

Clasificación de la densidad mineral ósea utilizando técnicas de aprendizaje automático en niños y adolescentes según edad y sexo

José Sulla-Torres, Dr¹, Alan Bedoya-Carrillo, Eng¹, Rossana Gomez-Campos, Dr², and Marco Cossio-Bolaños, Dr²

¹Universidad Católica de Santa María, Perú, jsullato@ucsm.edu.pe, abedoya@ucsm.edu.pe

²Universidad Católica de Maule, Chile, rossanagomez_c@hotmail.com, mcossio30@hotmail.com

Abstract— Bone health is a field that has become very important in recent years, especially in diseases related to bones, since they are becoming more common among humans. Osteoporosis currently causes an estimated 8.9 million fractures annually. Bone mineral density (BMD) and bone mineral content (BMC) are indicators that can diagnose the problem of bone health. The objective of this study is to classify BMD in children and adolescents using automatic learning techniques. A descriptive cross-sectional study was developed. We studied 660 schoolchildren from two educational centers with an age range of 6 to 18 years from the province of Arequipa (Peru). Anthropometric variables were evaluated. The BMD and CMO were determined. The Body Mass Index (BMI) was calculated, and a comparative study was made of 9 machine learning algorithms related to the subject. These include decision trees, bayesian networks, decision and regression tables. Random Forest's classification algorithm is 94.87%. This algorithm allowed to implement a software. This tool allows to calculate the bone health of schoolchildren between 6 to 18 years. The algorithm obtained can be implemented from a prediction software that allows the classification and prevention of the deterioration of the bone health of children and adolescents.

Keywords—Machine learning, decision tree, bone mineral density, children and adolescents, classification.

Resumen— La salud ósea es un campo que ha tomado mucha importancia en los últimos años, en especial en enfermedades relacionadas a los huesos, puesto que cada vez son más comunes entre los seres humanos. La osteoporosis actualmente causa un estimado de 8.9 millones de fracturas anualmente. La densidad mineral ósea (DMO) y contenido mineral óseo (CMO) son indicadores que permiten diagnosticar el problema de la salud ósea. El objetivo de este estudio es clasificar la DMO en niños y adolescentes utilizando técnicas de aprendizaje automático. Se desarrolló un estudio descriptivo de corte transversal. Se estudió a 660 escolares de dos centros educativos con un rango de edad de 6 a 18 años de la provincia de Arequipa (Perú). Se evaluó variables antropométricas. Se determinó la DMO y CMO. Se calculó el Índice de Masa Corporal (IMC), y se hizo un estudio comparativo de 9 algoritmos de aprendizaje automático relacionados al tema. Estos incluyen árboles de decisión, redes bayesianas, tablas de decisión y de regresión. El algoritmo que clasifica mejor la de DMO es el de Random Forest con un porcentaje de precisión de 94.87%. Este algoritmo permitió implementar un software. Esta herramienta permite calcular la salud ósea de los escolares entre 6 a 18 años. El algoritmo obtenido puede ser implementado de un software de predicción que permita la clasificación y prevención del deterioro de la salud ósea de niños y adolescentes.

Palabras clave—Aprendizaje automático, árbol de decisión, densidad mineral ósea, niños y adolescentes, clasificación.

Digital Object Identifier: (to be inserted by LACCEI).

ISSN, ISBN: (to be inserted by LACCEI).

I. INTRODUCCIÓN

La inteligencia artificial (IA) es una disciplina que se utiliza hoy en día en un amplio rango de aplicaciones. Un campo donde se ha venido desarrollando con una alta tendencia es el de la salud [1].

El aprendizaje automático es un área de investigación de inteligencia artificial que utiliza métodos estadísticos para la clasificación de datos. Se han aplicado varias técnicas de aprendizaje automático en entornos clínicos para predecir enfermedades y han demostrado una mayor precisión para el diagnóstico que los métodos clásicos [2]. Estos algoritmos matemáticos tienen la capacidad de clasificar grandes cantidades de datos en un formato útil [3].

La DMO se refiere a la cantidad de minerales que contiene cierto volumen de hueso. Esta cantidad de minerales depende de diversos estímulos, principalmente a la gravedad y las contracciones musculares [4]. La fase donde la DMO aumenta en los hombres es en la de crecimiento, llegando alcanzar, al final de la adolescencia, cerca del 95% [5].

La osteoporosis es una enfermedad que se desarrolla sin mostrar síntomas en sus primeras etapas. Esta enfermedad se caracteriza por una baja masa ósea y un deterioro microarquitectónico del tejido óseo. Desde un punto de vista económico, la osteoporosis tiene un alto costo para la sociedad, no solo para los gobiernos sino también para las familias afectadas [6]. En 1994, la OMS estableció cuatro categorías o criterios diagnósticos de la osteoporosis sobre la base de criterios epidemiológicos: Normal, Osteopenia, Osteoporosis, Osteoporosis grave o establecida, cuando al criterio de osteoporosis se añade la presencia de fracturas.

