementaria al porcentaje de masa muscular corporal. No obstante, ambos clústeres se pueden diferenciar por los valores promedio de masa muscular en las piernas, siendo menores a 7 kg para más de la mitad de los pacientes en el clúster 0 y menores a 5.8 kg para la mayoría de las personas adultas mayores en el primer clúster y, en consecuencia, presentan poca fuerza de agarre, baja velocidad de marcha y escasa masa muscular en las piernas, características que coinciden con los perfiles asociados a la sarcopenia [4][5][6][7][8].

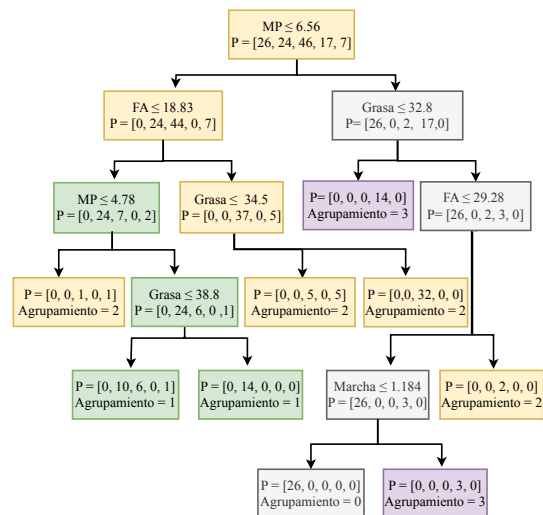


FIGURA 12. Árbol de decisión para clasificar pacientes de acuerdo con el sistema MP + FA + Marcha + Grasa.

En lo que respecta al modelo de *Random Forest* (Figura 12), las reglas que definen la rama principal de pacientes para el clúster 1 abarcan a 24 pacientes. La secuencia de condiciones implica que MP debe ser menor o igual a 6.57, FA menor o igual a 18.8, MP menor o igual a 4.77, y luego se bifurca en la última condición relacionada con el porcentaje corporal de grasa, donde 14 de los 24 pacientes presentan un porcentaje de grasa superior al 38.88 %, mientras que los 10 restantes cumplen la condición complementaria.

En cuanto al clúster 0, un total de 26 pacientes fueron clasificados por el modelo de *Random Forest*. Todos estos pacientes cumplen con las condiciones de MP > 6.57, Grasa > 32.8, FA <= 29.27 y Marcha <= 1.18.

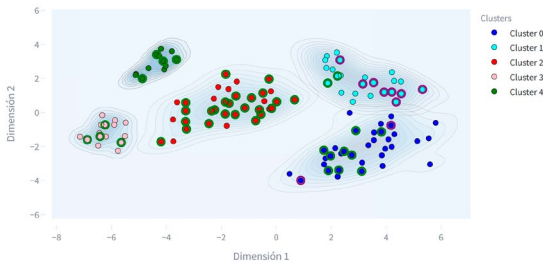


FIGURA 13. Gráfico bidimensional que muestra la densidad de puntos del sistema MP + FA + Marcha + Grasa.

Como se puede observar en la Figura 13, el clúster 1 contiene a 8 de los 10 diagnósticos confirmados, mientras que los otros 2 se clasifican dentro del clúster 0.

Sistema: MB + FA + Marcha + Grasa

Las Figuras 14 a 17 muestran los diagramas de cajas correspondientes al segundo sistema de clasificación, el cual se basa en la masa muscular de brazo (MB), fuerza promedio de agarre (FA), la velocidad de marcha y el porcentaje corporal de grasa.

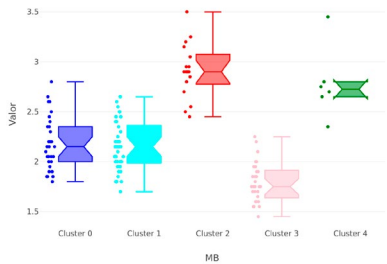


FIGURA 14. Diagrama de cajas para los valores de MB en los clústeres del sistema MB + FA + Marcha + Grasa.

