
Estimación de la edad de la muerte basada en aprendizaje conjunto y regresión simbólica a partir de múltiples anotaciones

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 Este estudio aborda el problema de la estimación semiautomática de la edad
2 de la muerte a partir de la sínfisis púbica, una tarea crucial aunque compleja
3 en antropología forense. Su precisión depende directamente de la calidad del
4 etiquetado de distintos rasgos en el hueso púbico desarrollado por los forenses, el
5 cual está afectado por una incertidumbre inherente a su definición.

6 Dado que la interpretabilidad es un requisito obligatorio, proponemos un enfoque
7 en el que el diseño del modelo se basa en aprendizaje evolutivo, considerando la
8 programación genética para resolver el problema mediante regresión simbólica.
9 Además, se usa el aprendizaje conjunto para abordar los retos que plantean el
10 ruido, la incertidumbre y las anotaciones contradictorias inherentes a los datos
11 recogidos de múltiples sujetos. El aprendizaje conjunto proporciona un enfoque
12 eficaz para superar estos retos, ya que facilita la obtención de consenso mediante
13 la toma de decisiones y la fusión de información. De ahí que se formen comités
14 de observadores, compuestos por múltiples especialistas forenses con diferentes
15 niveles de experiencia y conocimiento que proporcionan anotaciones alternativas.
16 Se prueban varias configuraciones de modelos de conjunto que combinan dife-
17 rentes estimadores base y operadores de agregación para evaluar la precisión y la
18 fiabilidad de sus estimaciones de la edad de la muerte. Se comparan con modelos
19 entrenados con anotaciones individuales, mostrando una mejora en su rendimiento.
20 Los resultados obtenidos también ponen de relieve las ventajas de incorporar diver-
21 sas perspectivas para abordar las complejidades asociadas a la variabilidad humana
22 y las evaluaciones anatómicas.

23 1. Introduction

24 Estimar la edad tanto de los vivos como de los muertos es una tarea crucial en Antropología Forense
25 (AF) [1]. Específicamente, la estimación de la edad de la muerte implica el examen de rasgos
26 específicos como la apariencia, la morfología y los patrones de osificación presentes en los restos
27 óseos de los sujetos. La sínfisis púbica es un hueso ampliamente utilizado para esta tarea debido a su
28 alta fiabilidad. El método pionero fue propuesto en 1920 por Todd [2], y aunque hasta la fecha se han
29 propuesto muchos otros enfoques, sigue siendo una técnica relevante e influyente. La metodología
30 estándar implica el análisis visual de la morfología de la superficie ósea y la categorización (basada
31 en expertos humanos) de diferentes características o rasgos para determinar un rango de edad
32 aproximado [2] o un valor numérico concreto [3].

33 Por lo tanto, la estimación de la edad de la muerte plantea un reto complejo caracterizado por la
34 incertidumbre inherente y la susceptibilidad al sesgo humano, a menudo exacerbado por los pocos

35 datos disponibles. Esto ha aumentado recientemente el interés por el desarrollo de métodos semi [4,5]
36 o totalmente automáticos [6, 7] para ayudar al antropólogo forense. Es obligatorio obtener modelos
37 interpretables y explicables, ya que este tipo de métodos suelen aplicarse en entornos médico-legales
38 y los forenses deben confiar en ellos. Las aproximaciones más interpretables existentes en la literatura
39 siguen basándose en la observación manual del hueso para caracterizar los diferentes rasgos de
40 la sínfisis púbica implicados en la toma de decisiones. Sin embargo, esto se considera una tarea
41 difícil debido a la incertidumbre que rodea a las definiciones de los rasgos esqueléticos [8]. En
42 consecuencia, la metodología para estimar la edad de la muerte resulta ser muy sensible a la calidad
43 de las anotaciones humanas de las variables, que dependen sobre todo de la experiencia y los
44 conocimientos del especialista forense.

45 La regresión simbólica (RS) surge como un enfoque ideal para abordar las complejidades de las
46 anotaciones incoherentes. La capacidad de la RS para extraer expresiones matemáticas de los datos
47 le permite adaptarse y aprender de diversos conjuntos de anotaciones, capturando los patrones
48 subyacentes y modelando relaciones únicas ocultas en los datos. En este trabajo, se encarga a una
49 serie de especialistas forenses que examinen visualmente y extraigan una serie de rasgos óseos
50 siguiendo una metodología basada en el método de Gilbert y McKern [3], que mejora la propuesta
51 canónica de Todd estimando directamente un valor numérico para la edad de la muerte a partir de una
52 combinación de evaluaciones parciales realizadas sobre cada rasgo óseo del pubis. A continuación,
53 las anotaciones se utilizan para aprender un estimador de RS basado en la programación genética
54 (PG). La interpretabilidad inherente de la PG, derivada de su naturaleza simbólica, permite el diseño
55 de modelos de estimación de la edad de la muerte flexibles, precisos y semiautomáticos, explorando
56 un espacio de modelos más extenso [9].