Hasta la fecha, hay varios estudios efectuados en poblaciones pediátricas, donde han confirmado que las variables antropométricas [7] y la maduración biológica son fuertes predictores de la salud ósea [8]. Además, las correlaciones que se encontraron entre la DMO y el CMO con las variables antropométricas, es porque el CMO depende de la longitud y del ancho del hueso y la densidad ósea, respectivamente [9]. Por lo tanto, la presencia de huesos cortos y estrechos podría dar lugar a una reducida DMO y CMO y posiblemente diversas consecuencias para la salud en general, sin embargo, el aumento de masa ósea máxima durante la infancia y la adolescencia se puede lograr por medio de constantes cambios de estilos de vida, los que deben ser instituidos a temprana edad [10].

Actualmente, se conocen algunas fórmulas que nos ayudan al cálculo de DMO [11] y el CMO [12], sin embargo, la precisión y exactitud son indicadores que se vienen estudiando con el afán de mejorar las predicciones.

Las técnicas de clasificadores toman los datos de cada paciente y predicen la presencia de enfermedades basadas en patrones subyacentes. Las máquinas de vectores de soporte (MVS) [13], los *random forest* (RF) [14] y las redes neuronales artificiales (RNA) [15] han sido enfoques ampliamente utilizados en el aprendizaje automático [2].

Algunos estudios han demostrado que los algoritmos MVS, RF y RNA podrían ayudar a predecir una DMO baja con datos de hábitos de dieta y estilo de vida [16], los árboles de decisión también ayudan a determinar la baja DMO [17], en general, las técnicas de clasificación del aprendizaje automático que permitan predecir las fracturas en los huesos [18], en personas en diferentes periodos de vida [19].

En ese contexto, la investigación en relación con el diagnóstico y clasificación de la DMO es de suma importancia a edades tempranas, puesto que podría ayudar a prevenir un deterioro de la salud ósea durante el crecimiento y desarrollo.

Por lo tanto, los objetivos de este estudio son clasificar la densidad mineral ósea de niños y adolescentes para predecir la salud ósea, basado en indicadores antropométricos y diseñar un software utilizando algoritmos que presentan la mejor precisión de una comparación de técnicas de minería de datos.

II. MATERIALES Y MÉTODOS

A. Sujetos

Se efectuó un estudio descriptivo de corte transversal. Los niños y adolescentes para esta investigación provienen de 2 centros escolares de la red pública de la provincia de Arequipa. La muestra está conformada por 660 escolares con un rango de edad de 6 a 18 años. Posteriormente, se realizó un preprocesamiento, donde se excluyeron a los escolares que fuman, a los que tuvieron una o más fracturas, a los que reciban suplementos vitamínicos, entre otros.

Se obtuvo el consentimiento de los padres y/o tutor de cada participante, además se consideró el asentimiento de cada participante. El protocolo experimental se basó de acuerdo con la declaración de Helsinki (Asociación Médica Mundial para seres Humanos). Además, el estudio contó con los respectivos permisos de las direcciones de los centros escolares y el comité de ética local.

B. Procedimientos

El proceso para la extracción de conocimiento utilizada es la del KDD (*Knowledge Discovery in Data*) [20]. Se han evaluado variables antropométricas que incluyen: peso (kg), altura de pie (cm), altura sentado (cm), longitud del antebrazo y el diámetro del fémur (cm). La DMO y el CMO fueron determinados a partir de fórmulas de regresión. Se calculó el índice de masa corporal (IMC) con su respectiva fórmula ($IMC = \text{peso} / \text{estatura}^2$). Se calculó la maduración somática por

medio de la ecuación propuesta por Mirwald et al (2002) [21]. Esta ecuación busca determinar los Años de pico de velocidad de Crecimiento (APVC).

Las 4 ecuaciones utilizadas tienen un poder explicativo del 76 al 84% [12], las fórmulas se muestran en las Tablas I y II.

TABLA I
ECUACIÓN DE REGRESIÓN PARA ESTIMAR LA DENSIDAD MINERAL ÓSEA Y CONTENIDO MINERAL ÓSEA BASADA EN LA MADURACIÓN BIOLÓGICA E INDICADORES ANTROPOMÉTRICOS EN HOMBRES

Ecuaciones	FIV	R	R ²	EEE	p
Hombres					
DMO=0.605+0.056*APHV+0.008*LongAntebrazo+0.022*DiamFemur					
APVC	4.034	0.89	0.79	0.10	0.00
LongAntebrazo	4.099				
DiametroFemur	1.867				
CMO=0.43+0.018*APHV+0.039*LongAntebrazo+0.06*DiamFemur					
APVC	4.034	0.91	0.84	0.29	0.00
LongAntebrazo	4.099				
DiametroFemur	1.867				

Leyenda: APVC=Años de pico de velocidad de crecimiento, FIV= factor de inflación de la varianza, EEE=Error estandar estimado [12].

