



UNIVERSIDAD DE GRANADA

TRABAJO FIN DE GRADO GRADO EN INGENIERÍA INFORMÁTICA

Estimación de la edad a partir de modelos 3D de la sínfisis pública por medio de deep learning

Autor
Alejandro Manzanares Lemus

Directores
Óscar Cordón García
Pablo Mesejo Santiago



**ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN**

Granada, 6 de julio de 2022

Estimación de la edad a partir de modelos 3D de la sínfisis pélvica por medio de deep learning

Alejandro Manzanares Lemus

Palabras clave: Aprendizaje automático, aprendizaje profundo, visión por computador, antropología forense, estimación del perfil biológico, estimación de la edad, regresión, proyección panorámica

Resumen

La estimación de la edad es una de las tareas más relevantes dentro del campo de la antropología forense y resulta de utilidad tanto para la identificación de personas vivas como muertas en problemáticas tan distintas como desapariciones de personas, crisis migratorias, catástrofes naturales, crímenes sin resolver, etc. Entre las diversas técnicas existentes para estimar la edad de la muerte en individuos fallecidos, destacan el estudio de las características morfológicas presentes en la sínfisis del pubis, método propuesto por T. W. Todd en 1921 y utilizado aún a día de hoy. Este método presenta un gran nivel de subjetividad, ya que el estudio de estas características depende casi en su totalidad de la experiencia del antropólogo forense. Destaca por tanto, el bajo nivel de aplicación de técnicas automáticas existentes para el apoyo de esta tarea.

Este TFG aborda el desarrollo de un sistema capaz de estimar la edad de un individuo a partir de un modelo 3D de su sínfisis pélvica. Nuestra propuesta presenta un enfoque totalmente nuevo y transgresor respecto a las técnicas actuales. A diferencia de estas, donde es necesaria la participación de un antropólogo forense experto para extraer los rasgos morfológicos de las muestras, el método que proponemos será capaz de extraer estos rasgos de forma automática. Conseguimos este objetivo enfocando este problema dentro del paradigma del aprendizaje profundo. Para poder aplicar técnicas relativas a este campo utilizamos un método de proyección panorámica que permite transformar los modelos 3D en imágenes 2D manteniendo la información morfológica de los mismos. Para entrenar los modelos profundos convolucionales propuestos utilizamos el conjunto de 1104 modelos 3D de sínfisis pélvicas propiedad de la Universidad de Granada. Este conjunto presenta las limitaciones de ser demasiado limitado en número y poseer clases altamente desbalanceadas, por lo que hemos aplicado estrategias de aumento y balanceo de datos.

Los resultados obtenidos por el método propuesto son comparables a los del estado del arte actual, por lo que el modelo obtenido es altamente competitivo y presenta una alternativa viable para el apoyo en la estimación de la edad de un individuo.

Age estimation from 3D models of the pubic symphysis using deep learning

Alejandro Manzanares Lemus

Keywords: Machine learning, deep learning, computer vision, forensic anthropology, biological profile estimation, age estimation, regression, panoramic projection

Abstract

Age estimation is one of the most relevant tasks in the field of forensic anthropology and is useful both for the identification of living and dead persons in problems as different as disappearances of persons, migratory crises, natural catastrophes, unsolved crimes, etc. Among the various existing techniques for estimating the age of death in deceased individuals, the study of the morphological traits present in the pubic symphysis, a method proposed by T. W. Todd in 1921 and still used today, stands out. This method presents a high level of subjectivity, since the study of these characteristics depends almost entirely on the experience of the forensic anthropologist. Therefore, the low level of application of existing automatic techniques to support this task stands out.

This TFG addresses the development of a system capable of estimating the age of an individual from a 3D model of the pubic symphysis. Our proposal presents a totally new and transgressive approach with respect to current techniques. Unlike these, where the participation of an expert forensic anthropologist is necessary to extract the morphological features of the samples, the method we propose will be able to extract these features automatically. We achieve this goal by approaching this problem within the deep learning paradigm. In order to apply techniques related to this field we use a panoramic projection method that allows us to transform 3D models into 2D images while maintaining the morphological information of the models. To train the proposed deep convolutional models we used the set of 1104 3D pubic symphysis models owned by the University of Granada. This set has the limitations of being too limited in number and having highly unbalanced classes, so we have applied data augmentation and balancing strategies.

The results obtained by the proposed method are comparable to those of the current state of the art, so the obtained model is highly competitive and presents a viable alternative to support the age estimation of an individual.

Yo, **Alejandro Manzanares Lemus**, alumno de la titulación Ingeniería Informática de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicaciónde la Universidad de Granada**, con DNI 77393031D, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Alejandro Manzanares Lemus

Granada a 6 de julio de 2022.

D. **Óscar Cordón García**, Profesor del Área de Ciencias de la Computación e Inteligencia Artificial del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada.

D. **Pablo Mesejo Santiago**, Profesor del Área de Ciencias de la Computación e Inteligencia Artificial del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada.

Informan:

Que el presente trabajo, titulado *Estimación de la edad a partir de modelos 3D de la sínfisis pélvica por medio de deep learning* , ha sido realizado bajo su supervisión por **Alejandro Manzanares Lemus**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a 6 de julio de 2022.

Los directores:

Óscar Cordón García Pablo Mesejo Santiago

Agradecimientos

En primer lugar, me gustaría agradecer a mis tutores Óscar Cordón García y Pablo Mesejo Santiago, por darme la oportunidad de trabajar en este proyecto, orientarme en su desarrollo, ayudarme a la hora de resolver todas las dudas y problemas que han surgido y por su paciencia y dedicación a la hora de desarrollar este informe.