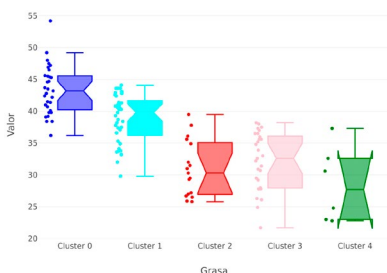


FIGURA 15. Diagrama de cajas para los valores de Grasa en los clústeres del sistema MB + FA + Marcha + Grasa.

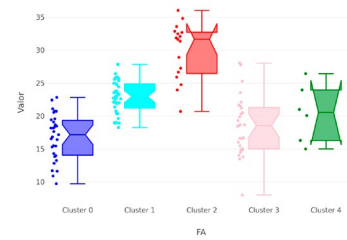


FIGURA 16. Diagrama de cajas para los valores de MB en los clústeres del sistema MB + FA + Marcha + Grasa.

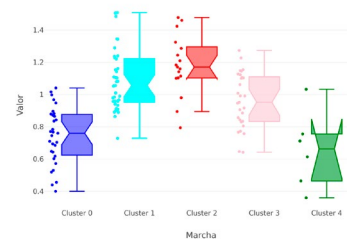


FIGURA 17. Diagrama de cajas para los valores de Marcha en los clústeres del sistema MB + FA + Marcha + Grasa.

Las Figuras 14 a 17 exhiben diagramas de cajas que resaltan las diferencias esenciales entre los clústeres. El clúster 0 se caracteriza por una marcha significativamente más lenta en comparación con los otros grupos. Además, este conjunto de pacientes muestra niveles más elevados de grasa corporal y una fuerza de agarre considerablemente menor en contraste con los otros clústeres, junto con valores intermedios de masa muscular en brazos. Ningún otro cumple con condiciones similares. El clúster 3 se acerca al clúster 0 en términos de los valores más bajos de masa muscular en brazos, fuerza reducida y valores normales de grasa, pero la mayoría de sus pacientes mantienen velocidades promedio de marcha superiores a 1.0 metro por segundo. Estas observaciones enfatizan las características distintivas de cada clúster con respecto a la marcha, la composición corporal y la fuerza muscular. A continuación, se reporta el árbol de decisión para este sistema.

En relación con el modelo de *Random Forest* (Figura 18), las reglas que definen la rama principal de pacientes para el clúster 0 involucran a 31 pacientes. La secuencia de condiciones incluye que la velocidad de

marcha sea menor o igual a 0.88, que la fuerza promedio de agarre esté en el rango de 8.9 a 22.95, que la masa muscular promedio en brazos sea menor o igual a 2.58 kg, y que el porcentaje de grasa corporal sea mayor al 36.86 %. Estas reglas describen las condiciones que cumplen 29 de estos 31 participantes. Como se puede apreciar en la Figura 19, el clúster 0 contiene a 8 de los 10 diagnósticos confirmados, mientras que los otros 2 se clasifican en los clústeres 1 y 4.

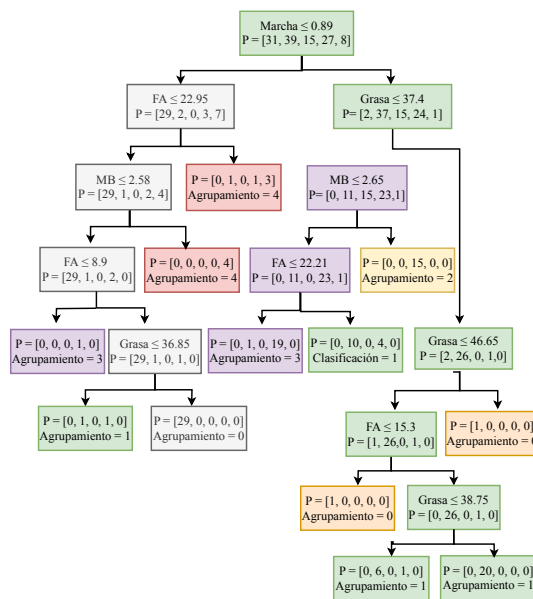


FIGURA 18. Árbol de decisión para clasificar pacientes de acuerdo con el sistema MB + FA + Marcha + Grasa.