TABLA II
ECUACIÓN DE REGRESIÓN PARA ESTIMAR LA DENSIDAD MINERAL ÓSEA Y CONTENIDO MINERAL ÓSEA BASADA EN LA MADURACIÓN BIOLÓGICA E INDICADORES ANTROPOMÉTRICOS EN MUJERES

Ecuaciones	FIV	R	R ²	EEE	p
Mujeres					
DMO=0.469+0.027*APHV+0.007*LongAntebrazo+0.019*DiamFemur					
APVC	3.150	0.87	0.76	0.08	0.00
LongAntebrazo	2.963				
DiametroFemur	1.781				
CMO=0.077+0.07*APHV+0.032*LongAntebrazo+0.48*DiamFemur					
APVC	3.150	0.91	0.83	0.19	0.00
LongAntebrazo	2.963				
DiametroFemur	1.781				

Leyenda: APVC= Años de pico de velocidad de crecimiento, FIV= factor de inflación de la varianza, EEE=Error estandar estimado [12].

1) Preprocesamiento de Datos.

El preprocesamiento de los datos incluye una serie de pasos que se muestra en la Fig. 1, que se puede utilizar para manejar nuestros datos y prepararlos para su modelamiento e implementación. Para realizar esta etapa tenemos los siguientes pasos en el modelo del proceso de extracción de información.

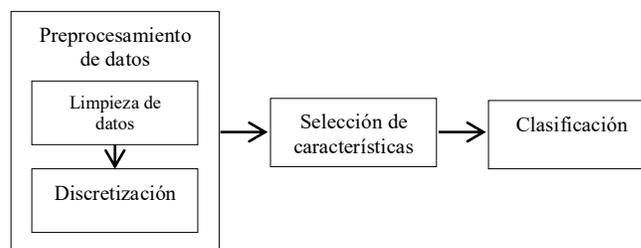


Fig. 1 Preprocesamiento de Datos.

La limpieza de datos es un proceso que implica remover la información incompleta o inconsistente que se tienen. Se comprobaron los campos sobre el rango de edades, así como las mediciones respecto a la estatura, la estatura sentado, la longitud del antebrazo entre otros. En este caso se encontraron diferentes medidas que estaban incorrectamente ingresadas, se procedieron a descartar algunas y arreglar otras utilizando las funciones de limpieza de la herramienta de Weka [22].

La discretización es un método que convierte una variable continua en una variable discreta. Se prefiere discretizar los datos en este estudio para mejorar los algoritmos utilizados. Según lo visto, para tener un mejor desempeño en la clasificación, los variables continuas en el conjunto de datos obtenido fue cambiado en valores nominales, usando el método de discretización en la herramienta Weka.

El proceso de selección de características permite reducir el número de variables con el fin de realizar una correcta clasificación del conjunto de datos. Este proceso permite incrementar la precisión y el desempeño de la clasificación. La herramienta Weka permite realizar la selección de características usando un método de búsqueda y evaluador específico de atributos, basado en la relevancia para la clase. El análisis de los datos, según la cantidad de variables seleccionadas para clasificar la salud ósea se muestra en la Tabla III.

TABLA III
TABLA DE LOS DATOS RECOPIADOS DE NIÑOS Y ADOLESCENTES DE LOS COLEGIOS DEL SECTOR DE AREQUIPA SEGÚN GÉNERO.

Variables	Hombres		Mujeres		Total	
	X	DE	X	DE	X	DE
n	294		366		660	
E.C. (años)	15	1.432	14.9	1.3465	14.9455	1.3937
Edad Biológ (APVC)	15.1	0.701	17	0.8553	16.1705	1.235
Peso (kg)	58.6	13.11	54.2	9.5793	56.1685	11.478
Estatura parado (cm)	164	7.884	156	6.9959	159.652	8.4302
Estatura Sentado (cm)	84.8	5.481	83	3.5072	83.7867	4.6487
Antebrazo (cm)	25.1	1.733	24.2	1.481	24.6039	1.6643
Diám de Rodilla (cm)	8.95	1.064	8.4	0.9458	8.64742	1.0361
Salud Ósea						
DMO	2.68	0.08	1.94	0.057	2.27162	0.3711
CMO	6.92	0.239	23.4	1.2904	16.0535	8.27

Leyenda: X= Promedio, DE=Desviación Estándar, EC= Edad cronológica, APVC= Años pico de velocidad de crecimiento, DMO = Densidad Mineral Ósea, CMO = Contenido Mineral Óseo.

Para la etapa de clasificación, se han incluido los siguientes algoritmos populares que se consideran para la clasificación de la salud ósea de niños y

adolescentes: Bosque Aleatorio (*Random Forest*), Bosque PA, Clasificación Bayesiana Multivariable Eficiente (EBMC), Algoritmo de inducción de reglas no ordenadas difusas (FURIA) y Árbol de redes bayesianas (NB *Tree*).