57 En el contexto del aprendizaje automático (AA), el campo conocido como aprendizaje a partir de
58 multitudes (*learning from crowds*) suele implicar el aprovechamiento de entradas o anotaciones de
59 múltiples sujetos para realizar predicciones/clasificaciones o tomar decisiones. A menudo, estas
60 anotaciones son ruidosas o contradictorias, ya que sufren de sesgos humanos y dependen de la
61 dificultad de la muestra y el nivel de habilidad o competencia de cada observador [10]. Por lo tanto,
62 este campo de investigación tiene una relación directa con la incertidumbre inherente presente en el
63 proceso de etiquetado de los rasgos de la sínfisis púbica en nuestro dominio. Nuestra hipótesis es que
64 una aplicación adecuada del aprendizaje a partir de multitudes nos permitirá superar el problema del
65 etiquetado de rasgos de la sínfisis púbica en la estimación de la edad de la muerte, logrando modelos
66 más robustos a pesar de la posible existencia de anotaciones de baja calidad.

67 Aunque existen diferentes aproximaciones para tratar escenarios con múltiples anotadores, en este
68 trabajo seguimos un enfoque de fusión de información basado en aprendizaje conjunto. Utilizando
69 modelos de AA diseñados con PG [11], este estudio busca mejorar la robustez del método introdu-
70 ciendo comités de observadores [12], con el objetivo de generar modelos predictivos más efectivos.
71 El uso de algoritmos evolutivos (AEs) y aprendizaje conjunto es un tema activo de investigación [13],
72 impulsado por los beneficios reconocidos de los modelos de conjunto en términos de eficiencia
73 computacional, rendimiento predictivo y especialización del espacio de características. En particular,
74 este trabajo se centrará en el proceso de toma de decisiones para alcanzar un consenso. Para ello,
75 evaluaremos diferentes operadores de agregación para combinar los modelos base [14].

76 El resto de este trabajo se estructura de la siguiente forma. La sección 2 explora el problema de
77 estimación de la edad de la muerte y los fundamentos de la RS con PG. La sección 3 detalla las
78 particularidades de la muestra de datos y el procedimiento de anotación. También introduce la
79 metodología adoptada para abordar nuestro problema, describiendo la configuración algorítmica del
80 método de PG empleado y los diferentes diseños de modelos de conjunto para fusionar la información.
81 Por último, en la sección 4 se presentan los experimentos desarrollados y su análisis, y en la sección 5
82 se exponen algunas observaciones finales.

83 **2. Preliminares**

84 **2.1. Estimación de la edad de la muerte a partir de la sínfisis púbica**

85 La AF juega un papel decisivo en las investigaciones criminales y los procesos judiciales, espe-
86 cializándose en el análisis de restos óseos para la identificación humana. Mediante un proceso
87 preliminar conocido como perfilado biológico, esta disciplina proporciona información crítica sobre
88 características biológicas clave como la edad, el sexo, la ascendencia, la estatura y evidencias relacio-

89 nadas con la causa de la muerte. En particular, la edad de la muerte es un factor crucial que reduce
90 significativamente el alcance de las posibles coincidencias durante el proceso de identificación.

91 Los métodos empleados por los forenses para abordar el problema de la estimación de la edad de la
92 muerte son obsoletos [2, 3, 15]. En los últimos años, se ha producido un notable aumento del interés
93 en el campo de la AF por el desarrollo de métodos precisos, robustos y automáticos para la estimación
94 de la edad de la muerte. En particular, los avances recientes se han centrado en la automatización
95 completa del proceso, integrando la Visión por Ordenador y el AA. Esta forma de proceder reduce la
96 dependencia de los antropólogos forenses, a costa de dar lugar a métodos excesivamente intrincados
97 y opacos. En consecuencia, la mayoría de los enfoques contemporáneos se basan en la extracción
98 automática de características a partir de modelos óseos escaneados en 3D [6, 7]. Teniendo en cuenta
99 la limitada interpretabilidad de las metodologías recientes y su compleja aplicabilidad, existe una
100 búsqueda continua de enfoques semiautomáticos destinados a mitigar los sesgos.

101 En 2022, Gámez et al. [5] desarrollaron un enfoque de aprendizaje transparente basado en reglas
102 evolutivas para la estimación de la edad de la muerte a partir de la sínfisis púbica, incorporando el
103 conocimiento experto del método de Todd e imitando su modo de funcionamiento. El objetivo del
104 AA era abordar un problema de clasificación ordinal, que implicaba la asignación de una de las diez
105 fases de Todd que delimitan un rango de edad estimado para cada hueso púbico.

106 **2.2. El método de Gilbert-McKern para estimar la edad de la muerte con la sínfisis púbica**

107 En este estudio presentamos un nuevo método interpretable y semiautomático inspirado en el método
108 de Gilbert y McKern [3]. La propuesta de Gilbert y McKern pretendía mejorar el método de Todd con
109 dos modificaciones clave. A los rasgos del hueso púbico se les asignaba un valor numérico dentro
110 del rango $\{0, 5\}$, valores que luego se agregaban para estimar directamente la edad de la muerte
111 numérica. La propuesta consideraba tres zonas diferentes del hueso (componentes): la demi-facial
112 dorsal, la rampante ventral y el borde sinfisal de la sínfisis púbica.

113 La agregación de los valores asignados a cada componente daba como resultado un número direc-
114 tamente asociado a un valor de edad de la muerte. En un estudio posterior [16], McKern llegó a la
115 conclusión de que sus resultados superaban a los del método pionero de Todd.