Me gustaría también agradecer a todos los compañeros, amigos, que me han acompañado durante mi paso por la ETSIIT, tanto a los que he conocido estos años como a los que ya tenía el placer de conocer. Siempre he encontrado en ellos un apoyo para seguir adelante y su presencia ha hecho de estos años algo inolvidable para mí. Quiero hacer especial mención a mis compañeros de la DEIIT, cuyo apoyo ha sido fundamental para poder dedicarme a desarrollar este TFG mientras ejercía mi cargo como Delegado de Centro. También me gustaría agradecer a todos los profesores que han sido capaces de sacar lo mejor de sus estudiantes y transmitir su pasión por lo que enseñan, animándonos a todos nosotros a continuar adelante.

Finalmente, quiero agradecer a mis padres todo el apoyo, la ayuda y la comprensión que me han brindado mientras desarrollaba este proyecto y durante todo mi paso por el grado.

Índice general

1. Introducción	19
1.1. Descripción del problema	19
1.2. Motivación	24
1.3. Objetivos	26
2. Fundamentos teóricos	29
2.1. Machine learning	29
2.2. Deep learning	31
2.2.1. Redes neuronales	31
2.2.2. Redes neuronales convolucionales	33
2.3. Tipos de representación de datos en 3D	38
2.3.1. Datos en bruto	39
2.3.2. Sólidos	41
2.3.3. Superficies	41
2.3.4. Estructuras de alto nivel	42
2.3.5. Datos de múltiples vistas	44
3. Estado del arte	47
3.1. Estimación del perfil biológico	47
3.1.1. Estimación de la edad a partir de restos óseos	48
3.1.2. Estimación de la edad a partir de la sínfisis pública . .	48
3.2. Representación de datos 3D	53
3.2.1. Representaciones basadas en proyecciones 3D	55
4. Métodos propuestos	57
4.1. Representación PANORAMA	57
4.1.1. Propuesta original de PANORAMA	57
4.1.2. Propuesta de adaptación	61
4.2. Arquitecturas de red	65
4.3. Deep imbalanced regression	69
5. Planificación e implementación	71
5.1. Planificación del proyecto	71
5.2. Detalles técnicos de la implementación	72

5.2.1. Implementación de la representación PANORAMA	72
5.2.2. Entrenamiento de los modelos	75
6. Experimentación	77
6.1. Conjunto de datos	77
6.1.1. Aumento de datos	79
6.1.2. Balanceo de datos	82
6.2. Métricas y protocolo de validación experimental	83
6.2.1. Métricas de calidad consideradas	83
6.2.2. Hiperparámetros utilizados	83
6.2.3. Protocolo de entrenamiento	84
6.2.4. Evolución del entrenamiento de los modelos	86
6.3. Resultados	90
6.3.1. Resultados totales	90
6.3.2. Resultados divididos en niños/adultos/ancianos	91
6.3.3. Resultados divididos en los rangos de Todd	92
6.4. Análisis de los resultados	94
6.5. Comparación con el estado del arte	95
7. Conclusiones	97
7.1. Trabajos futuros	99
Bibliografía	107
A. Tablas de resultados por ejecución	109
A.1. Resultados divididos en niños/adultos/ancianos	109
A.2. Resultados divididos en los rangos de Todd	112

Índice de figuras

1.1.	Diagrama de secuencia de los pasos realizados en la identificación forense basada en el esqueleto.	21
1.2.	Reconstrucción volumétrica por tomografía computerizada 3D de la síntesis del pubis.	21
1.3.	Características morfológicas en la síntesis pública.	22
1.4.	Representación del modelo a desarrollar.	26
2.1.	Ejemplo de red neuronal superficial y profunda.	32
2.2.	Procesamiento de señal a nivel de nodo.	32
2.3.	Operación de convolución.	34
2.4.	Movimiento del kernel sobre la imagen.	35
2.5.	Tipos de pooling.	36
2.6.	Arquitectura de CNN.	37
2.7.	Ejemplo de aplicación de dropout.	38
2.8.	Diferentes formas de representación de los datos 3D.	45
3.1.	Número de publicaciones relacionadas con la estimación de la edad y la AF por año.	48
4.1.	Cilindro utilizado para adquirir la proyección de un modelo 3D.	58
4.2.	Discretización de la superficie lateral del cilindro de proyección.	59
4.3.	Sección transversal más alta del cilindro junto con los rayos correspondientes.	60
4.4.	Ejemplo de la representación propuesta.	61
4.5.	Limitaciones de la representación propuesta.	61
4.6.	Ejemplo de vista panorámica aumentada de un modelo 3D.	62
4.7.	Ejemplo de imagen RGB (SDM-NDM-GNDM).	63
4.8.	Primera propuesta de representación.	63
4.9.	Segunda propuesta de representación.	64
4.10.	Red Panorama-CNN con entrada de imágenes RGB.	66
4.11.	Red Panorama-CNN con entrada de mapas de características.	66
4.12.	Esquema del funcionamiento de la técnica de transfer learning.	67
4.13.	Red Resnet-CNN.	67
4.14.	Ejemplo de la arquitectura ResNet34 para Imagenet.	68

