



Revista EIA, ISSN 1794-1237 /
e-ISSN 2463-0950
Año XVII/ Volumen 17/ Edición N.34
Julio-Diciembre de 2020
Reia34021 pág 1-19

Publicación científica semestral
Universidad EIA, Envigado, Colombia

PARA CITAR ESTE ARTÍCULO / TO REFERENCE THIS ARTICLE /

Becerra Álvarez, J.G.; Cortés Osorio, J.A. (2020). Propuesta supervisada de aprendizaje de máquina para la estimación de la edad biológica humana basada en odontología forense. Revista EIA, 17(34), Julio-Diciembre, Reia34021. <https://doi.org/10.24050/reia.v17i34.1392>

✉ Autor de correspondencia:

Becerra Álvarez, J.G. (Johana Gabriela): Universidad Tecnológica de Pereira, Colombia. Correo electrónico: gabriellabecerra@utp.edu.co

Recibido: 27-12-2019
Aceptado: 18-06-2020
Disponibile online: 09-11-2020

Propuesta supervisada de aprendizaje de máquina para la estimación de la edad biológica humana basada en odontología forense

✉ JOHANA GABRIELA BECERRA ÁLVAREZ¹
JIMY ALEXÁNDER CORTÉS OSORIO¹

1. Universidad Tecnológica de Pereira

Resumen

Código de Procedimiento Penal Colombiano avala el uso apropiado de las piezas dentales para la identificación de individuos, por lo que esta técnica resulta una herramienta de gran importancia en las instituciones que apoyan la administración de justicia. Los métodos clásicos de identificación humana, en especial, los usados para la determinación de la edad biológica de muerte en adultos, han sido desarrollados usando información de poblaciones con características morfológicas, métricas y culturales diferentes a las de Colombia. En consecuencia, se han obtenido tasas de error altas en sus resultados. En esta investigación, se calculó inicialmente la edad biológica usando el método clásico de Lamendin. Posteriormente, con las medidas directas e indirectas, también usadas en el método de Lamendin, se realizó el entrenamiento de las técnicas de aprendizaje de máquina conociendo la edad cronológica de los individuos de la muestra. Se compararon los resultados de las técnicas de regresión de aprendizaje de máquina supervisado: Máquinas de Soporte Vectorial (SVM), Procesos Gaussianos por Regresión (GPR) y Conjuntos de Árboles (Ensembles of Trees), entre otras. Se estudió una muestra de 48 piezas dentales de una sola raíz pertenecientes a 45 individuos de nacionalidad colombiana, con edades cronológicas de fallecimiento conocidas dentro del rango de edades desde los 19 años hasta los 81 años. Por otro lado, se utilizaron el pie de rey digital y el macroscopio de comparación para estimar las alturas de las características morfológicas en las piezas dentales. La mejor exactitud en la estimación de la edad biológica se alcanzó con GPR que presentó una **raíz del error medio cuadrático (RMSE) de 3,37 años** en la identificación, la cual se contrastó con la exactitud alcanzada por el método de Lamendin con un **RMSE de 15,52 años**. Esta investigación evidencia que el aprendizaje de máquina con regresión GPR, en especial para el caso colombiano, resulta una herramienta **válida para la estimación de la edad biológica de muerte** con errores muy inferiores a los entregados por técnicas tradicionales de origen europeo.

Palabras clave: aprendizaje de máquina, dientes, edad biológica, identificación humana, Lamendin.

Supervised Machine Learning Proposal For The Estimation Of Human Biological Age Based On Forensic Dentistry

Abstract

Colombian Criminal Procedure Code supports the appropriate use of dental pieces for the identification of individuals, so this technique is a tool of great importance for the administration of justice institutions. The classic methods of human identification, especially those used to estimate the biological age of death in adults such as Lamendin, have been developed using information from populations with morphological, metric and cultural characteristics different from those in Colombia. Consequently, Colombian forensic specialists have obtained high error rates in age estimation results. In this research, the biological age was initially calculated using the classic method of Lamendin. Subsequently, with the direct and indirect measures, also used in the Lamendin method, the training of machine learning techniques was carried out knowing the chronological age of the individuals in the sample. The results of the supervised machine learning regression techniques were compared: Vector Support Machines (SVM), Gaussian Regression Processes (GPR) and Ensembles of Trees (EoT), among others. A sample of 48 single-root dental pieces belonging to 45 individuals of Colombian nationality was studied, with chronological ages of death known within the age range from 19 to 81 years. Additionally, the digital caliper gauge and the comparison microscope were used to measure the heights of the morphological characteristics in the dental pieces. The best accuracy in the estimation of the biological age was achieved with the GPR regression that showed a root mean square error (RMSE) of 3.37 years in the identification, which was compared with the accuracy reached by the Lamendin method with a RMSE of 15.52 years. This research shows that the GPR machine learning regression, especially for the Colombian case, is a valid tool for estimating the biological age of death with much lower errors than those achieved by traditional European techniques.

Keywords: biological age, forensic dentistry, human identification, Lamendin, Machine learning, tooth.

1. Introducción

Si bien las técnicas de aprendizaje de máquina han sido usadas en diversas aplicaciones como lo son la identificación de instrumentos musicales (Tobón & Cortés, 2018). Métodos para clasificar las regiones promotoras en organismos eucariotas que toman como entrada las secuencias de ADN (Bedoya & Bustamante, 2011). Estas aproximaciones a la solución de problemas se pueden utilizar también para las ciencias forenses.

En Colombia el artículo 251 del Código de Procedimiento Penal, respalda el uso apropiado de las evidencias dentales para la identificación de individuos (Pretelt de la Vega, 2004). La existencia de víctimas mortales por parte del conflicto armado “a través de la historia” ha sido uno de los factores que promueven el uso de los dientes en la identificación; debido al mal estado en el cual se encuentran los cadáveres, en la mayoría de los casos, por diferentes factores como: el paso del tiempo, la humedad, fuego, etc. (Vilcapoma Guerra, 2012).

En muchos casos, los dientes son la única evidencia humana con la que se cuenta para poder realizar la identificación de un individuo. En el proceso, por medio del uso de las muestras dentales, se puede reconstruir la osteobiografía general: sexo, hábitos, ascendencia y edad (Suárez Ponce, 2014).

Uno de los primeros casos reconocidos del uso de los dientes para la identificación humana, ocurrió en París (1897). El doctor cubano Oscar Amoedo realizó la identificación de cadáveres haciendo uso de evidencias dentales (Krenzer, 2005).

Pronto, la odontología forense llega para definir sobre el manejo y el examen adecuado de la evidencia dental, teniendo como finalidad la identificación humana. Esta presenta aplicaciones en: el derecho laboral, penal y civil.

Como ya se indicó, la edad de un individuo se puede estimar haciendo uso de la evidencia dental. Esta se denomina edad biológica, pues esta es un indicativo del grado de desgaste en los órganos. En contraste la edad cronológica, se define en función del tiempo transcurrido desde el nacimiento del individuo (Alvarado García & Salazar Maya, 2014). El método clásico de estimación de la edad biológica en adultos mayores de 25 años es conocido como Lamendin. Este usa dos características dentales: la Periodontosis y la Translucidez Radicular en dientes de una sola raíz (Lamendin, et al., 1992).

A diferencia del método reconocido de Lamendin, las técnicas de aprendizaje de máquina supervisado por regresión tienen la ventaja de valerse de datos continuos que son verificables, para aprender de ellos y poder así dar soluciones pertinentes a necesidades y problemáticas (García Cambronero & Gómez Moreno, 2006).