116 **2.3. Regresión Simbólica con Programación Genética**

117 La RS es una forma de análisis de regresión que pretende representar la relación subyacente de los
118 datos sin conocimiento previo de la expresión matemática resultante. Una ventaja clave de la RS
119 reside en su imparcialidad, ya que no se ve afectada por los sesgos del diseñador del modelo ni por
120 las incertidumbres en el conocimiento del dominio. Muchos enfoques utilizan comúnmente EAs [17]
121 como potentes técnicas de optimización global para navegar por el intrincado espacio de búsqueda
122 de las expresiones matemáticas. La PG [18] se erige como un enfoque clásico que se alinea con los
123 principios básicos de los algoritmos genéticos, basándose en una gramática basada en árboles para
124 representar a los individuos. Así, los operadores genéticos se aplican tanto a los nodos del árbol como
125 a los subárboles, con las expresiones matemáticas codificadas como nodos del árbol y los operandos
126 representados como nodos terminales.

127 La RS y la PG son especialmente adecuadas para problemas en los que la interpretabilidad y
128 la transparencia son cruciales [9], como el actual. Estos enfoques destacan por proporcionar no
129 sólo modelos predictivos, sino también expresiones comprensibles para el ser humano que ofrecen
130 información sobre las relaciones subyacentes en los datos. Además, los modelos de PG resultantes
131 suelen ser más intuitivos y fáciles de aplicar que los complejos modelos de caja negra, lo que facilita
132 su aplicación práctica en los procedimientos forenses.

133 **3. Materiales y Métodos**

134 **3.1. Descripción del conjunto de datos**

135 El conjunto de datos empleado ha sido la colección de sínfisis públicas esqueléticas conservadas en el
136 laboratorio de Antropología Física de la Universidad de Granada (España). Los estudios de autopsia
137 se vienen recopilando desde 1991 conjuntamente con el Instituto de Medicina Legal y Ciencias
138 Forenses de Granada, dando lugar a uno de los mayores conjuntos de datos del mundo. De una

139 muestra total de 837 sujetos, se seleccionaron 497 para este estudio, filtrando aquellos casos con
 140 información ante-mortem poco fiable o condiciones de conservación inadecuadas. Dado que la edad
 141 de la muerte puede estimarse a partir de ambas lateralidades (o lados) de la sínfisis púbica, el número
 142 final de muestras es de 986, considerando sujetos en el rango de edad entre 18 y 60 años. Así pues, la
 143 variable de salida de nuestros modelos será directamente el valor numérico de la edad de la muerte
 144 asociado a cada sínfisis púbica, que se conoce en la muestra considerada.

Tabla 1: Rasgos característicos de la sínfisis púbica y valores categóricos asignados

Variable Name	Categorical Values (Numerical Values)	
x_1 Articular Face	Regular Porosity (1) Ridges and Grooves (3) Grooves Remains (5)	Ridges Formation (2) Grooves Shallow (4) No Grooves (6)
x_2 Irregular Porosity	Absence (1) Much (3)	Medium (2)
x_3 Upper Symphysial Ext.	Not Specified (1)	Defined (2)
x_4 Bony Nodule	Absent (1)	Present (2)
x_5 Lower Symphysial Ext.	Not Specified (1)	Defined (2)
x_6 Dorsal Margin	Absent (1)	Present (2)
x_7 Dorsal Plateau	Absent (1)	Present (2)
x_8 Ventral Bevel	Absent (1) Present (3)	In Process (2)
x_9 Ventral Margin	Absent (1) Formed Without Outgrowths (3) Formed With Recesses And Protrusions (5)	Partially Formed (2) Formed with Few Outgrowths (4)

145 Los aspectos morfológicos de la sínfisis púbica se analizaron a fondo en la contribución seminal
 146 de Todd [2]. Sin embargo, las descripciones originales eran demasiado genéricas y el proceso de
 147 inspección visual dependía en gran medida de la pericia del antropólogo forense. En un estudio
 148 reciente, Gámez et al. [5] sistematizaron estas descripciones en un intento de reducir el sesgo
 149 subjetivo del análisis. El atlas propuesto definía nueve rasgos asociados al desarrollo y a los cambios
 150 degenerativos de los huesos púbicos. Seguimos este estudio asociando los nueve rasgos (variables x_1
 151 a x_9) con dos o más valores categóricos, como se muestra en la Tabla 1. Luego transformamos los
 152 valores categóricos en valores numéricos, como se hace en el método de Gilbert-McKern [3], para
 153 diseñar métodos semiautomáticos de la edad de la muerte basados en RS.

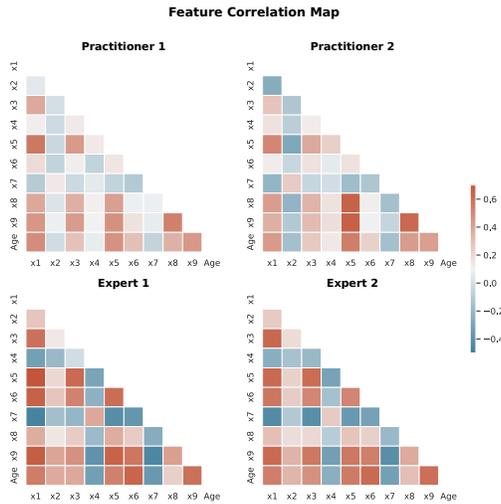


Figura 1: Análisis de correlación (coeficientes de Pearson) para los rasgos de la sínfisis púbica anotados por cada uno de los observadores.