4.15. Esquema de aplicación de LDS.	69
5.1. Grafo de llamadas de <code>calculate_panorama()</code>	73
5.2. Grafo de llamadas de <code>combine_panorama()</code>	73
5.3. Grafo de llamadas de <code>concat_panorama()</code>	74
5.4. Diagrama de la clase Mesh3D.	74
6.1. Numero de muestras Izquierda y Derecha.	77
6.2. Histograma de la edad de las muestras de síntesis públicas. . . .	78
6.3. Histograma de la edad de las muestras de síntesis públicas por rangos de Todd.	78
6.4. Proyecciones en el eje Y de una muestra del conjunto de datos después de aplicar la técnica de aumento de datos.	81
6.5. Histograma del conjunto de datos una vez aplicado LDS. . . .	82
6.6. Esquema de la división del conjunto de datos.	85
6.7. Histograma de la edad de las muestras de síntesis públicas divididas en conjuntos de train, validation y test.	85
6.8. Evolución del MAE en el conjunto de entrenamiento en Resnet-CNN durante transfer learning.	86
6.9. Evolución del MAE en el conjunto de validación en Resnet-CNN durante transfer learning.	86
6.10. Evolución del MAE en el conjunto de entrenamiento en Resnet-CNN durante fine tuning.	87
6.11. Evolución del MAE en el conjunto de validación en Resnet-CNN durante fine tuning.	87
6.12. Evolución del MAE en el conjunto de entrenamiento en Panorama-CNN con imágenes RGB.	88
6.13. Evolución del MAE en el conjunto de validación en Panorama-CNN con imágenes RGB.	88
6.14. Evolución del MAE en el conjunto de entrenamiento en Panorama-CNN con mapas de características.	89
6.15. Evolución del MAE en el conjunto de validación en Panorama-CNN con mapas de características.	89

Índice de tablas

1.1. Intervalos de edad de la muerte propuestos por Todd.	23
2.1. Diferentes clasificaciones para los distintos tipos de datos 3D.	39
3.1. Ejemplos de modelos 3D de síntesis púbicas pertenecientes a cada una de las fases de Todd.	50
3.2. Resumen de los resultados presentes en el estado del arte de los métodos de estimación de la edad.	53
3.3. Resumen de los resultados presentes en el estado del arte de los métodos de representación de datos 3D mediante proyecciones.	56
5.1. Planificación temporal inicial del proyecto.	71
5.2. Planificación temporal final del proyecto.	72
5.3. Costes del proyecto.	72
6.1. Rotaciones en el eje Y de una muestra del conjunto de datos.	80
6.2. Rotaciones en los ejes X y Z de una muestra del conjunto de datos.	80
6.3. Resultados RMSE y MAE totales para un conjunto de test con aumento de datos.	90
6.4. Resultados RMSE y MAE totales para un conjunto de test sin aumento de datos.	90
6.5. Resultados RMSE y MAE para niños/adultos/ancianos totales para un conjunto de test con aumento de datos.	91
6.6. Resultados RMSE y MAE para niños/adultos/ancianos totales para un conjunto de test sin aumento de datos.	91
6.7. Resultados RMSE y MAE para cada fase de Todd totales para un conjunto de test con aumento de datos.	92
6.8. Resultados RMSE y MAE para cada fase de Todd totales para un conjunto de test sin aumento de datos.	93
6.9. Comparación entre los resultados obtenidos y los resultados del estado del arte.	95

A.1.	Resultados MAE para niños/adultos/ancianos por ejecución para un conjunto de test con aumento de datos.	109
A.2.	Resultados RMSE para niños/adultos/ancianos por ejecución para un conjunto de test con aumento de datos.	110
A.3.	Resultados MAE para niños/adultos/ancianos por ejecución para un conjunto de test sin aumento de datos.	110
A.4.	Resultados RMSE para niños/adultos/ancianos por ejecución para un conjunto de test sin aumento de datos.	111
A.5.	Resultados MAE para cada fase de Todd por ejecución para un conjunto de test con aumento de datos.	112
A.6.	Resultados RMSE para cada fase de Todd por ejecución para un conjunto de test con aumento de datos.	113
A.7.	Resultados MAE para cada fase de Todd por ejecución para un conjunto de test sin aumento de datos.	114
A.8.	Resultados RMSE para cada fase de Todd por ejecución para un conjunto de test con aumento de datos.	115

Capítulo 1

Introducción

1.1. Descripción del problema

La **antropología forense** (AF) es un subcampo de la antropología¹, concretamente de la antropología física², que aborda el análisis de restos óseos humanos de importancia médica y legal [1]. Los expertos en dicha disciplina, dado su conocimiento de la biología del esqueleto humano y materias relacionadas, examinan los huesos humanos con el objetivo de obtener la mayor cantidad de información posible sobre las personas representadas por los restos óseos.

Los antropólogos forenses tratan de cumplir con cinco objetivos principales a la hora de realizar su trabajo [2]:

- Determinar el grupo poblacional, sexo, edad y estatura de la persona a la que pertenecen los restos óseos estudiados una vez los tejidos blandos se han deteriorado hasta tal punto que no es posible determinarlos mediante análisis visual.
- Cuando hay evidencia de heridas traumáticas en los restos óseos, identificar la naturaleza de estas heridas y sus posibles causas con la intención de recoger información para determinar la causa y circunstancias del fallecimiento.
- Establecer un intervalo *postmortem*, la cantidad de tiempo que ha pasado desde que la persona objeto de estudio falleció.

¹ La antropología (del griego ἄνθρωπος, «hombre (humano)», y λόγος, «conocimiento») es la ciencia que estudia al ser humano de una forma integral, de sus características físicas como animales y de su cultura.

² La antropología biológica o antropología física es una rama de la antropología y la biología que tiene como objeto el estudio de la evolución y variabilidad biológica humana, tanto pasada como actual.