En esta investigación, en la primera parte, se calculó el valor RMSE del método de Lamendin para predecir la edad biológica en la muestra dental propuesta. Posteriormente, se contrastó respecto al error obtenido del entrenamiento de las técnicas de aprendizaje de máquina supervisado: Máquinas de Soporte Vectorial (SVM), Procesos Gaussianos por Regresión (GPR) y Conjuntos de Árboles (Ensembles of Trees). Finalmente, en este artículo, se evaluó y comparó cuál de las técnicas de aprendizaje de máquina supervisado, predice la tasa de RMSE menor de la edad cronológica en comparación con la tasa de error de la edad biológica estimada por el método de Lamendin.

En términos de justicia, paz y verdad el uso de las técnicas de aprendizaje de máquina supervisado se propone como una herramienta que puede apoyar con elementos de verdad en un proceso de identificación y restitución de derechos.

A. Método de estimación de la edad biológica

H. Lamendin (1988), propuso una ecuación para estimar la edad biológica en adultos mayores de 25 años, en la cual se tenían como parámetros principales la periodontosis y la translucidez radicular. En este método se utilizan dientes monorradiculares que ya hayan emergido en su totalidad de las encías, que se encuentren en buen estado y que no tengan enfermedades periodontales. La muestra original utilizada por los autores fue de 306 dientes de una sola raíz, pertenecientes a 208 individuos franceses (135 hombres y 73 mujeres), y con edades cronológicas conocidas (22-90 años). El error medio obtenido fue de 10 años; este error era la diferencia promedio entre la edad cronológica y edad biológica.

B. Técnicas de aprendizaje de máquina supervisado

Aprendizaje de máquina radica en enseñar a la computadora para que aprenda de la experiencia a través de datos conocidos. Los algoritmos de aprendizaje automático utilizan métodos estratégicos para manejar la información directamente de los datos sin depender de una ecuación única predeterminada como modelo (García Cambronero & Gómez Moreno, 2006). Por lo tanto, los algoritmos se tienen que nutrir con

información que favorezca en el aprendizaje. Su rendimiento puede, en algunos casos, depender del tamaño de la muestra de datos proporcionada. Fundamentalmente, existen dos tipos de técnicas afines al aprendizaje de máquina. La aproximación que encuentra patrones ocultos o estructuras intrínsecas en los datos de entrada se denomina aprendizaje no supervisado. Por otro lado, el aprendizaje supervisado crea un modelo con los datos conocidos y sus salidas verdaderas para predecir resultados a la salida. Pueden ser de clase o de regresión (Pelaez Chávez, 2012). Como caso particular del aprendizaje supervisado, los algoritmos de regresión predicen respuestas continuas, para predecir salidas numéricas (Baviera, 2017).

Las máquinas de soporte vectorial (SVM) por regresión se consideran una técnica no paramétrica porque se basa en las funciones del kernel. Para ellas, el conjunto de datos de entrenamiento contiene variables predictores y valores de respuesta observados. Esta técnica encuentra una función que se desvíe poco de los valores de entrenamiento, y de ser posible, sea lo más plano posible (Drucker, et al., 1997; Vapnik, et al., 1997).

Los modelos de regresión de un proceso gaussiano (GPR) son probabilísticos basados en un kernel no paramétrico. A partir del entrenamiento de datos se obtienen funciones que siguen una distribución gaussiana, las cuales obtienen escalares que siguen el comportamiento de esta en conjunto. Este tipo de comportamiento aporta un número elevado de escalares que provienen de una función Gaussiana.

Los algoritmos de conjuntos de árboles pueden ser generados a partir de un conjunto de datos artificiales, mediante el muestreo aleatorio de puntos del espacio de la instancia. Las etiquetas de las clases predichas son asignadas por el clasificador de conjunto para aprender un conjunto de reglas o árboles de decisión de este nuevo conjunto de datos, y su complejidad aumenta mucho más con el tamaño del conjunto de datos (Witten, et al., 2011).

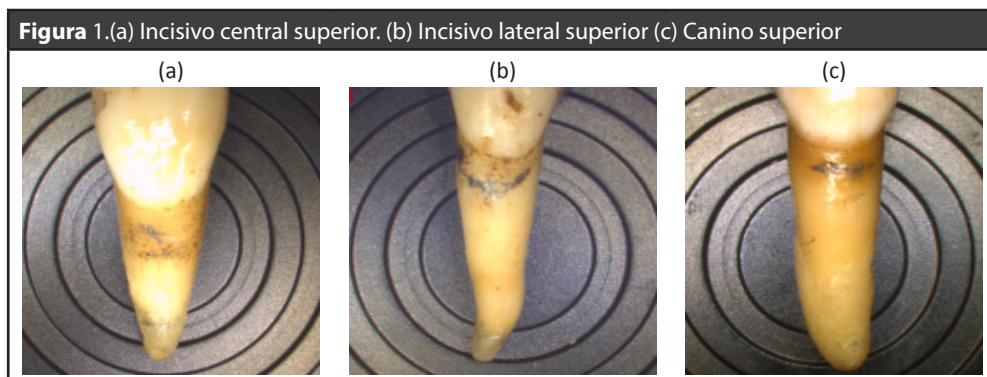
C. Métodos de reducción y selección de características

El método de validación cruzada proporciona una buena estimación de la precisión predictiva del modelo final entrenado utilizando el conjunto completo de datos. El método requiere múltiples ajustes, pero hace un uso eficiente de todos los datos, por lo que funciona bien para conjuntos de datos pequeños. La forma estándar de predecir la tasa de error de una técnica de aprendizaje dada una única muestra fija de datos es utilizar una validación cruzada estratificada diez veces. Los datos son divididos aleatoriamente en k partes en las que la clase se representa en aproximadamente las mismas proporciones que en el conjunto de datos completo. Así, el procedimiento de aprendizaje se ejecuta un total de k veces en diferentes conjuntos de entrenamiento (cada conjunto tiene mucho en común con los demás) (Witten, et al., 2011). Por otro lado, la selección de características disminuye la dimensionalidad de los datos al seleccionar un subconjunto de variables predictoras para crear un modelo. Los criterios de selección implican reducción de una medida específica de error predictivo para los modelos que se ajustan a diferentes subconjuntos. Restringiendo las características seleccionadas y el tamaño del subconjunto (Guyon & Elisseeff, 2003).

2. Materiales y métodos

La muestra dental utilizada para esta investigación estaba compuesta por caninos, incisivos centrales e incisivos laterales, véase la **Figura 1**. El número total de la muestra fue de 48 piezas dentales monorradiculares, pertenecientes a 45 individuos colombianos, proveniente de cuerpos exhumados en el año 2002 del cementerio central de Bogotá, con edades conocida entre los 19 a los 81 años. La doctora Luz Dary Escobar, servidora de la Fiscalía General con sede en Bogotá; facilitó la muestra

para este estudio bajo los parámetros de confidencialidad requeridos. Así mismo, cabe aclarar que los datos objeto a verificar, no fueron suministrados durante las prácticas, solo hasta que fueron requeridos dentro del proceso de desarrollo investigativo.



A. Protocolo de medición la transparencia radicular (T) y la periodontosis (P)

El protocolo de medición de la transparencia radicular (T) y la altura gingival o periodontosis (P) se realizó en el Laboratorio de Identificación Humana de la Fiscalía General de la Nación, Seccional Risaralda, donde se midieron: la altura de la raíz, altura de la periodontosis y altura de la translucidez radicular, por tres observadores distintos usando el mismo instrumento.