154 La muestra fue anotada por cuatro especialistas en AF que etiquetaron a ciegas los nueve rasgos
 155 analizando visualmente cada sínfisis púbica en orden aleatorio. Dos de ellos eran expertos en la
 156 evaluación de los procesos de desarrollo de la sínfisis púbica, mientras que los otros dos eran jóvenes
 157 profesionales con un máster y formación en el proceso de análisis. En un intento de identificar las
 158 relaciones entre los datos etiquetados por cada especialista, en la Fig. 1 se analizan los coeficientes de

159 correlación de Pearson entre la edad real de la muerte de los sujetos y los rasgos anotados. Este análisis
 160 muestra claramente la incertidumbre y el sesgo subjetivo incurrido durante el proceso de evaluación.
 161 La Fig. 1 también pone de manifiesto el diferente nivel de competencia de los observadores.

162 Para evaluar el rendimiento de los métodos, los datos se partitionaron en conjuntos de entrenamiento
 163 y de prueba siguiendo una división 80/20, seleccionando la misma muestra para cada anotación de
 164 especialista (experto (E) o profesional (P)). La Fig. 2 ilustra la cantidad de muestras consideradas
 165 para entrenar los modelos de cada observador y la agregación de datos en bruto (enfoque de fusión de
 166 datos) para evaluar la robustez del enfoque según los niveles de especialización de los anotadores.

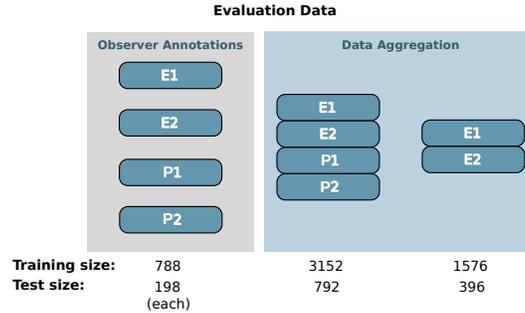


Figura 2: Enfoque de partición de datos. Bloque izquierda: datos utilizados para el aprendizaje de modelos base de los cuatro anotadores. Bloque derecha: agregación de las anotaciones *Todos* y (sólo) *Expertos* en un enfoque de fusión de datos mediante aprendizaje conjunto.

167 3.2. Diseño del método de Aprendizaje Automático basado en Programación Genética

168 En esta sección describimos las características específicas del diseño algorítmico de PG como modelos
 169 de regresión para identificar relaciones funcionales entre los rasgos óseos anotados [11]. La PG
 170 comparte los conceptos principales de los algoritmos genéticos, operando sobre una población de
 171 soluciones candidatas que representan expresiones matemáticas potenciales. La población se inicializa
 172 con soluciones aleatorias de tamaño pequeño y estas soluciones evolucionan iterativamente utilizando
 173 operadores genéticos como el cruce y la mutación.

174 *Esquema de codificación:* Las expresiones matemáticas se codifican en árboles de expresión, cuyos
 175 nodos terminales son una de las variables que representan los nueve rasgos identificados (x_1 a
 176 x_9 , véase la Tabla 1) o una constante numérica. Los nodos internos del árbol son los operadores
 177 matemáticos suma, resta, multiplicación, división, negación e inversa. Se seleccionó un conjunto
 178 de operadores sencillos para garantizar que los modelos obtenidos fueran interpretables por los
 179 antropólogos forenses. El resultado de la expresión matemática es el valor predicho de la edad de la
 180 muerte para los rasgos de la sínfisis púbica especificados como entradas.

181 *Mecanismo de selección:* La PG sigue el esquema generacional clásico tanto para los procedimientos
 182 de selección como de reemplazo. Se crea una población intermedia mediante selección por torneo y
 183 la población descendiente sustituye directamente a la población de padres considerando el elitismo.
 184 Se eligió un torneo de gran tamaño para mantener una alta presión selectiva y, al mismo tiempo,
 185 preservar la diversidad en la población. Como se sabe que la PG muestra una convergencia lenta,
 186 consideramos tanto un gran tamaño de población como un gran tamaño de torneo para aumentar la
 187 capacidad de explotación (véase la Tabla 2).

188 *Operadores genéticos:* Se considera el cruce GP habitual [18], donde se selecciona una arista
 189 aleatoria en cada padre, intercambiando ambos subárboles para formar un descendiente en la siguiente
 190 generación. Para mantener la diversidad se utilizan dos operadores de mutación diferentes: a) mutación
 191 de subárbol, selección aleatoria de una arista y generación aleatoria de un nuevo subárbol que sustituye
 192 al antiguo situado en esa arista, y b) mutación puntual, cambio aleatorio del valor de un nodo por otro
 193 del mismo tipo: bien un valor numérico aleatorio o bien otra variable/operador.

194 *Función de fitness:* Se basa en el error cuadrático medio (ECM) de la expresión codificada:

$$ECM = \sum_{i=1}^N (y_i - y'_i)^2 \quad ,$$

Tabla 2: Hiperparámetros de la PG después del ajuste

Parameter	Value
Número de generaciones	1000
Tamaño de población	1000
Tamaño del torneo	80
Probabilidad de cruce	0.75
Probabilidad de mutación de subárbol	0.11
Probabilidad de mutación puntual	0.05
Probabilidad de mutación de Hoist	0.08
Coefficiente de parsimonia	0.05

195 donde y_i es la edad de la muerte del ejemplo i e y'_i es el valor predicho por el modelo.