Para la carga del conjunto de datos se convirtió el formato de archivo de los registros a un archivo .csv. CSV es uno de los formatos permitidos por la herramienta Weka, que facilita la detección de campos y atributos de la información que se desea analizar. Es necesario realizar una adecuada preparación de los datos previa al proceso de minería, con el objetivo obtener conclusiones válidas y útiles al aplicar minería de datos.

2) Transformación

Se hizo un análisis de la calidad de los datos recolectados, donde se pudo reconocer cada uno de los atributos, el número de valores nulos, el valor máximo, media, moda, entre otros, para determinar que filtros de la herramienta Weka deben ser utilizados y aplicados.

Una vez realizado el análisis se pasó a la eliminación de los datos nulos y/o datos vacíos de entre todos los registros. Se descartaron los atributos que no aportaban valor en la clasificación de la información. Luego de la transformación de datos realizada, se obtiene que los atributos que aportan mayor valor en la predicción de la salud ósea.

3) Minería de datos

Para realizar la comparación y elección efectiva de los algoritmos de minería de datos para realizar la clasificación de salud ósea, se utilizó la herramienta Weka, principalmente porque cuenta con una colección amplia de algoritmos de aprendizaje automático para tareas de minería de datos, muestra resultados muy claros, pero a su vez, detalles de la clasificación que sirve para realizar el análisis del estudio. Por otro lado, es un software de libre acceso para los usuarios.

Los algoritmos que se van a evaluar son los siguientes: *Forest PA* [23], *J48* [24], *Random Forest* [14], *EBMC* [25], *TAN*, *NB Tree* [25], *DTNB* [26], *FURIA* [27] y *Classification Via Regression*, que son algoritmos populares en el campo de la clasificación de datos de salud.

III. RESULTADOS

En esta sección se presentan los resultados obtenidos del proceso de extracción del conocimiento, es decir, en base a todos los pasos descritos en secciones anteriores, se determinó qué algoritmo es el mejor para clasificar el estado de la salud ósea de escolares de ambos sexos. Utilizaremos cuadros

resultantes los cuales son: la validación cruzada, la matriz de confusión y la precisión por clase.

En las Tablas IV, V y VI se muestra la validación cruzada de los diferentes algoritmos para determinar cuál podría ser el mejor algoritmo en base a lo más importante que son las instancias clasificadas correctamente según el entrenamiento

aplicado. En la herramienta Weka se ha seleccionado cada uno de estos algoritmos con un porcentaje del 75.6% de datos para entrenar y el resto de los datos para clasificar, según este análisis se han obtenido los siguientes resultados representados en tablas.

TABLE IV
VALIDACIÓN CRUZADA EN LA CLASIFICACIÓN ENTRE LOS ALGORITMOS DE APRENDIZAJE AUTOMÁTICO FOREST PA, J48 Y RANDOM FOREST.

Detalle de resultados	Hombres			Mujeres			Todos		
	Forest PA	J48	Random Forest	Forest PA	J48	Random Forest	Forest PA	J48	Random Forest
Validación cruzada									
Instancias clasificadas correctamente	181	178	182	111	115	112	291	294	296
Instancias clasificadas incorrectamente	6	9	5	13	9	12	21	18	16
% clasificado correctamente	96.79%	95.19%	97.33%	89.52%	92.74%	90.32%	93.27%	94.23%	94.87%
% clasificado incorrectamente	3.21%	4.81%	2.67%	10.48%	7.26%	9.68%	6.73%	5.77%	5.13%
Estadística Kappa	0.2378	0.4477	0.4323	0.8191	0.8728	0.8317	0.8056	0.8476	0.8561
Error medio absoluto	0.0465	0.036	0.0369	0.1589	0.0728	0.1227	0.087	0.0547	0.0731
Error cuadrático medio	0.1371	0.1662	0.119	0.2452	0.2126	0.2275	0.1796	0.1941	0.01635
Error absoluto relativo	16.12%	4.90%	21.13%	24.42%	18.46%	31.11%	34.39%	21.63%	28.876%
Error cuadrático relativo	29.46%	26%	32.48%	38.63%	48.24%	51.64%	51%	55.12%	46.438
Total Número de instancias	187	187	187	124	124	124	312	312	312

TABLE V
VALIDACIÓN CRUZADA EN LA CLASIFICACIÓN ENTRE LOS ALGORITMOS DE APRENDIZAJE AUTOMÁTICO EBMC, TAN Y NB TREE.