- Asistir en la localización y recuperación de restos superficiales o enterrados, de manera que toda la evidencia correspondiente a la investigación forense sea recogida.
- Proveer información útil en la obtención de una identificación positiva de personas fallecidas.

Además de lo anteriormente mencionado, la AF cumple otras funciones dentro de la sociedad moderna. Se consulta a los especialistas en AF para la identificación de víctimas de catástrofes como accidentes aéreos, guerras, ataques terroristas o desastres naturales [3]. También asisten en el estudio de crímenes de guerra derivados de los numerosos arrestos de civiles producidos por la violencia política, particularmente en países de América central y del sur, África y Europa [4]. Por último, también participan en la estudio de personas históricamente relevantes pero no relevantes desde el punto de vista médico y legal [2].

De los objetivos de la AF citados con anterioridad, el que posee mayor relevancia para el desarrollo de este proyecto, es el campo de la **Identificación Humana (ID)** y la **estimación del perfil biológico (PB)**. La AF proporciona una alternativa con un mayor rango de aplicabilidad que otras técnicas de ID como análisis de huellas dactilares o análisis de ADN. Esta ciencia contribuye a la identificación de restos humanos de dos maneras: proporciona métodos que establecen identificaciones positivas y métodos que contribuyen a la identificación limitando las potenciales coincidencias con el individuo analizado [5]. La AF entra en escena cuando la descomposición del tejido blando está tan avanzada que los médicos forenses especialistas no pueden determinar características como la demografía, el tiempo, la causa y las circunstancias del fallecimiento. Sin embargo, los antropólogos forenses se involucran cada vez más como consultores, incluso cuando existe tejido blando presente.

La estimación del **PB** tiene un papel crucial a la hora de reducir el rango de posibles coincidencias a la hora de identificar determinados restos óseos. En la Figura 1.1 podemos ver que la estimación del PB es uno de los primeros pasos en el proceso de ID forense. El PB combina el estudio de los restos óseos con el objetivo de encontrar los rasgos característicos que determinan la identidad de un individuo. Entre estos rasgos característicos encontramos los siguientes:

- Estimación del sexo en individuos adultos y subadultos (niños).
- Estimación de la edad en individuos adultos y subadultos vivos o fallecidos.
- Estimación de la altura mediante el uso de huesos largos.

- Estimación de la ascendencia.

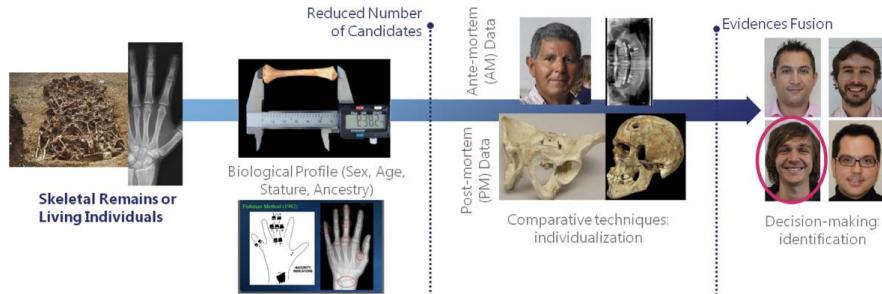


Figura 1.1: Diagrama de secuencia de los pasos realizados en la identificación forense basada en el esqueleto (skeleton-based forensic identification, SFI) [5].

Para llevar a cabo la tarea de estimar la edad de un individuo (o cadáver) existen numerosos métodos utilizados por los antropólogos forenses [6]. Podemos clasificarlos en métodos dentales y esqueléticos según si analizan la dentadura o se centran en cualquier otro hueso del esqueleto humano. Dentro de los métodos que analizan los huesos, la estimación de la edad a partir de la observación de características macroscópicas en las sínfisis públicas es uno de los enfoques más aceptados.

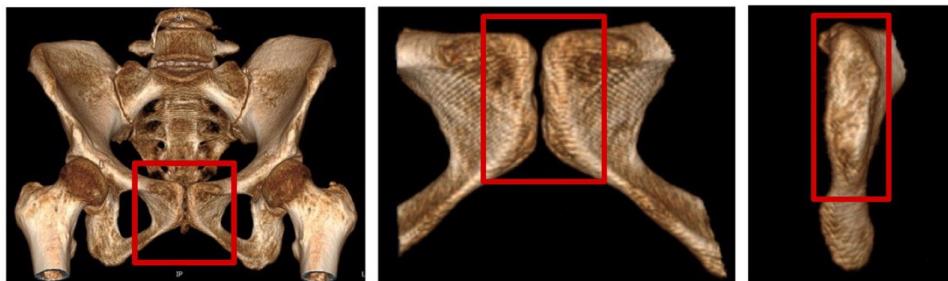


Figura 1.2: Reconstrucción volumétrica por tomografía computerizada 3D de la sínfisis del pubis en RadiAnt [7]. Las imágenes no tienen la misma escala. **Izquierda:** Pelvis humana. **Centro:** Ampliación de ambas partes del pubis. **Derecha:** Cara del pubis izquierdo. Encuadrado en rojo se ha marcado la sínfisis pública.