196 *Bloating*: El *bloat* es una limitación bien conocida de la PG, que provoca a un crecimiento excesivo
 197 en términos de tamaño de los individuos. El *bloat* conduce a un sobreajuste de los datos y a mayores
 198 costes de evaluación debido a la profundidad variable de los árboles. La presión de parsimonia [18, 19]
 199 es probablemente el método de control más frecuente, que penaliza explícitamente los programas
 200 más grandes disminuyendo su fitness proporcionalmente a su tamaño. También consideramos dos
 201 alternativas adicionales: i) el enfoque dinámico para calcular el coeficiente de parsimonia de acuerdo
 202 con la varianza de los tamaños de los programas dentro de la población en una generación introducida
 203 por Poli et al. [20]; y ii) un tercer tipo de mutación, la mutación *hoist* [21], que elimina material
 204 genético de la solución padre seleccionando y eliminando aleatoriamente un subárbol.

205 *Ajuste de parámetros*: Los hiperparámetros se ajustaron sobre un subconjunto de los datos donde otro
 206 experto proporcionó un conjunto adicional de anotaciones. En concreto, se consideró el estimador
 207 Parzen estructurado en árbol [22] como algoritmo de búsqueda por su buen rendimiento en la
 208 optimización de hiperparámetros frente a otros enfoques como la búsqueda aleatoria o en grid. Se
 209 basa en la optimización bayesiana, utilizando un modelo probabilístico para encontrar los parámetros
 210 óptimos de acuerdo con una función objetivo. Se siguió una validación cruzada de 5 subconjuntos
 211 durante la fase de ajuste y la puntuación de validación cruzada se utilizó para guiar la optimización.
 212 La Tabla 2 recoge la configuración final de parámetros obtenida.

213 3.3. Diseño de los modelos de conjunto

214 Dada la complejidad y la incertidumbre inherentes al problema, consideramos el concepto de comité
 215 de especialistas con experiencia variada como medio para alcanzar una solución mejorada con
 216 respecto a la obtenida por un único profesional. Este concepto está bien establecido en Inteligencia
 217 Artificial en forma de aprendizaje conjunto [12], que permite la combinación estratégica de múltiples
 218 modelos para resolver un problema de ML mejorando su rendimiento individual.

219 Concretamente, seguiremos el enfoque habitual de tres etapas [13] para la fusión de información a
 220 través de comités basados en modelos independientes, también conocidos como modelos base (véase
 221 la Fig. 3). En primer lugar, consideramos una etapa de generación basada en datos. Se entrenarán
 222 cuatro modelos base mediante PG en todo el espacio de características utilizando cada conjunto de
 223 anotaciones por separado. Para establecer una referencia inferior, compararemos su rendimiento con
 224 los modelos entrenados utilizando datos agregados (es decir, todo el conjunto de datos compuesto por
 225 cada anotación de los cuatro especialistas o sólo de los dos expertos), como se muestra en la Fig. 2.

226 La segunda etapa corresponde a la selección o poda del modelo [23]. Para evaluar el impacto del
 227 nivel de pericia en la anotación visual de los rasgos óseos, seguiremos dos enfoques: sin poda y poda
 228 por pericia. En este último caso, en la tercera etapa sólo se tienen en cuenta los modelos generados a
 229 partir de las observaciones de los anotadores expertos.

230 La última etapa es la de consenso o toma de decisiones, que se consigue combinando las salidas de
 231 los modelos base. La fusión de modelos suele hacerse con operadores de agregación como el máximo,
 232 la media aritmética o el voto mayoritario [23]. Aunque estos operadores son más eficaces cuando los
 233 errores de los modelos base son independientes, esta condición no suele estar garantizada [24].

234 Proponemos cuatro enfoques distintos para la etapa de toma de decisiones del modelo de conjunto
 235 considerando diferentes funciones de agregación [14]:

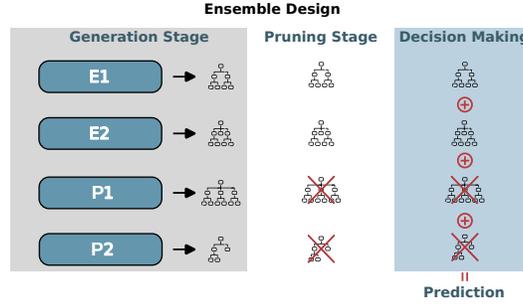


Figura 3: Representación de la metodología de diseño de modelos de conjunto en tres etapas.

- 236 ■ Media simple (MS): Como aproximación naive, los modelos base de RS (M_i) se combinan
 237 utilizando el operador de media aritmética:

$$MS = \frac{1}{n} \sum_{i=1}^n M_i$$

- 238 ■ Media ponderada (MP): Para considerar el nivel de experiencia de los observadores, se
 239 asignan pesos manualmente para potenciar la contribución de los modelos base entrenados
 240 sobre anotaciones realizadas por expertos. Así, a los *modelos de expertos* se les asigna el
 241 doble de peso que a los *modelos de jóvenes profesionales*.

$$MP = \sum_{i=1}^n w_i M_i, \text{ where } w_{E_i} = 2 \cdot w_{P_i}, \sum_{i=1}^n w_i = 1$$