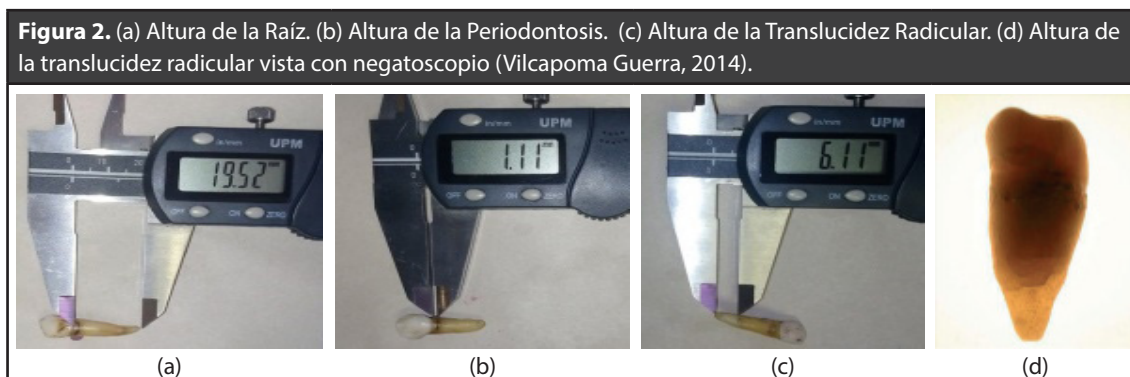
Las medidas directas realizadas fueron:

- * La medida de la altura de la raíz (HR (mm)): se realizó desde el ápice hasta la unión del cemento-esmalte en la superficie vestibular. Véase **Figura 2(a)**.

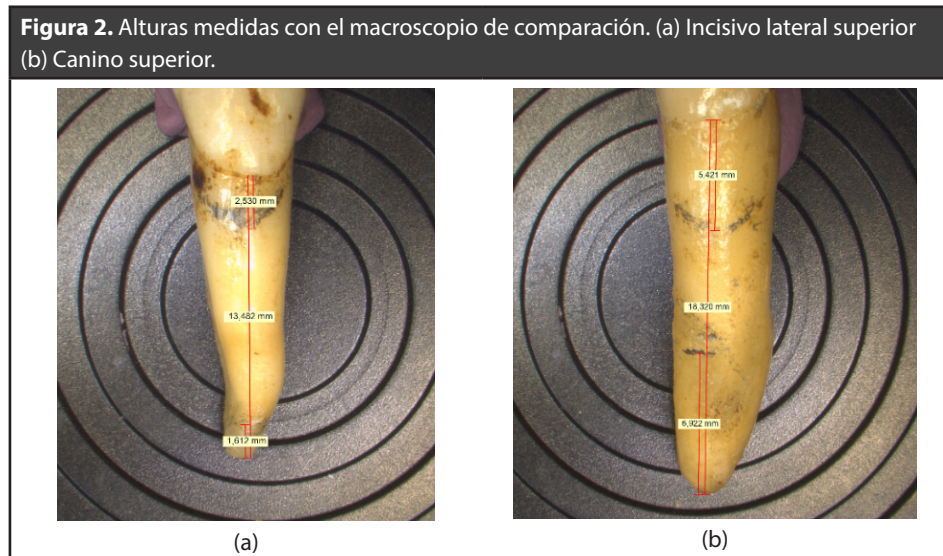
- * La medida de la altura de la periodontosis (HPER (mm)): se aprecia como un área suave y amarillenta debajo del esmalte y es más oscura que éste, pero más clara que el resto de la raíz. Se realizó entre la unión cemento-esmalte y la línea de inserción del tejido blando. Véase **Figura 2(b)**.

- * La medida de la altura de la translucidez radicular (HTRANS (mm)): esta no aparece antes de los 20 años y se produce por el depósito de cristales de hidroxiapatita dentro de los túbulos dentinarios. Se realizó desde el ápice hasta la unión del cemento-esmalte en la superficie, haciendo uso de un negatoscopio para poder apreciar la zona de medida. Véase **Figura 2(c)** y **2(d)**.

Para realizar cada una de estas medidas se debe marcar con lápiz las zonas mencionadas; lo anterior sin destruir el diente. Cada una de las alturas medidas, se muestran en la **Figura 2**.



Posteriormente las medidas fueron realizadas con fines comparativos con el Macroscopio de Comparación marca LEICA Software: LAS V 4.10 (Rango: (0,4-8) X), como se observa en la **Figura 3**. Cabe mencionar que cada instrumento contaba con su debida calibración a la fecha. Adicionalmente, el Comité de Bioética de la Universidad Tecnológica de Pereira, avaló la realización del proyecto.



A partir de las medidas realizadas se calcularon la Periodontosis (P) y la Translucidez radicular (T) usando (1) y (2), respectivamente.

$$P = \frac{\text{Altura de la periodontosis} \times 100}{\text{Altura de la raíz}} \tag{1}$$

$$T = \frac{(\text{Altura de la transparencia radicular} \times 100)}{\text{Altura de la raíz}} \tag{2}$$

La ecuación propuesta por Lamendin para estimar la edad biológica es presentada en (3):

$$\text{Edad Biológica} = (0,18 \times P) + (0,42 \times T) + 25,53 \tag{3}$$

En la **Tabla 1**, se presentan la información parcial tabulada para la construcción de cada una de las bases de datos de los instrumentos de medida propuestos.

TABLA 1. BASE DE DATOS PARCIAL E ILUSTRATIVA, REALIZADA PARA CADA INSTRUMENTO CON TRES OBSERVADORES. LOS DATOS CORRESPONDEN A UN CANINO INFERIOR

Observador	Muestra	Tipo de Diente	HR (mm)	HPER (mm)	HTRANS (mm)	P	T	Edad Biológica	Edad Cronológica	Rotulado
Observador 1	1	Canino Inferior	17,59	1,48	3,11	8,41	17,68	35	26	143
	2		17,57	1,58	3	8,99	17,07	34	26	
	3		17,56	1,46	3,3	8,31	18,79	35	26	
Observador 2	1		17,53	1,27	1,56	7,24	8,90	31	26	
	2		17,67	1,58	2,36	8,94	13,36	33	26	
	3		17,54	1,85	2,33	10,55	13,28	33	26	
Observador 3	1		17,69	1,72	3,55	9,72	20,06	36	26	
	2		17,22	1,8	3,27	10,45	18,99	35	26	
	3		17,61	1,75	3,45	9,94	19,59	36	26	

Las edades cronológicas de los individuos y las edades biológicas, obtenidas por medio del método de Lamendin, fueron utilizadas para realizar el cálculo de la raíz del error medio cuadrático (RMSE), error absoluto y el cálculo del coeficiente de determinación.

B. Implementación de las 3 técnicas de aprendizaje de máquina supervisado

Las tres técnicas de aprendizaje de máquina supervisado básicas de regresión fueron: máquinas de soporte vectorial (SVM), procesos gaussianos por regresión (GPR) y conjuntos de árboles (Ensembles of Trees). Las bases de datos se construyeron a partir de las medidas directas realizadas por tres observadores expertos igualmente utilizadas para el método de Lamendin: HR (mm), HPER (mm), HTRANS (mm). Seguidamente, se calcularon las medidas indirectas para P y T. Se tomó un total 2592 medidas directas (3 medidas x 3 observadores x 3 atributos directos x 48 piezas dentales monorradiculares x 2 instrumentos) y 1728 indirectas (3 medidas x 3 observadores x 2 atributos indirectos x 49 piezas dentales monorradiculares x 2 instrumentos) para un total de 4320. Esto permitió obtener 432 instancias para cada instrumento de medida. En la **Tabla 2** se presenta una muestra del conjunto de características usadas para construir la base de datos de cada uno de los instrumentos.

TABLA 2. INFORMACIÓN PARCIAL ILUSTRATIVA DE LAS BASES DE DATOS, UTILIZADA PARA IMPLEMENTAR LAS TÉCNICAS DE APRENDIZAJE DE MÁQUINA SUPERVISADO. LOS DATOS CORRESPONDEN A UN CANINO INFERIOR.