Como podemos apreciar en la Figura 1.2, la sínfisis del pubis es la conexión entre las dos partes del pubis. Esta puede observarse tanto en la cara del pubis derecho como del izquierdo. Es en estas caras donde los antropólogos forenses centran su atención a la hora de identificar características

morfológicas. Estas características, discriminatorias y observables a simple vista, fueron descritas por T. W. Todd en su artículo *Ages changes in the pubic bone* (1921) [8]. Aún a día de hoy, 100 años después de la publicación de este artículo, los antropólogos forenses localizan y emplean las características descritas por Todd a la hora de estimar la edad de los restos óseos de personas muertas en edad adulta. Entre dichas características se encuentran: la porosidad, definición y profundidad de las crestas y surcos; la porosidad y regularidad de la superficie y la presencia de nódulos o bordes; entre otras. En la Figura 1.3 podemos ver un ejemplo de algunas de las características definidas por Todd.

V1: Articular face: Ridges & grooves	V2: Irregular Porosity: Much	V3: Upper symphysial extremity: defined	V4: Bony nodule: Present	V5: Lower symphysial extremity: defined	V6: Dorsal margin: Present	V7: Dorsal Plateau: Present

Figura 1.3: Algunos ejemplos de características que definen el desarrollo típico y los cambios morfológicos degenerativos en la síntesis pública durante el crecimiento. Las regiones de interés están resaltadas en rojo

Todd también estableció varios rangos de edad para los que los restos óseos presentaban características similares. Como se puede observar en la Tabla 1.1 estos rangos o fases abarcan de 1 o 2 años para muestras jóvenes (18-26 años) a 4 años de diferencia en las últimas fases. Además el último rango de edad es catalogado como 50 años o más. Esto era adecuado para la época de publicación del artículo, pero queda bastante obsoleto en el momento actual en el que la esperanza de vida es superior.

Para realizar el proceso de estimación de la edad a partir de restos humanos, podemos diferenciar dos enfoques distintos: realizar la estimación aportando un valor continuo para la edad de la muerte, como el enfoque propuesto por Gilbert-McKern [9], o determinar la edad en un intervalo, como los enfoques propuestos por el propio Todd [8] o el propuesto por Suchey-Brooks [10]. En el estado del arte actual ha habido diversos intentos de automatizar estos métodos de estimación de la edad (ver Capítulo 3).

Fase 1	19 años	Fase 6	31 a 34 años
Fase 2	20 a 21 años	Fase 7	35 a 39 años
Fase 3	22 a 24 años	Fase 8	40 a 44 años
Fase 4	25 a 26 años	Fase 9	45 a 49 años
Fase 5	27 a 30 años	Fase 10	50 años o más

Tabla 1.1: Intervalos de edad de la muerte propuestos por Todd.

Los métodos de automatización propuestos han aplicado diversos métodos de aprendizaje automático (regresores, modelos Bayesianos, redes neuronales, etc.) consolidados en cuanto a resolución de problemas de clasificación y regresión.

Sin embargo, en este proyecto proponemos un nuevo enfoque. Aplicar técnicas de **aprendizaje profundo** (DL) a la tarea de la estimación de la edad a partir de las síntesis públicas. Lo haremos, además, sin tomar como base ninguno de los métodos de extracción de características morfológicas actuales, si no que será el propio modelo quién aprenderá a extraer estas características de forma automática. Como datos de entrada contaremos con el conjunto de modelos 3D de las síntesis públicas humanas escaneadas manualmente por personal del laboratorio de Antropología Física del Departamento de Medicina Legal, Toxicología y Antropología Física de la Universidad de Granada. Por tanto, el desarrollo del proyecto presenta un reto doble. En primer lugar, el uso de modelos 3D en el campo del DL supone un reto en sí mismo, ya que no existen precedentes en su aplicación en el campo de la medicina legal. En segundo lugar, las técnicas de DL aplicadas deben extraer automáticamente características morfológicas discriminatorias de los modelos 3D de las síntesis públicas sin la intervención de un experto humano. Todo ello configura un TFG de enorme originalidad y carácter disruptivo, en donde se presenta una propuesta transgresora en relación al resto de modelos y técnicas existentes en el estado del arte. Los resultados proporcionados por la propuesta de este TFG se compararán con aquellos obtenidos con métodos pre-existentes.

Finalmente, en base a todo lo anteriormente descrito y a modo de resumen, el presente TFG aborda la tarea de **estimar la edad de un individuo mediante la aplicación de técnicas de DL a modelos 3D de la síntesis pública**.

1.2. Motivación

Los antropólogos forenses utilizan las características macroscópicas descritas en [8] para poder realizar una estimación de la edad de los restos analizados. La extracción de estas características (ver Figura 1.3) depende en gran medida de la experiencia y criterio del antropólogo que realiza el análisis. Esto en ocasiones puede concurrir en errores a la hora de realizar una estimación correcta de la edad de unos restos.

Actualmente uno de los principales retos a los que se enfrenta la AF es responder con precisión, robustez y rapidez al enorme número de casos que deben ser abordados. La estimación de la edad de restos óseos es una tarea fundamental en casos de personas desaparecidas, accidentes múltiples, incendios, desastres naturales, actos de terrorismo, fosas comunes, crímenes de lesa humanidad³, etc [1, 11]. Por esto la minimización de los errores cometidos por los antropólogos forenses en el proceso de ID es clave.

A continuación se muestran algunos ejemplos donde la estimación del PB y la aplicación de técnicas de AF resultan indispensables:

- Según el Centro Nacional de Desaparecidos (CNDES), desde el año 2011 han sido hallados 2.749 cadáveres o restos humanos sin identificar en España. De estos solo han podido ser identificados 656.
- Según datos ofrecidos por el Ministerio del Interior, en España hay en la actualidad 12.330 casos de personas desaparecidas sin resolver.
- Según el Ministerio de Justicia, en España se han contabilizado 2.457 fosas comunes. En estas, la Plataforma de Víctimas de Desapariciones Forzadas por el Franquismo estima que fueron enterradas más de 140.000 personas víctimas de ejecuciones extrajudiciales y asesinatos en custodia durante el período de 1936 a 1975.
- En junio de 2019 había contabilizados en España por el Ministerio del Interior un total de 12.301 menores migrantes no acompañados (MENAS). Entre los motivos que llevan a estos niños y niñas a salir de sus países de origen se encuentran la pobreza y la falta de futuro y expectativas; situaciones de desestructuración familiar y desprotección institucional; catástrofes naturales; la guerra, la persecución, la violencia y situaciones de violación generalizada de los derechos humanos.