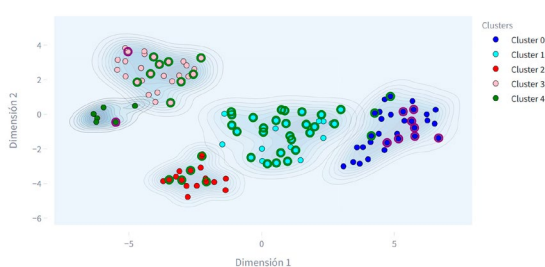


FIGURA 19. Gráfico bidimensional que muestra la densidad de puntos del sistema MB + FA + Marcha + Grasa.

CONCLUSIONES

Este estudio explora el uso de mediciones antropométricas en adultos mayores para desarrollar una aplicación que asista en el diagnóstico de sarcopenia.

Se identificaron reglas de decisión basadas en estas mediciones para la construcción de la metodología propuesta.

Este estudio inició con la depuración de los participantes eliminando registros repetidos, incompletos y datos atípicos. Se realizó la reducción de los parámetros antropométricos eligiendo la mayor varianza, seguida por una alta correlación de Spearman y se validó con la prueba de *t-Student*, obteniendo una lista de cinco parámetros que son consistentes con los reportados en la sintomatología de sarcopenia.

Para establecer los puntos de corte de los parámetros minimizados se propusieron tres sistemas que fueron validados respecto a diagnósticos confirmados de sarcopenia. Los resultados mostraron que, es posible realizar el diagnóstico primario empleando el modelo japonés. El segundo sistema pudo categorizar la sarcopenia clásica y la relacionada con la obesidad, aunque este diagnóstico requiere evaluación adicional con pruebas específicas. Mientras que, el tercer sistema identificó síntomas consistentes con sarcopenia, como marcha lenta, debilidad en la fuerza de agarre, baja masa muscular en los brazos y exceso de grasa corporal.

Un hallazgo interesante fue que los parámetros que no dependen de medir los brazos o las pantorrillas son más simples y menos intrusivos, lo que los hace más accesibles y cómodos para los pacientes, así como más económicos y rápidos de aplicar. Sin embargo, esta simplificación puede llevar a una pérdida de información y menor precisión en la evaluación de la sarcopenia. La elección de utilizar estos sistemas dependerá de los objetivos clínicos y las limitaciones de recursos en un entorno médico específico. Además, en situaciones clínicas reales, los pacientes a veces están postrados, deformaciones óseas, inflamaciones articulares, o bien el personal no dispone del equipo de medición, limita la posibilidad de extraer mediciones antropométricas precisas.

Como trabajo a futuro se plantea emplear esta metodología de trabajo con bases de datos como los generados en la Encuesta Nacional de Salud y Nutrición para determinar la factibilidad teórica y técnica de la actual propuesta.

DECLARACIÓN ÉTICA

El proyecto fue aprobado por el Comité de Bioética del Instituto Estatal de Cancerología, verificado por la Comisión Nacional de Bioética en México bajo el registro: CEICANCL08012021-MARCOGP-01-RA1.

AGRADECIMIENTOS

Los autores agradecen al Consejo Nacional de Humanidades, Ciencias y Tecnologías, por la beca posdoctoral otorgada al CVU 231269, que permitió el desarrollo de este trabajo.