HR (mm)	HPER (mm)	HTRANS (mm)	P	T	Edad Cronológica
17,59	1,48	3,11	8,41	17,68	26
17,57	1,58	3	8,99	17,07	26
17,56	1,46	3,3	8,31	18,79	26
17,53	1,27	1,56	7,24	8,90	26
17,67	1,58	2,36	8,94	13,36	26
17,54	1,85	2,33	10,55	13,28	26
17,69	1,72	3,55	9,72	20,06	26
17,22	1,8	3,27	10,45	18,99	26
17,61	1,75	3,45	9,94	19,59	26

En cuanto al método de validez, se usó validación cruzada con K=5. También se realizó selección de características para su reducción. Esta se hizo, dividiendo la información en tres grupos: medidas directas (HR (mm), HPER (mm), HTRANS (mm)), medidas indirectas (P y T) y el número total de características. La evaluación de los métodos se llevó a cabo en una computadora de escritorio con procesador core i5, memoria RAM de 16Gb, 1T disco duro y Windows 8. para la implementación y evaluación de los algoritmos de regresión de aprendizaje de maquina se usó MATLAB R2017a.

C. Técnicas de evaluación

Para evaluar el comportamiento de cada uno de las técnicas de aprendizaje de máquina supervisado utilizadas, se consideraron los errores: la raíz del error medio cuadrático (RMSE), error medio cuadrático (MSE), el coeficiente de determinación (R^2), el error medio absoluto (MAE), error absoluto (E) y el error relativo (E_r).

La raíz del error medio cuadrático (RMSE), se define en (4) como (Willmott & Matsuura, 2005):

$$RMSE = \sqrt{\left(\frac{\sum_{i=1}^N (A_i - A_r)^2}{N}\right)} \tag{4}$$

El error medio cuadrático (MSE), se presenta en (5) como:

$$MSE = \frac{\sum_{i=1}^N (A_i - A_r)^2}{N} \tag{5}$$

El coeficiente de determinación o R^2 , se determina como la medida del grado de fiabilidad o bondad del ajuste del modelo adaptado a un conjunto de datos (Martinez Rodriguez, 2005).

El error medio absoluto se ilustra en (6) (Willmott & Matsuura, 2005).

$$MAE = \left| \frac{\sum_{i=1}^N (A_i - A_r)}{N} \right| \tag{6}$$

Por otro lado, error absoluto (7) es el resultado de una medición menos un valor verdadero del mensurando (JCGM 100, 2008).

$$E = A_i - A_r \tag{7}$$

El error relativo (8) es el error de medición dividido por un valor verdadero del mensurando (JCGM 100, 2008).

$$E_r = \frac{|A_i - A_r|}{A_r} \times 100\% \tag{8}$$

A_i : es el valor obtenido de la magnitud medida.

A_r : es el valor verdadero de la magnitud medida.

3. Resultados y discusión

A. Estimación de la edad biológica por medio del Método de Lamendin

Las edades biológicas obtenidas por medio del método de Lamendin, utilizando el pie de rey digital y el macroscopio de comparación se muestran en las **Tablas 3 y 4**.

TABLA 3. EDADES BIOLÓGICAS CALCULADAS POR DÉCADAS (35 POSIBLES EDADES CRONOLÓGICAS ÚNICAS) USANDO EL MÉTODO DE LAMENDIN CON EL PIE DE REY DIGITAL COMO INSTRUMENTO DE MEDICIÓN.

Consecutivo de muestra	Edad Biológica (años)	Edad Cronológica (años)	Consecutivo de muestra	Edad Biológica (años)	Edad Cronológica (años)
1	58	19	18	39	43
2	28	21	19	46	45
3	35	22	20	40	46
4	33	24	21	40	49
5	37	26	22	42	50

6	34	27	23	45	51
7	34	28	24	49	53
8	41	29	25	40	55
9	37	30	26	42	58
10	35	32	27	45	62
11	44	33	28	37	64
12	37	36	29	46	70
13	50	37	30	52	71
14	42	38	31	56	72
15	35	40	32	46	75
16	48	41	33	46	77
17	41	42	34	43	79
			35	59	81

TABLA 4. EDADES BIOLÓGICAS CALCULADAS POR DÉCADAS (35 POSIBLES EDADES CRONOLÓGICAS ÚNICAS) USANDO EL MÉTODO DE LAMENDIN CON EL MACROSCOPIO DE COMPARACIÓN COMO INSTRUMENTO DE MEDICIÓN.

Número de Edades	Edad Biológica (años)	Edad Cronológica (años)	Número de Edades	Edad Biológica (años)	Edad Cronológica (años)
1	54	19	18	41	43
2	40	21	19	34	45
3	37	22	20	45	46
4	46	24	21	37	49
5	36	26	22	40	50
6	28	27	23	45	51
7	33	28	24	47	53
8	31	29	25	35	55
9	35	30	26	40	58
10	31	32	27	44	62
11	31	33	28	35	64
12	39	36	29	44	70
13	34	37	30	49	71
14	31	38	31	53	72
15	42	40	32	46	75
16	35	41	33	43	77
17	49	42	34	42	79
			35	51	81

En la gráfica de la **Figura 4**, se aprecian las barras de la raíz del error medio cuadrático, obtenido de la estimación de la edad biológica en la muestra.

En la gráfica de la **Figura 5**, se muestra el coeficiente de determinación, el cual se calculó para las edades biológicas obtenidas del método de Lamendin.

Figura 3. Comparación de la raíz del error medio cuadrático usando el método de Lamendin estimado por cada instrumento en años.

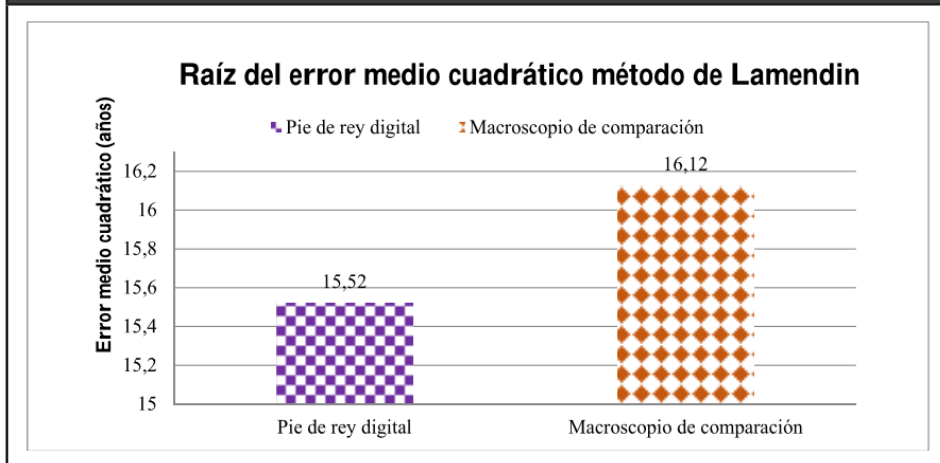
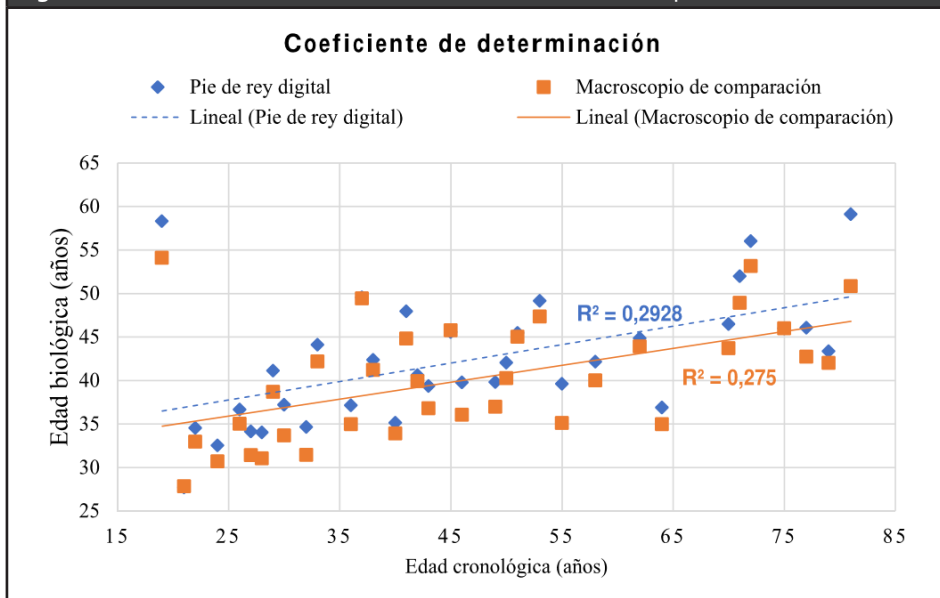


Figura 4. Coeficiente de determinación del método de Lamendin para cada instrumento.