³ Se considera crímenes de lesa humanidad —o contra la humanidad— a aquellos delitos especialmente atroces y de carácter inhumano, que forman parte de un ataque generalizado o sistemático contra una población civil, cometidos para aplicar las políticas de un Estado o una organización.

- Según los cálculos de la Agencia de la ONU para los Refugiados (AC-NUR), para fines de noviembre de 2021 la cantidad de personas muertas o desaparecidas en travesías a través del Mediterráneo ya superaba los 2500.
- Entre 2017 y 2020 a nivel mundial murieron 970 personas como consecuencia de ataques terroristas.
- En 2020, el número de víctimas mortales registradas a nivel mundial como consecuencia de las catástrofes naturales ascendió a aproximadamente 8.200. La AF tiene un papel fundamental en sucesos de víctimas múltiples, siendo responsable de la individualización de fragmentos, comparación radiográfica, estimación del PB, etc.
- La AF está fuertemente vinculada a la investigación de crímenes de lesa humanidad, siendo la principal responsable de la identificación de personas como consecuencia de desapariciones forzadas, ejecuciones extrajudiciales, masacres y otras formas de violencia.

A pesar de estos dramáticos datos, y del consiguiente interés humanitario, político y social que existe a la hora de resolver muchos de estos casos, sorprende constatar el bajo desarrollo tecnológico de la AF en comparación con otras disciplinas médicas [5]. Aún a día de hoy, las herramientas más utilizadas por los antropólogos forenses para el desempeño de su práctica diaria son el calibre y la hoja de cálculo.

Como ingenieros, tenemos la capacidad de dotar a la sociedad con herramientas que ayuden a desempeñar tareas complejas de forma que los errores cometidos por el factor humano sean mínimos. Existen numerosos ejemplos de sistemas de inteligencia artificial que adaptan métodos basados en el análisis esquelético para desempeñar la tarea de estimar el PB de un individuo [5].

Además en años recientes, las técnicas de DL aplicadas directamente sobre modelos 3D han aportado buenos resultados en resolución de tareas como clasificación de objetos, recuperación de la pose y segmentación [12, 13]. Dado que la tendencia dentro de la AF es que los antropólogos forenses trabajen cada vez más con modelos 3D de los restos óseos en lugar de trabajar directamente con estos, un sistema capaz de asistir en la tarea de la estimación de la edad a partir de la información obtenida de forma automática directamente de estos modelos 3D podría ser una herramienta de gran utilidad para los antropólogos forenses, guiándolos a la hora de realizar su labor.

La motivación de este TFG viene dada por la necesidad de dotar a los antropólogos forenses de herramientas para el apoyo en la estimación de la edad de restos humanos.

1.3. Objetivos

El objetivo general de este proyecto es, como se ha comentado anteriormente, estimar la edad de un individuo mediante la aplicación de técnicas de DL a modelos 3D de la sínfisis pública. Para esto definiremos un modelo que extraiga la información necesaria para realizar la estimación directamente a partir de los modelos 3D. Para poder desarrollar este modelo, encontraremos una representación de la información de los modelos 3D entre las opciones disponibles en el estado del arte. En la Figura 1.4 vemos un esquema de la propuesta presentada.

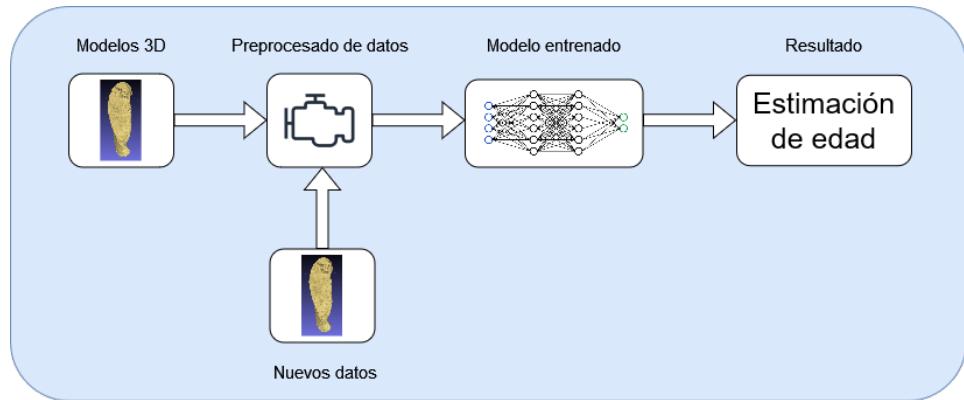


Figura 1.4: Representación del modelo a desarrollar.

Para el desarrollo de este TFG, dividiremos el objetivo general en los siguientes objetivos parciales:

- Dado que la fuente principal de datos se trata de una serie de modelos 3D, realizaremos un análisis pormenorizado y sistemático de las formas presentes en el estado del arte de representación de datos en 3D. De todas las formas existentes, enumeraremos las ventajas y desventajas de cada una, eligiendo finalmente una única forma de representar los datos.
- Desarrollaremos las herramientas necesarias para poder implementar la representación de los datos.
- Diseñaremos un modelo de red neuronal profunda que permita trabajar con la representación de datos 3D.