CONTRIBUCIÓN DE CADA AUTOR

S.A.D. conceptualizó la propuesta, desarrolló los algoritmos, analizó los resultados y participó en la redacción del documento. X.A.R.T.T. gestionó la obtención de los registros antropométricos de los participantes ante la Comisión Nacional de Bioética de México, y además colaboró en la revisión y retroalimentación del manuscrito. J.A.B.B. recolectó y curó los datos y contribuyó en la conceptualización de la propuesta. E.E.B.B. conceptualizó y desarrolló la metodología, y participó en la redacción del manuscrito. M.R.S. y M.H.V. contribuyeron en la redacción de la discusión y revisión bibliográfica. J.R.G.F. realizó pruebas de campo e ingresó las mediciones al software para evaluar la usabilidad y compatibilidad con los sistemas operativos Android e iOS.

REFERENCIAS

- [1] B. Ballina Negrete, "La construcción institucional de la educación inclusiva en México," Tesis de pregrado, Univ. Nac. Aut. Mex., Ciudad de México, México, 2023. [En línea]. Disponible en: <https://ru.dgb.unam.mx/bitstream/20.500.14330/TES01000842167/3/0842167.pdf>
- [2] L. F. Faúndez-Pino, "Envejecimiento saludable y enfoque salutogénico," *Rev. Mex. Med. Fam.*, vol. 10, no. 1, pp. 39-45, 2023. [En línea]. Disponible en: https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2696-12962023000100039
- [3] M. M. Hechavarría Ávila, M. Ramírez Romaguera, H. García Hechavarría y A. García Hechavarría, "El envejecimiento. Repercusión social e individual," *Rev. Inf. Cient.*, vol. 97, no. 6, pp. 1173-1188, 2018. [En línea]. Disponible en: http://scielo.sld.cu/scielo.php?pid=S1028-99332018000601173&script=sci_arttext
- [4] M. Casati, A. S. Costa, D. Capitanio, L. Ponzoni, E. Ferri, S. Agostini y E. Lori, "The Biological Foundations of Sarcopenia: Established and Promising Markers," *Front. Med.*, vol. 6, art. no. 184, 2019, doi: <http://dx.doi.org/10.3389/fmed.2019.00184>
- [5] H. Simsek, R. Meseri, S. Sahin, A. Kilavuz, et al., "Prevalence of sarcopenia and related factors in community-dwelling elderly individuals," *Saudi Med. J.*, vol. 40, no. 6, pp. 568-574, 2019, doi: <https://doi.org/10.15537%2Fsmj.2019.6.23917>
- [6] P. B. V. Fernández, A. E. Domínguez, I. B. Vázquez, C. A. Acosta, S. C. Diaz, D. D. A. Navarro, "Sarcopenia and related factors in women over 40 years," *Rev. Cuba. Endoc.*, vol. 31 no. 3, art. no. e198, 2020. [En línea]. Disponible en: <https://www.medigraphic.com/cgi-bin/new/resumenl.cgi?IDARTICULO=100636>
- [7] A. Casirati, G. Vandoni, S. Della Valle, G. Greco, M. Platania, S. Colatruccio, L. Lalli, C. Gavazzi, "Nutritional status and body composition assessment in patients with a new diagnosis of advanced solid tumour: Exploratory comparison of computed tomography and bioelectrical impedance analysis," *Clin. Nutr.*, vol. 40, no. 3, pp. 1268-1273, 2021, doi: <http://dx.doi.org/10.1016/j.clnu.2020.08.009>
- [8] W. A. Sepulveda Loyola, G. A. Luna Corrales, F. Ganz, H. Gonzalez Caro y V. Suziane Probst, "Sarcopenia, definición y diagnóstico: ¿Necesitamos valores de referencia para adultos mayores de Latinoamérica?," *Rev. Chil. Ter. Ocup.*, vol. 20, no. 2, pp. 259-267, 2020, doi: <https://doi.org/10.5354/0719-5346.2020.53583>
- [9] T. M. Nicolalde Cifuentes y S. I. Heredia Aguirre, "Parámetros antropométricos y su relación con la actividad física en una población adulta," *Polo Conoc.*, vol. 6 no. 9, pp. 723-736, 2021. [En línea]. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=8094563>
- [10] K. Keller, "Sarcopenia," *Wien. Med. Wochenschr.*, vol. 169, pp. 157-172, 2019, doi: <https://doi.org/10.1007/s10354-018-0618-2>