Detalle de resultados	Hombres			Mujeres			Todos		
	EBMC	TAN	NB Tree	EBMC	TAN	NB Tree	EBMC	TAN	NB Tree
Validación cruzada									
Instancias clasificadas correctamente	181	181	181	63	111	104	241	279	293
Instancias clasificadas incorrectamente	6	6	6	61	13	20	71	33	19
% clasificado correctamente	96.79%	96.79%	96.790%	50.81%	89.52%	83.87%	77.24%	89.42%	93.91%
% clasificado incorrectamente	3.21%	3.21%	3.21%	49.19%	10.48%	16.13%	22.76%	10.58%	6.09%
Estadística Kappa	0.015	0.125	0.2378	0	0.8176	0.7195	0.7254	0.7156	0.8365
Error medio absoluto	0.0662	0.0585	0.0501	0.3941	0.1016	0.1326	0.2525	0.1003	0.0711
Error cuadrático medio	0.1467	0.1525	0.1407	0.4406	0.2301	0.2912	0.3521	0.2293	0.1916
Error absoluto relativo	4.34%	8.21%	34.75%	49.95%	25.76%	33.63%	4.47%	39.64%	28.11%
Error cuadrático relativo	24.83%	33%	72.53%	89.86%	52.21%	66.09%	26.74%	65.13%	54.41%
Total Número de instancias	187	187	187	124	124	124	312	312	312

VALIDACIÓN CRUZADA EN LA CLASIFICACIÓN ENTRE LOS ALGORITMOS DE APRENDIZAJE AUTOMÁTICO DTNB, FURIA Y CVR.

Detalle de resultados	Hombres			Mujeres			Todos		
	DTNB	FURIA	CVR	DTNB	FURIA	CVR	DTNB	FURIA	CVR
Validación cruzada									
Instancias clasificadas correctamente	181	181	181	109	111	110	289	290	295
Instancias clasificadas incorrectamente	6	6	6	15	13	14	23	22	17
% clasificado correctamente	96.79%	96.79%	96.79%	87.90%	89.51%	88.71%	92.63%	93.98%	94.55%
% clasificado incorrectamente	3.21%	3.21%	3.21%	12.10%	10.49%	11.29%	7.37%	6.02%	5.45%
Estadística Kappa	0.05	0.4398	0.2791	0.7901	0.8175	0.8066	0.7984	0.8444	0.8499
Error medio absoluto	0.0646	0.0256	0.0371	0.1047	0.0805	0.108	0.1081	0.036	0.0777
Error cuadrático medio	0.1454	0.1498	0.1205	0.2346	0.2489	0.2218	0.2197	0.1573	0.1594
Error absoluto relativo	24.37%	3.08%	26.93%	26.56%	20.42%	27.39	42.72%	14.24%	30.73%
Error cuadrático relativo	33.26%	19%	36.07%	53.24%	56.48%	50.33%	62.39%	44.68%	45.27%
Total Número de instancias	187	187	187	124	124	124	312	312	312

Según todas las Tablas vistas se tomaron en cuenta los siguientes resultados: El algoritmo con más instancias clasificadas correctamente para el conjunto de datos es el *Random Forest* con un 94.87% de instancias clasificadas. El algoritmo que menos depende del azar es el *Random Forest* (0.8561), seguido del CVR y FURIA, según su estadística Kappa de todos los algoritmos. No hay mucha relevancia entre la clasificación entre géneros específicos (hombres y mujeres individualmente). En cuanto a errores medios y absolutos se observó que el algoritmo FURIA tuvo el menor índice de error con una pequeña diferencia frente a otros como el *Random Forest*. Todos los algoritmos mostraron un gran porcentaje de instancias clasificadas correctamente.

En función a estos resultados, en la Tabla VII se muestra la matriz de confusión para el algoritmo *Random Forest* que mostraron los mejores resultados analizados que nos describen la clasificación de los datos:

TABLA VII

INSTANCIAS CLASIFICADAS ENTRE LOS DIFERENTES PERCENTILES QUE DETERMINAN LA SALUD ÓSEA CON EL ALGORITMO RANDOM FOREST.

Algoritmo Random Forest	Hombres			Mujeres			Todos		
	<i>f_i</i>	%	<i>T</i>	<i>f_i</i>	%	<i>T</i>	<i>f_i</i>	%	<i>T</i>
Normal	180	99%	0.05	58	52%	0.12	240	81%	0.21
Osteopenia	2	1%		44	39%		40	14%	
Osteoporosis	0	0%		10	9%		16	5%	
Total	182	100%		112	100%		296	100%	

Leyenda *f_i*: frecuencia, %: porcentaje, *T*: Tiempo de ejecución.

En la evaluación de la matriz de confusión se observó principalmente dos cosas: la clasificación que se le asignó a cada instancia y el tiempo de ejecución de cada algoritmo

trabajando con diferentes números de datos. Tomando en cuenta estos dos puntos se tienen los siguientes resultados de importancia: Se han tomado en cuenta solamente el número de instancias que fueron clasificadas correctamente para todos los algoritmos. El algoritmo *Random Forest* clasifican el conjunto de instancias en un 81%, variando entre un porcentaje mínimo con los otros algoritmos.