- Comprobaremos si la representación de la información escogida y el modelo desarrollado permiten aprender de forma automática y satisfactoria las características representativas de los restos óseos, de forma que la estimación del modelo pueda servir como apoyo para el trabajo de un antropólogo forense. Para esto, compararemos los resultados obtenidos con aquellos aportados por los métodos actuales presentes en el estado del arte.

Para ajustarnos a estos objetivos, y documentar el trabajo realizado de la forma más clara y completa posible, esta memoria se estructura de la siguiente forma: en el Capítulo 2 presentaremos los fundamentos teóricos en los que se basa el desarrollo del proyecto; en el Capítulo 3 analizaremos el estado del arte tanto para la estimación de la edad a partir de restos óseos como para la representación de datos 3D en el ámbito del DL; en el Capítulo 4 presentaremos los métodos utilizados para desarrollar los objetivos propuestos; en el Capítulo 5 se proporcionarán detalles sobre la planificación y la implementación del proyecto, en el Capítulo 6 veremos el protocolo de validación experimental del modelo y el proceso de entrenamiento del mismo, y analizaremos los resultados finales obtenidos y los compararemos con el estado del arte y, finalmente, en el Capítulo 7 presentaremos las conclusiones principales de este TFG.

Capítulo 2

Fundamentos teóricos

Esta sección tiene como objetivo presentar y explicar los fundamentos teóricos en los que se basan los métodos utilizados en el desarrollo de este proyecto, así como justificar su relevancia en la resolución del problema que se plantea.

2.1. Machine learning

El **machine learning** (ML) o **aprendizaje automático** [14, 15, 16] es una rama de la inteligencia artificial centrada en el uso de datos y algoritmos para desarrollar y entender modelos que sean capaces de aprender patrones existentes en los datos de forma progresiva. La forma en que estos modelos aprenden es mejorando gradualmente su rendimiento al resolver una tarea o problema concretos.

Los modelos de ML pueden aprender a partir de datos de cualquier tipo: números, palabras, estadísticas, imágenes, etc. Cualquier información almacenada digitalmente puede utilizarse como entrada para un modelo. Los modelos aprenden información de los patrones extraídos de los datos de entrenamiento, mejorando su rendimiento con el tiempo. Una vez entrenado, un modelo podrá identificar los patrones aprendidos en nuevos datos, distintos a los utilizados en el proceso de entrenamiento.

Estos modelos tienen una amplia variedad de aplicaciones: visión por computador, reconocimiento del lenguaje natural, aplicaciones en el campo de la medicina, filtrado automático de contenido, etc. Son particularmente útiles en tareas que no pueden ser abordadas por algoritmos convencionales.

Diferenciamos tres aproximaciones o problemáticas principales dentro del ML: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo.

El **aprendizaje supervisado** consiste en presentar los datos de entrenamiento como un conjunto de datos de entrada (*inputs*) etiquetados con la información que debe determinar (*output*). Esta aproximación requiere un menor número de datos de entrenamiento y facilita el proceso de aprendizaje, ya que los resultados obtenidos por el modelo pueden compararse con los datos ya etiquetados. Sin embargo, este proceso de etiquetado puede ser costoso, ya que hay que conocer previamente como se deben etiquetar los datos para que el modelo pueda aprender y aplicar el conocimiento adquirido en la clasificación de nuevos datos. Además, el modelo es muy sensible al etiquetado de datos, pudiendo obtener modelos sesgados en función al etiquetado de los datos de entrenamiento. Las problemáticas de aprendizaje supervisado incluyen algoritmos de clasificación y regresión. Los algoritmos de clasificación se aplican en tareas donde las etiquetas están limitadas a un conjunto de valores discretos, mientras que los de regresión se aplican en casos en los que las etiquetas presentan un dominio continuo de valores.

A diferencia del aprendizaje supervisado, en el **aprendizaje no supervisado** los datos de entrada no se encuentran etiquetados. Esta aproximación se centra en la identificación de patrones y relaciones existentes en el conjunto de datos de entrenamiento, de forma que pueda etiquetar y clasificar los datos sin la intervención humana. Requiere un mayor número de datos y es menos utilizada frente a otras aproximaciones. El aprendizaje semi-supervisado es un enfoque intermedio que consiste en utilizar un subconjunto de datos etiquetado para guiar la extracción y detección de patrones en el conjunto no etiquetado.

Finalmente, el **aprendizaje por refuerzo** se basa en determinar las acciones que debe llevar a cabo un algoritmo o agente en un determinado entorno, con el objetivo de maximizar la recompensa acumulada. El agente es recompensando positiva o negativamente en función de su desempeño. Esta técnica se aplica en campos como los algoritmos genéticos o la teoría de juegos.

Aunque estas tres problemáticas son las principales, existen muchas aproximaciones distintas como el aprendizaje débilmente supervisado, auto-supervisado, multitarea, etc. Muchas de estas reúnen características de varias aproximaciones, como es el caso del aprendizaje semi-supervisado. Las aproximaciones que hemos comentado no son departamentos estancos, por lo que un algoritmo de ML puede ubicarse en varias de ellas en función de como se aborde el problema a resolver.