- 242 ■ Media ponderada ordenada (MPO): Una ligera variación consiste en ordenar cada modelo
 243 en función de su rendimiento. En lugar de asignar manualmente las ponderaciones, en este
 244 enfoque las ponderaciones se asignan proporcionalmente según el coeficiente de determina-
 245 ción de entrenamiento del modelo. R^2 relaciona la suma de residuos al cuadrado (SS_{res}),
 246 calculada como las diferencias entre los valores observados y predichos, con la suma total de
 247 cuadrados (SS_{tot}), calculada como las diferencias entre los valores observados y la media
 248 de los valores observados:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

- 249 ■ Agregación fuzzy (AGF): Las agregaciones fuzzy son adecuadas para extraer conocimiento
 250 y hacer frente a la incertidumbre del modelo. Algunas de las más populares son la integral de
 251 Choquet, la integral de Sugeno y la agregación ponderada ordenada (OWA) [14], considerada
 252 en este estudio. La OWA se diferencia de los enfoques anteriores por realizar una agregación
 253 ponderada en la que los modelos base M_i se ordenan descendientemente por rendimiento y
 254 los pesos se asocian según ese orden de la siguiente forma:

$$w_i = Q_{a,b} \left(\frac{i}{n} \right) - Q_{a,b} \left(\frac{i-1}{n} \right), \text{ donde}$$

$$Q_{a,b}(x) = \begin{cases} 0, & \text{if } x < a \\ 1, & \text{if } x > b \\ \frac{x-a}{b-a}, & \text{otherwise} \end{cases}$$

256 Los vectores de peso para la función OWA se obtuvieron empleando los parámetros $a = 0.3$
 257 y $b = 0.8$.

258 4. Experimentos y Análisis de Resultados

259 En esta sección realizamos un análisis de rendimiento de los cuatro modelos base, cada uno entrenado
 260 con las anotaciones específicas de cada especialista. Comparamos sus resultados con modelos entre-
 261 nados con anotaciones combinadas siguiendo un enfoque de fusión de información. La Tabla 3 recoge

Tabla 3: Resultados de modelos de referencia y aprendizaje conjunto al estimar la edad de la muerte.

Modelo		EAM	ERCM	
Modelos Base	E1	6.61	8.15	
	<i>E2</i>	<i>6.19</i>	<i>7.76</i>	
	P1	7.69	9.20	
	P2	7.39	8.91	
Agregación de Datos		Todos	7.05	
		<i>Expertos</i>	<i>6.39</i>	
Modelos de Conjunto	Sin Poda	MS	6.56	7.82
		MP	6.35	7.65
		<i>MPO</i>	<i>6.27</i>	<i>7.59</i>
		AGF	6.43	7.76
	Con Poda	MS	6.14	7.54
		MP	6.11	7.52
		MPO	6.08	7.52
		AGF	6.11	7.53

262 los resultados obtenidos por cada modelo en la experimentación para dos medidas de evaluación: el
 263 error absoluto medio (EAM) y el error de raíz cuadrada media (ERCM), diferentes de la función de
 264 fitness utilizada para el proceso de aprendizaje, el ECM (véase la Sec. 3.2).

265 Los resultados de los modelos base subrayan la importancia de los niveles de competencia de los
 266 antropólogos forenses. Aunque el sesgo subjetivo sigue presente en el proceso de anotación de rasgos,
 267 los modelos entrenados con etiquetas de expertos superaron a los basados en etiquetas de jóvenes
 268 profesionales en aproximadamente 1 año tanto en EAM como en ERCM. El mejor resultado es el del
 269 modelo diseñado con las anotaciones del segundo experto (E2, resaltado en cursiva en la tabla).

270 Esas diferencias se propagan también a los modelos que consideran un enfoque de fusión de informa-
 271 ción, mostrando un rendimiento cercano a la media de los cuatro modelos base. La incorporación de
 272 las anotaciones de los expertos ayuda a mejorar el rendimiento de los modelos base de los profesio-
 273 nales, mientras que la consideración conjunta de las anotaciones de ambos expertos no mejora los
 274 resultados individuales del modelo base E2.

275 Analizando los modelos de conjunto que consideran todos los modelos base generados (es decir, sin
 276 estrategia de poda). En general, las cuatro variantes presentan un rendimiento competitivo, destacando
 277 la MPO como la agregación con el mejor rendimiento (en cursiva). Sin embargo, es el único modelo
 278 de conjunto no podado capaz de superar al mejor modelo base (E2) y sólo en la medida ERCM.

279 Concluimos que la incertidumbre y las anotaciones potencialmente erróneas de los observadores
 280 profesionales (con nivel de máster) penalizan el rendimiento de los modelos generados. Confirmamos
 281 esta suposición mediante los modelos de conjunto podados que solo consideran las anotaciones de los
 282 dos expertos, los cuales fueron capaces de proporcionar resultados robustos, mejorando los modelos
 283 base, independientemente del operador de agregación considerado. De nuevo, la MPO destaca como
 284 el mejor enfoque al obtener los errores de predicción más bajos, EAM=6.08 años y ERCM=7.52
 285 años, que constituyen un estado del arte ya que [5] obtuvo MAE=13.19 y ERCM=10.38 en error de
 286 prueba para un conjunto de pubis con el mismo rango de edad (18-60 años) y [4] obtuvo EAM=12.1
 287 y ERCM=9.7 en error de entrenamiento para un rango de edad de 19-100 años.