La Tabla VIII muestra la capacidad de precisión por clase (Normal, Osteopenia, Osteoporosis) del algoritmo *Random Forest*.

Tomando en cuenta que el algoritmo *Random Forest* es el mejor algoritmo para la clasificación de salud ósea en escolares, se utilizará la siguiente representación de árbol de decisión en la herramienta para el modelado.

La Fig 2. muestra el árbol de decisión para la clasificación de la salud ósea.

TABLA VIII

CAPACIDAD DE PRECISIÓN POR CLASE DEL RANDOM FOREST.

Random Forest	Normal	Osteopenia	Osteoporosis
TP Tasa	0.996	0.755	0.889
FP Tasa	0.127	0.012	0.014
Precisión por Clase	0.964	0.930	0.800
Sensibilidad	0.996	0.755	0.889
Valor-F	0.980	0.833	0.842
MCC	0.908	0.810	0.833
ROC Área	0.993	0.981	0.993
PRC Área	0.997	0.910	0.934

Leyenda TP: Verdadero Positivo, FP: Falso Positivo, MCC: Coeficiente de Correlación de Matthew (phi), ROC Area: Curva de la Característica Operativa del Receptor, PRC Área: Curva de Precisión y Sensibilidad.

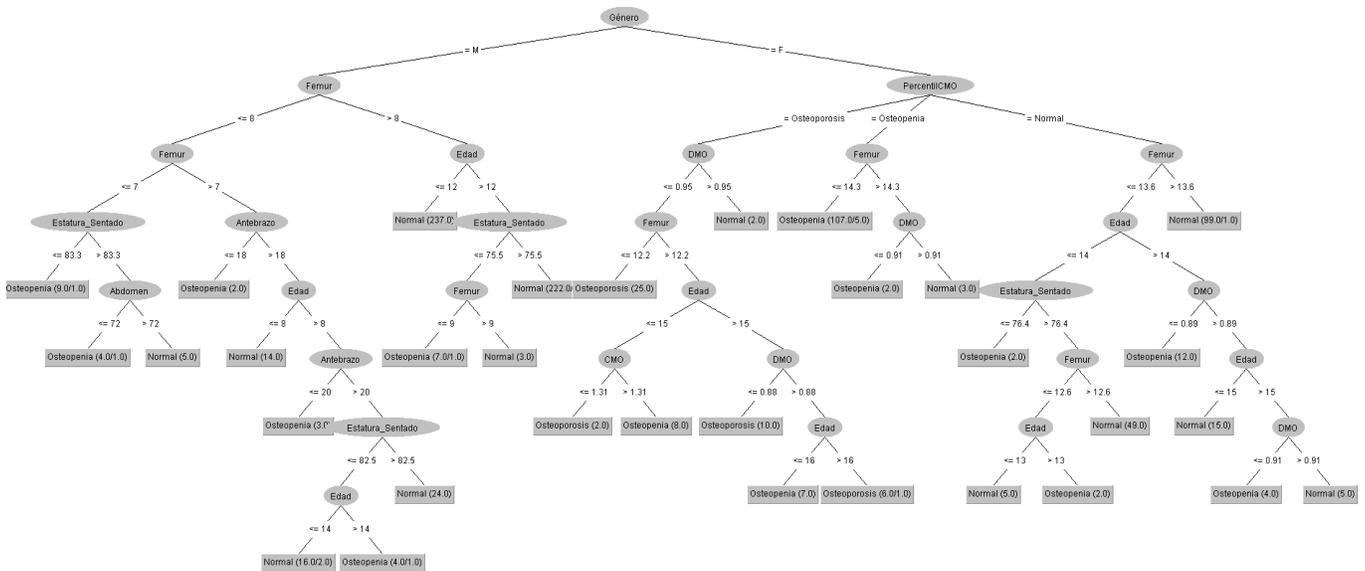


Fig. 2 Árbol de decisión para la clasificación de la salud ósea utilizado en la herramienta Weka.

En base a los resultados presentados y tomando en cuenta todos los indicadores que se obtuvo con Weka, se tiene muy claro que el algoritmo que mejor clasifica este conjunto de datos es el *Random Forest*, ya que en un principio es el que tiene un mayor número de instancias clasificadas, no depende del factor de aleatoriedad y tiene una precisión muy alta. Sin embargo, también hay que tener en cuenta que los otros algoritmos de clasificación no se quedan muy alejados del *Random Forest*, por lo que también se deberían tomar en cuenta para similares estudios realizados en un futuro.

Para aplicar los análisis realizados, se desarrolló una herramienta web denominada BHPProject donde se aplicó las fórmulas de regresión, calculan el estado de la salud ósea de los usuarios.

En la Fig 3, se muestra la pantalla que permiten el ingreso de los datos de un usuario y la realización de los cálculos utilizando el árbol de decisión.