Según lo expuesto en la **Tablas 3 y 4**, de las edades biológicas estimadas por el método de Lamendin y las **Figuras 4 y 5**; se evidenció que el instrumento de medición más pertinente para realizar la toma de medida de cada una de las alturas HR, HPER y HTRANS, en base al método de Lamendin, fue el pie de rey digital, la raíz de su error medio cuadrático se estimó de 15,52 años; este instrumento permitió realizar la medida de los dientes en sus tres dimensiones. Por el contrario, el macroscopio de comparación toma una imagen en dos dimensiones, lo que marca un leve aumento en la estimación de su RMSE el cual fue de 16,00 años.

B. Implementación de las 3 técnicas de aprendizaje de máquina supervisado

En este apartado se muestra la información obtenida de realizar el entrenamiento de las bases de datos haciendo uso de la aplicación Regression Learner y de un pie de rey como instrumento de medición en las **Tablas 4, 5 y 6**.

Los resultados obtenidos del entrenamiento, utilizando la selección de características: periodontitis y translucidez radicular, que se muestran en la **Tabla 6**. Estos evidencian que la técnica de aprendizaje de máquina procesos Gaussianos por regresión, en específico el kernel Exponential, se destacó por tener una raíz del error medio cuadrático de 8,99 años un coeficiente de determinación de 0,75; siendo esta la técnica que realizó la mejor predicción. En este caso se tiene que, para dos características, el que presentó una mayor velocidad de predicción fue: SVM Quadratic de 17000 obs/s. El modelo que menor tiempo de entrenamiento tuvo fue SVM Coarse Gaussian con 0,28701s.

TABLA 4. RESULTADOS DEL ENTRENAMIENTO CON LA BASE DE DATOS DEL PIE DE REY CON 2 CARACTERÍSTICAS P Y T.

MODELOS		RESULTADOS					
características: P y T.		RMSE (años)	Coefficiente de determinación	MSE (años) ²	MAE (años)	Velocidad de predicción (obs/s)	Tiempo de entrenamiento (s)
ET	Bagged Tres	10,02	0,69	100,34	7,37	4100	1,9613
	Boosted Trees	9,7	0,71	94,06	7,03	4500	2,4638
GPR	Exponential	8,99	0,75	80,84	5,86	11000	2,5631
	Matern 5/2	9,61	0,71	92,36	6,48	11000	2,5381
	Rational Quadratic	9,39	0,73	88,21	6,22	5300	9,3124
	Squared Exponential	9,96	0,69	99,16	6,85	12000	2,6984
SVM	Coarse Gaussian	14,54	0,34	211,54	10,68	15000	0,28701
	Cubic	13,3	0,45	177,01	9,77	14000	1,6235
	Fine Gaussian	9,87	0,7	97,36	6,05	15000	0,34771
	Medium Gaussian	10,93	0,63	119,49	7,69	15000	0,3195
	Linear	15,91	0,22	253,17	11,08	9700	1,4376
	Quadratic	14,54	0,35	211,35	10,86	17000	0,72458

Una vez más los resultados obtenidos del entrenamiento utilizando la selección de características: HR, HPER y HTRANS, que son mostrados en la **Tabla 7**, permiten verificar que la técnica de aprendizaje de máquina GPR Rational Quadratic arrojó un RMSE de 3,37 años y un R-squared de 0,96, siendo ella la que realizó la mejor predicción. En este caso se tiene que, para tres características, los modelos que presentaron una mayor velocidad de predicción fueron: SVM Quadratic y SVM Coarse Gaussian, ambos con 16000 obs/s. Por otro lado, el modelo que menor tiempo de entrenamiento entregó fue SVM Coarse Gaussian con 0,29378s.

TABLA 5. RESULTADOS DEL ENTRENAMIENTO CON LA BASE DE DATOS DEL PIE DE REY CON 3 CARACTERÍSTICAS HR, HPER Y HTRANS.

MODELOS		RESULTADOS					
Características: HR, HPER y HTRANS.		RMSE (años)	Coefficiente de determinación	MSE (años) ²	MAE (años)	Velocidad de predicción (Obs./s)	Tiempo de entrenamiento (s)
ET	Bagged Tres	9,22	0,74	85,04	7,25	3800	2,2513
	Boosted Trees	8,45	0,78	71,48	6,25	3700	2,6788

GPR	Exponential	3,58	0,96	12,85	1,88	8300	14,185
	Matern 5/2	3,69	0,96	13,63	2,01	11000	4,1169
	Rational Quadratic	3,37	0,96	11,34	1,75	2700	18,71
	Squared Exponential	3,84	0,95	14,73	2,21	12000	3,3164
SVM	Coarse Gaussian	15,18	0,29	230,52	11,05	16000	0,29378
	Cubic	11,74	0,57	137,73	8,5	15000	2,6862
	Fine Gaussian	4,49	0,94	20,2	2,95	14000	0,40458
	Medium Gaussian	8,62	0,77	74,31	6,37	14000	0,34922
	Linear	15,81	0,23	249,99	11,22	7600	1,8356
	Quadratic	13,83	0,41	191,37	10,49	16000	0,6515

Los resultados obtenidos del entrenamiento en la aplicación Regression Learner utilizando selección de características: HR, HPER, HTRANS, P y T. Mostrados en la **Tabla 8**, medidos con el pie de rey digital, demuestran que la técnica de aprendizaje de máquina GPR exponential arrojó un RMSE de 3,94 años y un R-squared de 0,95, siendo esta la técnica que realizó la mejor predicción. En este caso se tiene que, para cinco características, el que presentó una mayor velocidad de predicción fue: Support Vector Machine Regression Coarse Gaussian con 16000 obs/s. Por otro lado, modelo que menor tiempo de entrenamiento tuvo fue Support Vector Machine Regression Coarse Gaussian con 0,27344s.

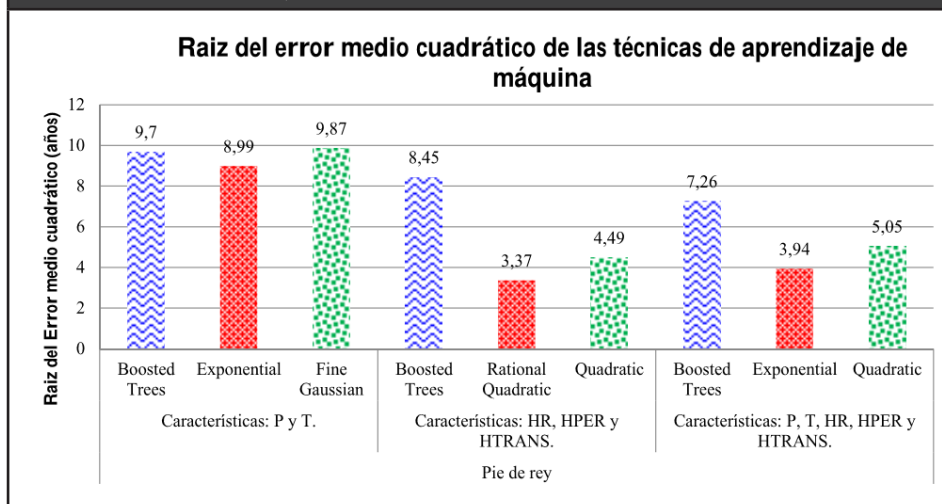
TABLA 6. RESULTADOS DEL ENTRENAMIENTO CON LA BASE DE DATOS DEL PIE DE REY CON 5 CARACTERÍSTICAS HR, HPER, HTRANS, P Y T.