- [11] M. C. Espinel-Bermúdez, S. Sánchez-García, C. García-Peña, X. Trujillo, M. Huerta-Viera, V. Granados-García, S. Hernández-González, y E. D. Arias-Merino, "Factores asociados a sarcopenia en adultos mayores mexicanos: Encuesta Nacional de Salud y Nutrición 2012," *Rev. Med. Inst. Mex. Seguro Soc.*, vol. 56, no. 1, pp. 46-53, 2018. [En línea]. Disponible en: <https://www.redalyc.org/articulo.oa?id=457754907008>
- [12] G. S. Collins y K. G. Moons, "Reporting of artificial intelligence prediction models," *The Lancet*, vol. 393, no. 10181, pp. 1577-1579, 2019, doi: [https://doi.org/10.1016/s0140-6736\(19\)30037-6](https://doi.org/10.1016/s0140-6736(19)30037-6)
- [13] M. Banerjee, E. Reynolds, H. B. Andersson y B. K. Nallamothu, "Tree-based analysis: a practical approach to create clinical decision-making tools," *Cir. Cardiovasc. Qual. Outcomes*, vol. 12, no. 5, art. no. e004879, 2019, doi: <https://doi.org/10.1161%2FCIRCOUT-COMES.118.004879>
- [14] S. O. Lugo-Reyes, G. Maldonado-Colín y C. Murata, "Inteligencia artificial para asistir el diagnóstico clínico en medicina," *Rev. Alerg. Mex.*, vol. 61, no. 2, pp. 110-120, 2014, doi: <https://doi.org/10.29262/ram.v61i2.33>
- [15] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5-32, 2001, doi: <https://doi.org/10.1023/A:1010933404324>
- [16] G. T. Reddy, M. P. K. Reddy, K. Lakshmana, R. Kaluri, D. S. Rajput, G. Srivastava y T. Baker, "Analysis of dimensionality reduction techniques on big data," *IEEE Access*, vol. 8, pp. 54776-54788, 2020, doi: <https://doi.org/10.1109/access.2020.2980942>
- [17] R. Krzyżmińska-Siemaszko, E. Deskur-Śmielecka, A. Kaluźniak-Szymanowska, A. Styszyński y K. Wieczorowska-Tobis, "Polish version of SARC-F to assess sarcopenia in older adults: An examination of reliability and validity," *PLoS One*, vol. 15, no. 12, art. no. e0244001, 2020, doi: <http://dx.doi.org/10.1371/journal.pone.0244001>
- [18] S. Huang, J. Yang, S. Fong y Q. Zhao, "Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges," *Cancer Lett.*, vol. 471, pp. 61-71, 2020, doi: <http://dx.doi.org/10.1016/j.canlet.2019.12.007>
- [19] F. Morandín-Ahuerma, "Diez recomendaciones de la Unesco sobre ética de la inteligencia artificial," En: *Principios normativos para una ética de la inteligencia artificial*. Puebla, México: Consejo de Ciencia y Tecnología del Estado de Puebla (CONCYTEP). pp. 86-94.
- [20] Cámara de Diputados del H. Congreso de la Unión. (7 de febrero de 1984). *Ley General de Salud*. [En línea]. Disponible en: <https://www.diputados.gob.mx/LeyesBiblio/pdf/LGS.pdf>
- [21] J. W. Osborne, *Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data*, Thousand Oaks, CA: Sage Publications Inc., 2013, doi: <https://doi.org/10.4135/9781452269948>
- [22] J. F. Ávila-Tomás, M. A. Mayer-Pujadas y V. J. Quesada-Varela, "La inteligencia artificial y sus aplicaciones en medicina II: importancia actual y aplicaciones prácticas," *Aten. Primaria*, vol. 53, no. 1, pp. 81-88, 2020, doi: <https://doi.org/10.1016/j.aprim.2020.04.014>
- [23] S. H. Park y K. Han, "Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction," *Radiology*, vol. 286, no. 3, pp. 800-809, 2018, doi: <http://dx.doi.org/10.1148/radiol.2017171920>
- [24] L. Lera, C. Albala, H. Sánchez, B. Angel, M. J. Hormazabal, C. Márquez y P. Arroyo, "Prevalence of Sarcopenia in Community-Dwelling Chilean Elders According to an Adapted Version of the European Working Group on Sarcopenia in Older People (EWGSP) Criteria," *J. Frailty Aging.*, vol. 6, no. 1, pp. 12-17, 2017, doi: <https://doi.org/10.14283/jfa.2016.117>
- [25] J. Moncada Jiménez, *Estadística para ciencias del movimiento humano*. Costa Rica: Editorial de la Universidad de Costa Rica, 2005.
- [26] S. Huang, J. Yang, S. Fong y Q. Zhao, "Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges," *Cancer Lett.*, vol. 471, pp. 61-71, 2020, doi: <http://dx.doi.org/10.1016/j.canlet.2019.12.007>