Resultados de la Herramienta

Estos son los resultados para YOVANA IBIA:

DMO (Densidad mineral ósea): 0.93 (g/cm ²)
Estado según DMO: Normal
CMO (Contenido Mineral Óseo): 1.66(kg)
Estado según CMO: Normal
IMC (Índice de Masa Corporal): 20.46(kg/m ²)
Estado según IMC: Peso normal
Estado de Adiposidad: Normal
Clasificación salud ósea: Normal
Su identificador es : 3968

Fig. 3 Vista de resultados ejemplo de la Herramienta BHPProject

En la Fig 4, se muestra la ubicación dentro de un cuadro de percentiles que determina la salud ósea de los escolares estudiados.

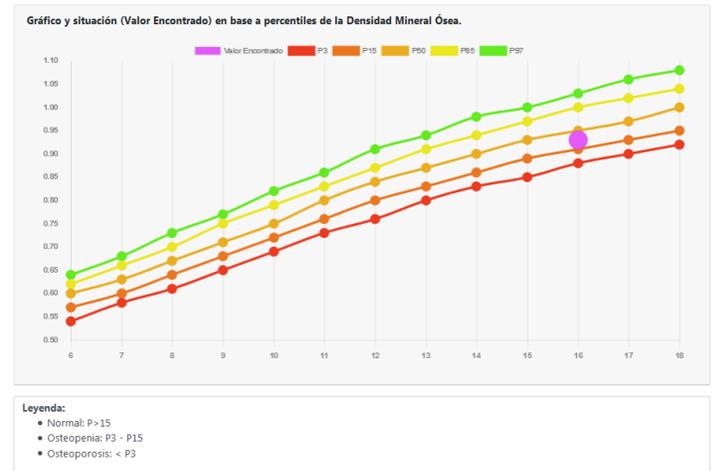


Fig. 4 Resultado ejemplo de crecimiento y situación en base a los percentiles de la densidad mineral ósea de la Herramienta BHPProject.

CONCLUSIONES

Se ha obtenido dos resultados consistentes. Inicialmente el análisis adecuado de los algoritmos de clasificación de salud ósea en base al proceso de extracción del conocimiento y el desarrollo de una herramienta en un ambiente web donde se aplique dicho análisis.

La salud ósea se ha analizado mediante técnicas de minería de datos, donde se ha determinado que el mejor algoritmo entre los analizados es el *Random Forest*, con un 94.87% de instancias correctamente clasificadas. Estos resultados sugieren clasificar los datos de DMO y CMO con dicho algoritmo

A su vez, se ha realizado el desarrollo de una herramienta de software que, utilizando las fórmulas de regresión, calculan el estado de la salud ósea de un niño y adolescente entre los 6 y 18 años.

Las técnicas y herramientas utilizadas se consideran las correctas para hacer un análisis de predicción de la salud ósea en términos de DMO y CMO para niños y adolescentes.

La herramienta de software desarrollado es práctico y fácil de usar, pudiendo ser usada por profesores de colegio y especialistas de la salud, lo que beneficiará a la población en general para determinar su estado en su salud ósea en una edad temprana de vida.

AGRADECIMIENTO

Se agradece a la financiación de los fondos para la investigación UCSM 2015-II con Resolución N° 241554-R-2017.de la Universidad Católica de Santa María.

REFERENCIAS

- [1] A. S. Cruz, H. C. Lins, R. V. A. Medeiros, J. M. F. Filho, and S. G. da Silva, "Artificial intelligence on the identification of risk groups for osteoporosis, a general review," *Biomed. Eng. Online*, 2018.
- [2] C. H. Hsieh, R. H. Lu, N. H. Lee, W. T. Chiu, M. H. Hsu, and Y. C. Li, "Novel solutions for an old disease: Diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks," *Surgery*, 2011.
- [3] P. Larranaga, "Machine learning in bioinformatics," *Brief. Bioinform.*, 2006.
- [4] A. Andreoli, M. Monteleone, M. Van Loan, L. Promenzio, U. Tarantino, and A. De Lorenzo, "Effects of different sports on bone density and muscle mass in highly trained athletes," *Med. Sci. Sports Exerc.*, 2001.
- [5] R. Marcus, "Role of exercise in preventing and treating osteoporosis," *Rheumatic Disease Clinics of North America*. 2001.
- [6] F. J. De Cos Juez, F. S. Lasheras, P. J. García Nieto, and M. A. S. Suárez, "A new data mining methodology applied to the modelling of the influence of diet and lifestyle on the value of bone mineral density in post-menopausal women," *Int. J. Comput. Math.*, vol. 86, no. 10–11, pp. 1878–1887, Nov. 2009.
- [7] F. L. Fonseca *et al.*, "A relação entre a pressão arterial e índices antropométricos na infância/adolescência e o comportamento das variáveis de risco cardiovascular na fase adulta jovem, em seguimento de 17 anos: estudo do Rio de Janeiro," *Rev SOCERJ*, vol. 21, no. 5, pp. 281–290, 2008.
- [8] R. Gómez-Campos, R. C. Huenul, C. U. Albornoz, C. Luarte-Rocha, V. I. Quispe, and M. A. C. Bolaños, "Densidad mineral ósea en una muestra de jóvenes chilenos practicantes de diversas modalidades deportivas," *Salud Uninorte*, 2017.
- [9] C. Mølgaard, B. L. Thomsen, A. Prentice, T. J. Cole, and K. F. Michaelsen, "Whole body bone mineral content in healthy children and adolescents," *Arch. Dis. Child.*, 1997.
- [10] K. Rubin, V. Schirduan, P. Gendreau, M. Sarfarazi, R. Mendola, and G. Dalsky, "Predictors of axial and peripheral bone mineral density in healthy children and adolescents, with special attention to the role of