MODELOS		RESULTADOS					
Características: HR, HPER, HTRANS, P y T.		RMSE (años)	Coefficiente de determinación	MSE (años) ²	MAE (años)	Velocidad de predicción (obs/s)	Tiempo de entrenamiento (s)
ET	Bagged Tres	8,02	0,8	64,26	5,94	3800	2,5854
	Boosted Trees	7,26	0,84	52,65	5,41	2300	6,5732
GPR	Exponential	3,94	0,95	15,49	2,09	11000	4,6326
	Matern 5/2	4,35	0,94	18,94	2,23	10000	2,7825
	Rational Quadratic	4	0,95	15,98	2,08	3200	17,749
	Squared Exponential	4,85	0,93	23,47	2,5	12000	2,6514
SVM	Coarse Gaussian	15,07	0,3	227,07	10,85	16000	0,27344
	Cubic	11,89	0,56	141,37	7,78	15000	61
	Fine Gaussian	5,05	0,92	25,51	3,18	15000	0,40331
	Medium Gaussian	8,33	0,79	69,32	6,05	15000	0,32848
	Linear	15,82	0,22	250,36	11,16	2900	4,3699
	Quadratic	13,43	0,44	180,32	9,77	14000	2,4344

En la **Figura 6**, se muestra la tasa de error obtenida del entrenamiento de cada técnica de aprendizaje de máquina supervisado por regresión, haciendo uso del pie de

rey como instrumento de medición. En este gráfico de barras, las azules representan los conjuntos de árboles, las rojas representan los procesos gaussianos y las verdes representan las máquinas de soporte vectorial. Para cada técnica de aprendizaje de máquina se tiene la información del kernel destacado según el número de características del entrenamiento.

Figura 5. Mejores tasas de la raíz del error medio cuadrático presentadas por las 3 técnicas de aprendizaje de máquina supervisado, utilizando los tres grupos de características de la base de datos del pie de rey digital



Conforme a la **Figura 6**, claramente se puede apreciar que los procesos gaussianos representados por las barras rojas, se destacan por que su estimación de la tasa de error. Finalmente las máquinas de soporte vectorial y los conjuntos de árboles obtuvieron resultados de predicción menos favorables.

La información obtenida del entrenamiento realizado con las mediciones adquiridas con el macroscopio de comparación se observan en las **Tablas 9, 10 y 11**.

Los resultados obtenidos del entrenamiento utilizando selección de características: P y T. Mostrados en la **Tabla 9**, demuestran que la Técnica de aprendizaje de máquina GPR Rational Quadratic arrojó un RMSE de 10,33 años y un Coeficiente de determinación de 0,67. Siendo esta la técnica que realizó la mejor predicción. En este caso se tiene que, para dos características, los que tuvieron una mayor velocidad de predicción fueron: SVM Quadratic y SVM Cubic de 19000obs/s. El modelo que menor tiempo de entrenamiento tuvo fue SVM Coarse Gaussian con 0,25708s.

TABLA 7. RESULTADOS DEL ENTRENAMIENTO CON LOS DATOS DEL MACROSCOPIO DE COMPARACIÓN CON 2 CARACTERÍSTICAS P Y T.

MODELOS		RESULTADOS					
Características: P y T		RMSE (años)	Coeficiente de determinación	MSE (años) ²	MAE (años)	Velocidad de predicción (obs/s)	Tiempo de entrenamiento (s)
ET	Bagged Tres	12,06	0,55	145,34	8,63	4300	2,3693
	Boosted Trees	12,21	0,54	149,04	8,65	2500	7,7079

GPR	Exponential	10,52	0,66	110,6	6,63	11000	2,9393
	Matern 5/2	10,52	0,66	110,71	6,62	12000	3,1148
	Rational Quadratic	10,33	0,67	106,73	6,44	5000	10,174
	Squared Exponential	10,77	0,64	116	6,89	12000	2,818
SVM	Coarse Gaussian	15,32	0,27	234,56	10,89	18000	0,25708
	Cubic	15,14	0,29	229,18	10,57	19000	5,04
	Fine Gaussian	11,48	0,59	131,82	6,96	17000	0,30777
	Medium Gaussian	14,89	0,31	221,29	10,29	16000	0,29039
	Linear	15,8	0,23	249,67	11,3	11000	1,9312
	Quadratic	15,6	0,25	243,25	11,16	19000	0,80896

Los resultados obtenidos del entrenamiento en la aplicación Regression Learner utilizando selección de características: HR, HPER y HTRANS; son mostrados en la **Tabla 10**, demuestran que la técnica de aprendizaje de máquina GPR Matern 5/2 arrojó un RMSE de 5,21 años y un R-squared de 0,92, siendo esta la técnica que realizó la mejor predicción. Para este caso se tiene que, para tres características, los modelos que tuvieron una mayor velocidad de predicción fueron: SVM Quadratic y SVM Coarse Cubic, con 17000 obs/s. El modelo que menor tiempo de entrenamiento tuvo fue SVM Coarse Gaussian con 0,3135s.

TABLA 8. RESULTADOS DEL ENTRENAMIENTO CON LA BASE DE DATOS DEL MACROSCOPIO DE COMPARACIÓN, EN LA APLICACIÓN REGRESSION LEARNER CON 3 CARACTERÍSTICAS HR, HPER Y HTRANS.

MODELOS		RESULTADOS					
Características: HR, HPER y HTRANS		RMSE (años)	Coefficiente de determinación	MSE (años) ²	MAE (años)	Velocidad de predicción (obs/s)	Tiempo de entrenamiento (s)
ET	Bagged Tres	9,45	0,72	89,31	6,88	3100	5,3777
	Boosted Trees	9,23	0,74	85,14	6,54	2600	4,9079
GPR	Exponential	5,36	0,91	28,73	2,56	12000	3,1059
	Matern 5/2	5,2	0,92	27,1	2,44	10000	2,8498
	Rational Quadratic	5,39	0,91	29,05	2,46	6000	8,771
	Squared Exponential	5,35	0,91	28,6	2,76	11000	2,7535
SVM	Coarse Gaussian	15,05	0,3	226,49	10,7	16000	0,3135
	Cubic	13,16	0,46	173,17	9,26	17000	12,334
	Fine Gaussian	5,91	0,89	34,93	3,38	13000	0,36159
	Medium Gaussian	11,23	0,61	126,13	7,62	16000	0,45644
	Linear	15,33	0,27	234,88	10,96	11000	1,3106
	Quadratic	14,83	0,32	220,07	10,4	17000	1,5036

Los resultados obtenidos del entrenamiento en la aplicación Regression Learner utilizando selección de características: HR, HPER, HTRANS, P y T, mostrados en la

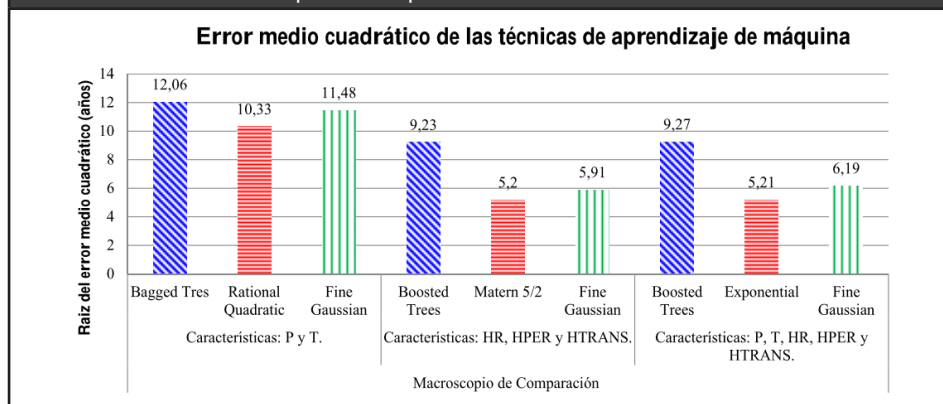
Tabla 11, demuestran que la técnica de aprendizaje de máquina GPR Exponential arrojó un RMSE de 5,2 años y un R-squared de 0,92. En este caso se tiene que, para cinco características, los modelos que tuvieron una mayor velocidad de predicción fueron: SVM Quadratic y SVM Coarse Cubic, con 18000 obs/s. El modelo que menor tiempo de entrenamiento tuvo fue SVM Medium Gaussian con 0,29635s.