288 La Fig. 4 muestra la distribución de las estimaciones de los cuatro modelos base y los dos modelos de
 289 conjunto con agregación MPO (podados y no podados). Identificamos un patrón claro: se sobreestima
 290 la edad en los sujetos jóvenes y se subestima en los sujetos mayores. Este comportamiento es
 291 un problema bien conocido en el área [25]. Los resultados demuestran la relevancia del nivel de
 292 competencia y la fiabilidad de las anotaciones de los especialistas forenses. Las diferencias entre
 293 observadores son más acusadas en el caso de los más novatos, dado que los modelos entrenados con
 294 las etiquetas de los expertos tienden a proporcionar resultados más precisos y una menor variabilidad,
 295 aunque no existe un consenso aparente entre ambos expertos. Las etiquetas inexactas penalizan
 296 fuertemente las predicciones conjuntas, especialmente en el menor rango de edad (18-23 años). No
 297 obstante, la fusión de las predicciones de los expertos ha sido un enfoque acertado, que permitió al
 298 modelo de conjunto ofrecer estimaciones más precisas que los modelos individuales.

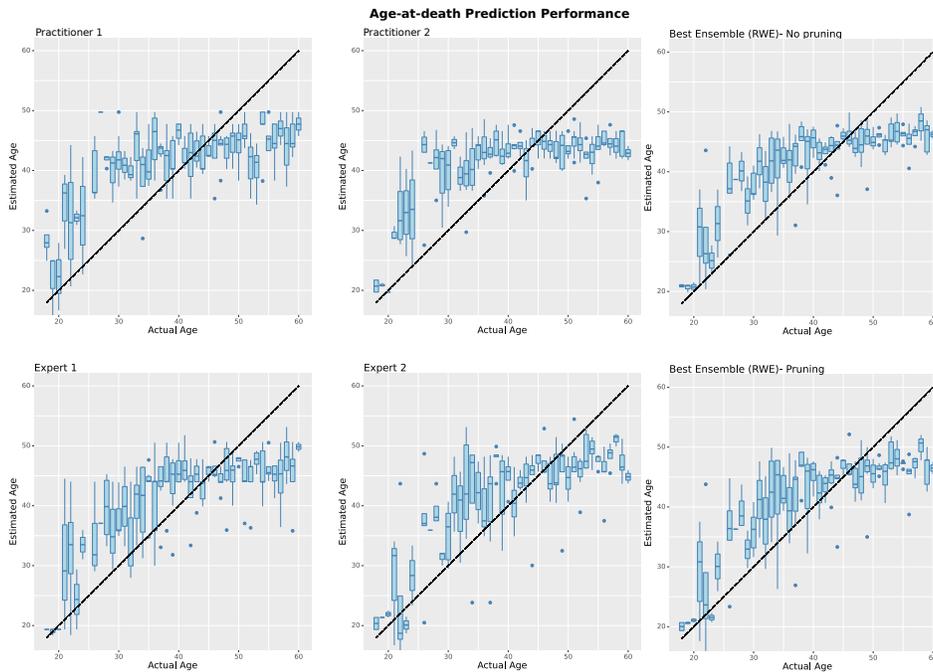


Figura 4: Análisis visual del rendimiento de varios modelos. Cada boxplot muestra el error de estimación para cada edad. La línea de puntos negra es la referencia de la estimación real.

299 5. Conclusiones

300 En este artículo hemos explorado un enfoque de fusión de información de múltiples anotadores
 301 de características en una tarea compleja de ML caracterizada por incertidumbre y sesgo subjetivo.
 302 Para abordar el problema de la estimación de la edad de la muerte a partir de la sínfisis púbica
 303 adoptamos un enfoque de RS utilizando PG. Este EA produce expresiones matemáticas interpretables
 304 y fácilmente aplicables, asistiendo a los antropólogos forenses en el análisis.

305 Hemos considerado el aprendizaje conjunto y la toma de decisiones empleando diferentes operadores
 306 de agregación, logrando una mejora del rendimiento y aumentando la robustez de las predicciones en
 307 comparación con los modelos individuales. Siguiendo este enfoque, conseguimos tratar adecuada-
 308 mente la incertidumbre inherente al proceso de anotación de muestras por parte de los antropólogos
 309 forenses, necesario para el diseño del método semiautomático de estimación de la edad de la muerte.

310 Obtuvimos dos conclusiones relevantes. El sesgo subjetivo de la evaluación visual para determinar
 311 los rasgos óseos está estrechamente relacionado con la experiencia de los especialistas humanos. Así,
 312 puede dar lugar a anotaciones poco fiables, que dificultan la agregación de datos de múltiples especia-
 313 listas. No obstante, el uso de modelos de conjunto nos permitió armonizar diversas observaciones de
 314 antropólogos forenses muy competentes, mejorando el rendimiento de los modelos base.

315 Agradecimientos

316 Esta investigación ha sido subvencionada por el proyecto CONFIA (PID2021-122916NB-I00) finan-
 317 ciado por MCIN/AEI/10.13039/501100011033 y por “FEDER: Una manera de hacer Europa”.

318 Referencias

- 319 [1] E. Cunha, E. Baccino, L. Martrille, F. Ramsthaler, J. Prieto, Y. Schuliar, N. Lynnerup, and C. Cattaneo,
 320 “The problem of aging human remains and living individuals: A review,” *Forensic Science International*,
 321 vol. 193, no. 1, pp. 1–13, 2009.
- 322 [2] T. W. Todd, “Age changes in the pubic bone,” *American Journal of Physical Anthropology*, vol. 3, no. 3,
 323 pp. 285–328, 1920.