- puberty," *J. Pediatr.*, 1993.
- [11] R. Gomez-Campos *et al.*, "Fat-Free Mass and Bone Mineral Density of Young Soccer Players: Proposal of Equations Based on Anthropometric Variables," *Front. Psychol.*, 2019.
- [12] R. Gómez-Campos, C. L. Andruske, M. De Arruda, C. Urra Albornoz, and M. Cossio-Bolaños, "Proposed equations and reference values for calculating bone health in children and adolescent based on age and sex," *PLoS One*, 2017.
- [13] V. Saphthagirivasan and M. Anburajan, "Diagnosis of osteoporosis by extraction of trabecular features from hip radiographs using support vector machine: An investigation panorama with DXA," *Comput. Biol. Med.*, 2013.
- [14] N. Kilic and E. Hosgormez, "Automatic Estimation of Osteoporotic Fracture Cases by Using Ensemble Learning Approaches," *J. Med. Syst.*, 2016.
- [15] F. J. de Cos Juez, M. A. Suárez-Suárez, F. Sánchez Lasheras, and A. Murcia-Mazón, "Application of neural networks to the study of the influence of diet and lifestyle on the value of bone mineral density in post-menopausal women," *Math. Comput. Model.*, vol. 54, no. 7–8, pp. 1665–1670, Oct. 2011.
- [16] C. Ordóñez, J. M. Matías, J. F. de Cos Juez, and P. J. García, "Machine learning techniques applied to the determination of osteoporosis incidence in post-menopausal women," *Math. Comput. Model.*, 2009.
- [17] F. Firouzi *et al.*, "A decision tree-based approach for determining low bone mineral density in inflammatory bowel disease using WEKA software," *Eur. J. Gastroenterol. Hepatol.*, 2007.
- [18] C. Kruse, P. Eiken, and P. Vestergaard, "Machine Learning Principles Can Improve Hip Fracture Prediction," *Calcif. Tissue Int.*, 2017.
- [19] T. K. Yoo *et al.*, "Osteoporosis Risk Prediction for Bone Mineral Density Assessment of Postmenopausal Women Using Machine Learning," *Yonsei Med. J.*, vol. 54, no. 6, p. 1321, 2013.
- [20] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," *Commun. ACM*, vol. 39, no. 11, pp. 27–34, 1996.
- [21] R. L. Mirwald, A. D. G. Baxter-Jones, D. A. Bailey, and G. P. Beunen, "An assessment of maturity from anthropometric measurements," *Med. Sci. Sport. Exerc.*, vol. 34, no. 4, pp. 689–694, 2002.
- [22] M. A. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explor.*, vol. 11, no. 1, pp. 10–18, 2009.
- [23] M. N. Adnan and M. Z. Islam, "Forest PA: Constructing a decision forest by penalizing attributes used in previous trees," *Expert Syst. Appl.*, 2017.
- [24] C. P. De Lira *et al.*, "Use of data mining to predict the risk factors associated with osteoporosis and osteopenia in women," *CIN - Comput. Informatics Nurs.*, 2016.
- [25] X. Jiang, B. Cai, D. Xue, X. Lu, G. F. Cooper, and R. E. Neapolitan, "A comparative analysis of methods for predicting clinical outcomes using high-dimensional genomic datasets," *J. Am. Med. Inform. Assoc.*, 2014.
- [26] J. Thongkam, V. Sukmak, G. Xu, and Y. Zhang, "Building Decision Rules for Decision-Making: A Study of Thai Breast Cancer Survivability Prediction," in *2010 THE 3RD INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE AND INDUSTRIAL APPLICATION (PACIIA2010)*, VOL V, 2010, pp. 213–216.
- [27] A. Palacios, L. Sánchez, I. Couso, and S. Destercke, "An extension of the FURIA classification algorithm to low quality data through fuzzy rankings and its application to the early diagnosis of dyslexia," *Neurocomputing*, 2016.