TABLA 9. RESULTADOS DEL ENTRENAMIENTO CON LA BASE DE DATOS DEL MACROSCOPIO DE COMPARACIÓN CON 5 CARACTERÍSTICAS HR, HPER, HTRANS, P Y T.

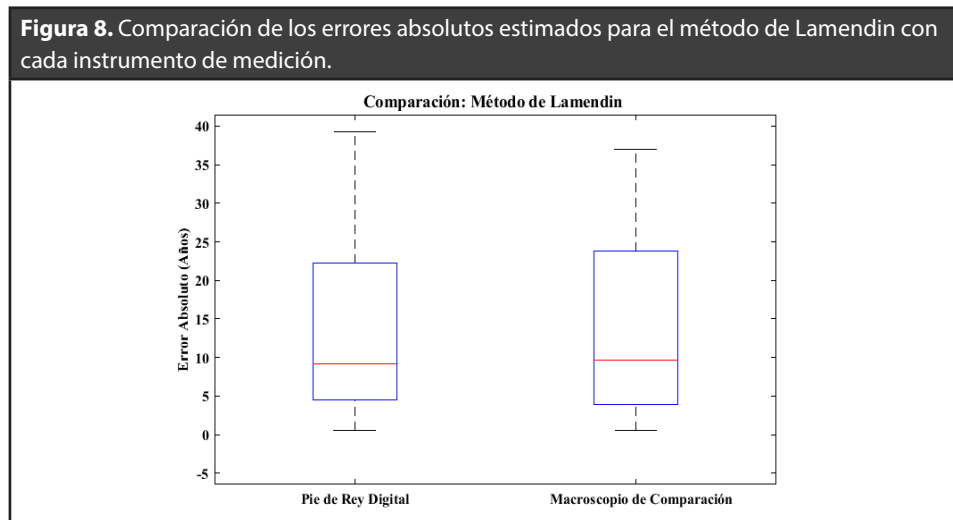
MODELOS		RESULTADOS					
Características: HR, HPER, HTRANS, P y T		RMSE (años)	Coefficiente de determinación	MSE (años) ²	MAE (años)	Velocidad de predicción (obs/s)	Tiempo de entrenamiento (s)
ET	Bagged Tres	9,42	0,72	88,81	6,71	3900	2,2984
	Boosted Trees	9,27	0,73	85,84	6,43	2900	3,91186
GPR	Exponential	5,21	0,92	26,99	2,51	12000	3,027
	Matern 5/2	5,29	0,91	27,94	2,64	11000	2,452
	Rational Quadratic	5,23	0,92	27,4	2,58	5300	8,7505
	Squared Exponential	5,51	0,91	30,37	2,87	11000	2,5522
SVM	Coarse Gaussian	15,11	0,29	228,39	10,67	15000	0,30469
	Cubic	11,83	0,57	139,9	7,99	18000	65,652
	Fine Gaussian	6,19	0,88	38,37	3,47	16000	0,339
	Medium Gaussian	11,71	0,58	137,02	7,79	16000	0,29635
	Linear	15,44	0,26	238,41	10,94	10000	2,1526
	Quadratic	13,45	0,44	181,02	9,21	18000	7,5129

Al igual que en la **Figura 6**, en la **Figura 7**, se muestra la tasa de error obtenida del entrenamiento de cada técnica de aprendizaje de máquina supervisado por regresión, haciendo uso en este caso del macroscopio de comparación, como instrumento de medición. En este gráfico de barras, las azules representan los conjuntos de árboles, las rojas representan los procesos Gaussianos y las verdes representan las máquinas de soporte vectorial. Para cada técnica de aprendizaje de máquina se tiene la información del kernel destacado según el número de características del entrenamiento.

Figura 6. Mejores tasas de la raíz del error medio cuadrático presentadas por las 3 técnicas de aprendizaje de máquina supervisado, utilizando los tres grupos de características de la base de datos del macroscopio de comparación



En la **Figura 8**, se muestra la proyección del histograma de los datos representado en los boxplot. De esta se puede apreciar que los datos del Pie de Rey están menos dispersos y que su error promedio es menor comparado con el macroscopio de comparación.

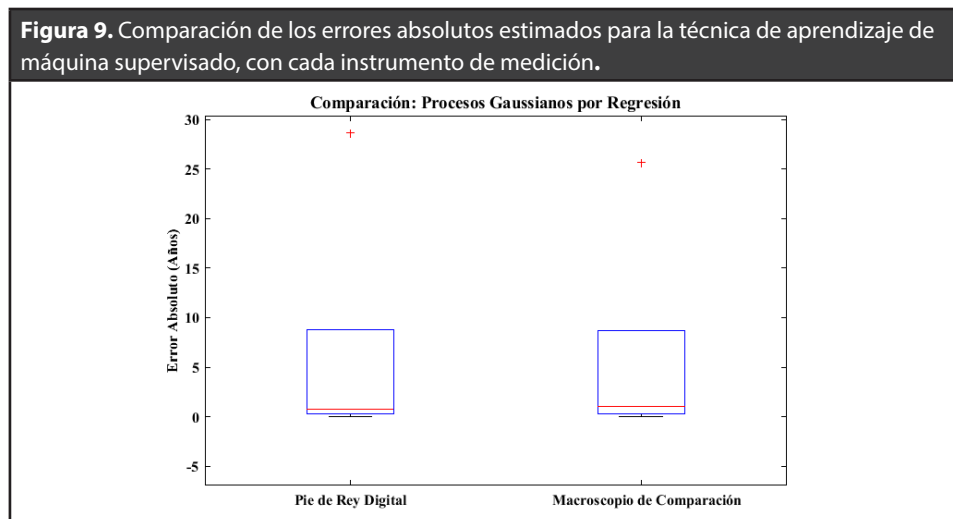


En la **Tabla 12**, están los datos obtenidos para cada instrumento, obtenidos de la comparación del método de Lamendin, respecto al error absoluto estimado, el mínimo, primer cuartil, la mediana, tercer cuartil y el máximo.

TABLA 10. RESULTADOS DEL ERROR ABSOLUTO, OBTENIDO DE LA COMPARACIÓN DEL MÉTODO DE LAMENDIN CON CADA INSTRUMENTO.

Instrumento	Mínimo	Cuartil 1	Mediana	Cuartil 3	Máximo
Pie De Rey Digital	0,14	5,23	8,18	18,61	38,46
Macroscopio De Comparación	0,07	5,62	8,80	16,91	41,04

En la **Figura 9**, se tiene la proyección del histograma de los datos representado en los boxplot. Nuevamente, el Pie de Rey presenta un error absoluto promedio menor que el Macroscopio de comparación para la técnica de regresión Procesos Gaussianos.