- 324 [3] B. Gilbert and T. W. McKern, "A method for aging the female os pubis," *American Journal of Physical*
325 *Anthropology*, vol. 38, no. 1, pp. 31–38, 1973.
- 326 [4] A. Kotěrová, D. Navega, M. Štepanovský, Z. Buk, J. Brůžek, and E. Cunha, "Age estimation of adult
327 human remains from hip bones using advanced methods," *Forensic Science International*, vol. 287, pp.
328 163–175, 2018.
- 329 [5] J. C. Gámez, J. Irurita, A. González, S. Damas, I. Alemán, and O. Cordón, "Automating the decision
330 making process of Todd's age estimation method from the pubic symphysis with explainable machine
331 learning," *Information Sciences*, vol. 612, pp. 514–535, 2022.
- 332 [6] D. Slice and B. Algee-Hewitt, "Modeling bone surface morphology: A fully quantitative method for
333 age-at-death estimation using the pubic symphysis," *Journal of Forensic Sciences*, vol. 60, no. 4, pp.
334 835–843, 2015.
- 335 [7] A. Kotěrová, M. Štepanovský, Z. Buk, J. Brůžek, N. Techataweewan, and J. Velemínská, "The compu-
336 tational age-at-death estimation from 3d surface models of the adult pubic symphysis using data mining
337 methods," *Scientific Reports*, vol. 12, no. 1, p. 10324, Jun 2022.
- 338 [8] B. Dudzik and N. Langley, "Estimating age from the pubic symphysis: A new component-based system,"
339 *Forensic Science International*, vol. 257, pp. 98–105, 2015.
- 340 [9] Y. Mei, Q. Chen, A. Lensen, B. Xue, and M. Zhang, "Explainable artificial intelligence by genetic
341 programming: A survey," *IEEE Transactions on Evolutionary Computation*, vol. 27, no. 3, pp. 621–641,
342 2022.
- 343 [10] D. Karger, S. Oh, and D. Shah, "Iterative learning for reliable crowdsourcing systems," in *Advances in Neu-
344 ral Information Processing Systems*, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger,
345 Eds., vol. 24. Curran Associates, Inc., 2011.
- 346 [11] E. Bermejo, A. D. Villegas, J. Irurita, S. Damas, and O. Cordón, "Interpretable machine learning for
347 age-at-death estimation from the pubic symphysis," *Expert Systems*, vol. 42, p. e70021, 2025.
- 348 [12] B. Barua and E. Corchado, *The Committee of Experts Approach: Ensemble Learning*. Berlin, Heidelberg:
349 Springer Berlin Heidelberg, 2011, pp. 31–47.
- 350 [13] H. E. L. Cagnini, S. C. N. Das Dôres, A. A. Freitas, and R. C. Barros, "A survey of evolutionary algorithms
351 for supervised ensemble learning," *The Knowledge Engineering Review*, vol. 38, p. e1, 2023.
- 352 [14] G. Beliakov, H. B. Sola, and T. C. Sánchez, *A practical guide to averaging functions*. Springer, 2016, vol.
353 329.
- 354 [15] S. Brooks and J. M. Suchey, "Skeletal age determination based on the os pubis: A comparison of the
355 acsádi-nemeskéri and suchey-brooks methods," *Human Evolution*, vol. 5, no. 3, pp. 227–238, 1990.
- 356 [16] T. McKern, "Sexual dimorphism in the maturation of the human public symphysis," in *The Measures of*
357 *Man*. Cambridge: Harvard University Press, 1976, pp. 433–450.
- 358 [17] J. E. Smith and A. E. Eiben, *Introduction to Evolutionary Computing*. Heidelberg: Springer Berlin, 2010.
- 359 [18] J. R. Koza, *Genetic programming: on the programming of computers by means of natural selection*.
360 Cambridge: MIT Press, 1992.
- 361 [19] B.-T. Zhang and H. Mühlenbein, "Balancing Accuracy and Parsimony in Genetic Programming," *Evolutionary*
362 *Computation*, vol. 3, no. 1, pp. 17–38, 03 1995.
- 363 [20] R. Poli and N. McPhee, "Parsimony pressure made easy," in *GECCO'08: Proceedings of the 10th Annual*
364 *Conference on Genetic and Evolutionary Computation*. Atlanta, GA, USA: ACM, 2008, pp. 1267–1274.
- 365 [21] K. Kinneer, "Fitness landscapes and difficulty in genetic programming," in *Proceedings of the First IEEE*
366 *Conference on Evolutionary Computation (CEC)*. *IEEE World Congress on Computational Intelligence*
367 *(WCCI)*, vol. 1, 1994, pp. 142–147.
- 368 [22] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in
369 *Advances in Neural Information Processing Systems*, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and
370 K. Weinberger, Eds., vol. 24. Curran Associates, Inc., 2011.
- 371 [23] D. Ruta and B. Gabrys, "Classifier selection for majority voting," *Information Fusion*, vol. 6, no. 1, pp.
372 63–81, 2005.
- 373 [24] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship
374 with the ensemble accuracy," *Machine Learning*, vol. 51, pp. 181–207, 2003.
- 375 [25] R. G. Aykroyd, D. Lucy, A. M. Pollard, and T. Solheim, "Technical note: regression analysis in adult age
376 estimation," *American Journal of Physical Anthropology*, vol. 104, no. 2, pp. 259–265, Oct. 1997.