- [27] A. T. Azar y S. M. El-Metwally, "Decision tree classifiers for automated medical diagnosis," *Neural Comput. Applic.*, vol. 23, pp. 2387-2403, 2013, doi: <http://dx.doi.org/10.1007/s00521-012-1196-7>
- [28] D. Blázquez García, "Análisis de datos de expresión en retina para la identificación de biomarcadores de DMAE: posible herramienta diagnóstica," Tesis de pregrado, Fac. Med., Univ. Vall., Valladolid, España, 2023.
- [29] F. T. T. Lai, P. E. Beeler, B. H. K. Yip, M. Cheetham, et al., "Comparing multimorbidity patterns among discharged middle-aged and older inpatients between Hong Kong and Zurich: A hierarchical agglomerative clustering analysis of routine hospital records," *Front. Med.*, vol. 8, art. no. 651925, 2021. <https://doi.org/10.3389/fmed.2021.651925>
- [30] P. Lemenkova, "R Libraries {dendextend} and {magrittr} and Clustering Package scipy.cluster of Python For Modelling Diagrams of Dendrogram Trees," *Carpathian J. Electr. Comp. Eng.*, vol. 13, no. 1, pp. 5-12, 2020, doi: <https://doi.org/10.2478/cjece-2020-0002>
- [31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825-2830, 2011. [En línea]. Disponible en: <http://jmlr.org/papers/v12/pedregosa11a.html>
- [32] A. K. Castañeda Escobar, "Diseño e implementación de un algoritmo para determinar la mejor agrupación utilizando k-means y los índices de validación," Tesis de maestría, Univ. Aut. Edo. Mor., Morelos, México, 2023. [En línea]. Disponible en: <http://riaa.uaem.mx/handle/20.500.12055/3115>
- [33] E. Goffinet, M. Lebbah, H. Azzag, L. Giraldi y A. Coutant, "Multivariate time series multi-coclustering: Application to advanced driving assistance system validation," in *Europ. Symp. Artif. Neural Netw. Comput. Intell. Mach. Learn.*, Oct. 2021, pp. 463-468, doi: <https://doi.org/10.14428/esann/2021.es2021-119>
- [34] S. J. Rigatti, "Random forest," *J. Insur. Med.*, vol. 47, no. 1, pp. 31-39, 2017, doi: <http://dx.doi.org/10.17849/insm-47-01-31-39.1>
- [35] M. Bakator y D. Radosav, "Deep Learning and Medical Diagnosis: A Review of Literature," *Multimodal Technol. Interact.*, vol. 2, no. 3, art. no. 47, 2018, doi: <https://doi.org/10.3390/mti2030047>
- [36] A. Yuki, F. Ando, R. Otsuka, Y. Matsui, A. Harada, H. Shimokata, "Epidemiology of sarcopenia in elderly Japanese," *J. Phys. Fitness Sports Med.*, vol. 4, no. 1, pp. 111-115, 2015, doi: <http://dx.doi.org/10.7600/jpfsm.4.111>