En la **Tabla 13**, los datos obtenidos para cada instrumento, obtenidos de la comparación de la técnica de procesos gaussianos por regresión., respecto al error absoluto estimado, el mínimo, primer cuartil, la mediana, tercer cuartil y el máximo.

TABLA 11. RESULTADOS DEL ERROR ABSOLUTO, OBTENIDO DE LA COMPARACIÓN DE LA TÉCNICA DE PROCESOS GAUSSIANOS POR REGRESIÓN.

Instrumento	Mín.	Cuartil 1	Mediana	Cuartil 3	Máx.
Pie De Rey Digital	0,04	0,32	0,74	2,14	28,68
Macroscopio De Comparación	0,00	0,42	0,98	2,97	25,69

4. Conclusiones

El método clásico de Lamendin, en la literatura científica presenta usualmente un error medio de 10,00 años cuando se usaron muestras dentales monorradiculares de origen francés. En contraste, al utilizar muestras al utilizar muestras dentales monorradiculares de origen colombiano se presenta una tasa de error de 15,52 años, es decir, este método se ve afectado por las variaciones morfológicas, fisiológicas y ancestrales, según las características de la población. Como se evidencia en los resultados obtenidos por medio del Método de Lamendin, al usar el pie de rey y el macroscopio de comparación, como instrumento de medición se demuestra que el RMSE entre la edad biológica y la edad cronológica fue mucho mayor, en comparación con el error promedio obtenido por Lamendin et al.

Se puede concluir que el método de Lamendin es una ecuación estática simple, que usa mediciones directas para estimar la edad biológica de un sujeto; sin embargo, no es muy exacta en términos de predecir un valor cercano de la edad cronológica de la muestra de estudio colombiana.

Se concluye que para la predicción de la edad cronológica, fue posible ampliar el rango de estimación de la edad en adultos, de 19 años hasta los 81 años para adultos. En contraste con el rango de error definido por Lamendin que se encuentra definido entre 25 y 60 años.

Se concluye que el uso de aprendizaje de máquina supervisado, especialmente el modelo de regresión de procesos gaussianos es el más indicado para estimar la edad cronológica, porque permiten calcular la distribución predictiva, lo cual corresponde a la predicción media, al intervalo de confianza de las predicciones y la alta exactitud del modelo.

El realizar un análisis de selección de características previo, permitió evidenciar que este influye de manera significativa en los resultados de la estimación en términos de disminución del error. Se comprobó que las características más relevantes se encuentran en las medidas directas realizadas sobre la raíz del diente, y no en las características indirectas que usa el método de Lamendin.

Si bien la técnica de Lamendin es universalmente aceptada en la odontología forense, su raíz del error medio cuadrático en la estimación de la edad biológica está por encima de los valores propuestos cuando se aplica a una población colombiana. Por el contrario, la propuesta de aprendizaje de máquina utilizando procesos gaussianos por modelos de regresión, redujo la raíz del error medio cuadrático (RMSE) a 3,37 años en la estimación. Los autores de esta investigación proponen considerar esta técnica alternativa como una herramienta que brinda soporte científico a la estimación de la edad biológica en adultos a los administradores de justicia, y en especial a los colombianos donde las técnicas tradicionales entregan un marcado error.

Agradecimientos

El presente trabajo de investigación fue realizado bajo la asesoría, de las peritas odontólogas María del Pilar Ramírez Plitt y Yadira Roa Ruiz, servidoras del Laboratorio de Identificación Humana de la Fiscalía General de la Nación, Seccional Risaralda, a quienes los autores brindan su más sincero agradecimiento por el tiempo y disposición para guiar en todo momento el desarrollo y evolución de esta investigación.

De igual forma, se agradece a la perita odontóloga Luz Dary Escobar, por facilitar la muestra dental de estudio del presente proyecto.

Referencias

- Morales, D., Betancourt, C. & Mesa, L., 2010. Cálculo de la incertidumbre de medida al equipo de movimiento rectilíneo uniforme del grupo de investigación dicoped. *Scientia et Technica*, Issue 46, pp. 206-210.
- Pretelt de la Vega, S., 2004. *Código de Procedimiento Penal*. [En línea] Available at: https://www.oas.org/juridico/mla/sp/col/sp_col-int-text-cpp-2005.html [Último acceso: Noviembre 2017].
- Prince, D. A. & Ubelaker, D. H., 2002. Application of Lamendin's adult dental aging Technique to a diverse skeletal sample. *Journal Forensic Science*, pp. 107-116.
- Willmott, C. & Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, Volumen 30, pp. 79-82.
- Alvarado Garcia, A. M. & Salazar Maya, A. M., 2014. Análisis del concepto de envejecimiento. *Gerokomos*, 25(2), pp. 57-62.
- Baviera, T., 2017. Técnicas para el análisis del sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength. *Revista Dígitos*, 1(3), pp. 33-50.
- Bedoya, O. & Bustamante, S., 2011. CNN-PROMOTER, NEW CONSENSUS PROMOTER PREDICTION PROGRAM BASED ON NEURAL NETWORKS. *Revista EIA*, 8(15), pp. 153-164.
- Drucker, H. y otros, 1997. Support vector machine. *Advances in neural information processing systems*, Enero, pp. 155-161.
- Foti, B. y otros, 2001. Limits of the Lamendin method in age determination. *Forensic Science International*, pp. 101-106.
- García Cambrónero, C. & Gómez Moreno, I., 2006. Algoritmos de aprendizaje: KNN & KMEANS. [Inteligencia en Redes de Telecomunicación, Universidad Carlos III de Madrid].
- Guyon, I. & Elisseeff, A., 2003. An Introduction to Variable and Feature Selection. *Journal of Machine Learning*, pp. 1157-1182.
- JCGM 100, 2008. Guide to the expression of uncertainty in measurement. En: J. C. f. i. Metrology, ed. *Evaluation of measurement data*. s.l.:s.n., pp. 36-37.
- Krenzer, U., 2005. Odontología forense. En: *Tomo VI Antropología Dental*. Guatemala: CAFCA, pp. 52-58.
- Krenzer, U., 2006. Cambio en las dientes. En: *Tomo III Estimación de la edad Osteológica en adultos*. Guatemala: CAFCA, pp. 14-16.
- Lamendin, H. y otros, 1992. A simple Technique for age estimation in adult corpses: The two criteria dental method. *Journal of forensic sciences*, pp. 1373-1379.
- Martínez Rodríguez, E., 2005. Errores frecuentes en la interpretación del coeficiente de determinación lineal. *Anuario jurídico y económico escurialense*, Issue 38, pp. 315-331.
- Pelaez Chávez, N., 2012. *Aprendizaje no supervisado y el algoritmo WAKE-SLEEP en redes neuronales*. s.l.:s.n.
- Suárez Ponce, D. G., 2014. Reconstrucción osteobiográfica. *Odontología Sanmarquina*, pp. 44-46.

- Tobón, I. & Cortés, J., 2018. Identificación de instrumentos musicales de cuerdas pulsadas de la región andina colombiana en solo, mediante técnicas de aprendizaje de máquina. *Revista EIA*, 15(30), pp. 177-193.
- Vapnik, V., Golowich, S. & Smola, A., 1997. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. *Advances in neural information processing systems*, pp. 281-287.
- Vilcapoma Guerra, H. J., 2012. Método dental modificado para la estimación de la edad en individuos adultos. *Odontología Sanmarquina*, pp. 27-30.
- Witten, I. H., Frank, E. & Hall, M. A., 2011. Cross Validation. En: *Data Mining Practical Machine Learning*. USA: Morgan Kaufmann, pp. 152-154.
- Witten, I. H., Frank, E. & Hall, M. A., 2011. Ensemble learning. En: *Data Mining Practical Machine Learning*. s.l.:Morgan Kaufmann, pp. 352-362.