2.2. Deep learning

El **deep learning** (DL) o **aprendizaje profundo** [17, 18, 19] es un subcampo del ML. Los métodos de DL tienen como objetivo obtener representaciones más abstractas (en última instancia, más útiles) mediante la combinación de múltiples transformaciones no lineales [20]. ML y DL difieren en cómo aprende cada algoritmo. DL automatiza el proceso de extracción de características de los datos, eliminando la intervención humana y consiguiendo que los algoritmos de DL sean aplicables en conjuntos de datos mucho mayores. También se puede entender el DL como una versión escalable del ML clásico [21]. El ML depende en gran medida de la intervención de un experto en el conjunto de datos capaz de extraer las características del mismo. Los métodos empleados en DL pueden aplicarse en conjuntos de datos etiquetados, de la misma forma que se aplicarían los métodos de ML, sin embargo, no es una condición necesaria. Los modelos DL pueden extraer conjuntos de características discriminatorias para elementos de distinta clase a partir de datos en bruto como texto o imágenes.

2.2.1. Redes neuronales

Las **redes neuronales** [22], también conocidas como redes neuronales artificiales (ANN), son un subcampo del ML y forman el núcleo del DL. Su nombre y su estructura toma inspiración del cerebro humano. Se busca imitar la forma en que las neuronas biológicas interactúan entre sí.

Las redes neuronales están formadas por capas de nodos: una capa de entrada, una o varias capas de nodos ocultas y una capa de salida. Cada neurona (o nodo) se conecta a las de la capa siguiente como puede verse en la Figura 2.1. Esta conexión tiene un peso y un umbral de activación asociado. Si el valor de salida de un nodo es superior al umbral de activación, la información se transmite hacia delante entre las capas de nodo. De lo contrario esta información no se transmite entre las capas. Los valores producidos como salida de una red neuronal pueden ser tanto valores continuos, binarios o categorías específicas en función de la tarea que se deseé realizar.

Cada conexión entre dos nodos tiene un peso asignado, de forma que se modela la “fuerza” de cada una de las señales de entrada a un nodo. El correcto ajuste de estos pesos será el que permitirá a la red neuronal aprender a realizar la tarea que debe llevar a cabo. Los datos introducidos a la capa de entrada que pueden verse en la Figura 2.2 se definen como variables independientes. Serán procesados por el siguiente nodo para propagar la información hacia delante en la red. Todas las variables independientes de la red corresponden a una única muestra del conjunto de datos disponible. La forma en la que los datos se procesan en los nodos es multiplicando las señales de entrada por el peso asignado a cada señal y combinando li-

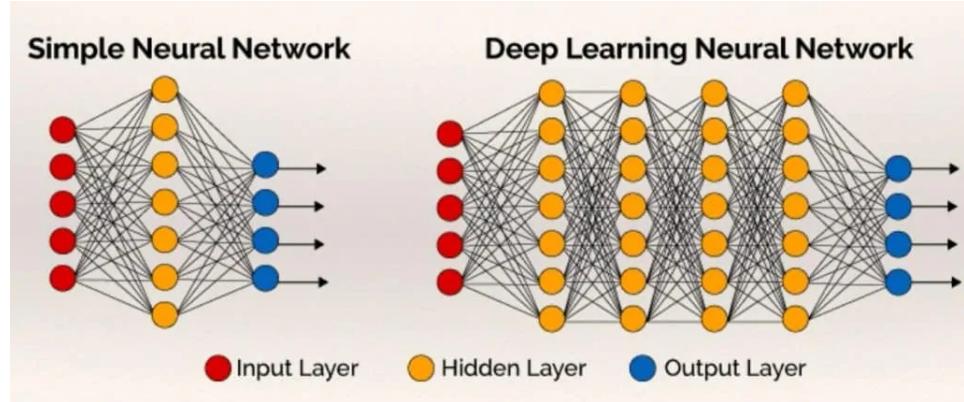


Figura 2.1: Ejemplo de red neuronal superficial y profunda [23].

nealmente estas operaciones. Posteriormente a esta combinación lineal se le aplica una función de activación.

Las funciones de activación son una parte fundamental de las redes neuronales, pues son las que introducen la no linealidad en el modelo y en última instancia, permiten a la red aprender. De aplicarse funciones de activación lineales, la red sería equivalente a una red sin capas ocultas. Las funciones de activación más comunes suelen ser la función threshold, la función sigmoide, la función rectificadora (ReLU) y la función tangente hiperbólica. Sin embargo existen más funciones de activación distintas a las mencionadas. La elección de una función de activación u otra esta motivada por el problema a resolver y por los resultados experimentales obtenidos.

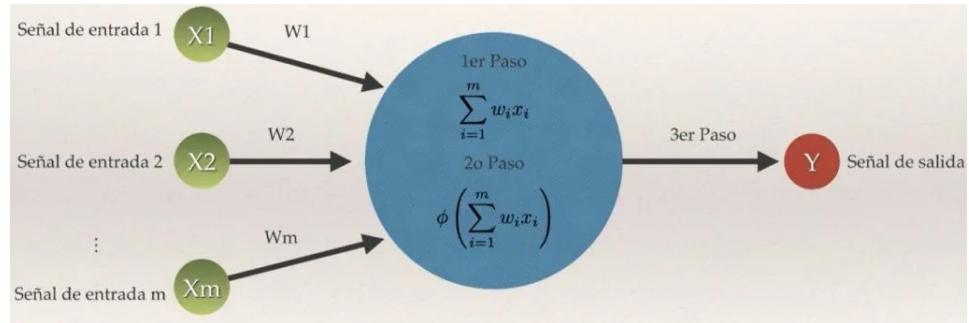


Figura 2.2: Procesamiento de señal a nivel de nodo [23].

Además de la función de activación, otra función imprescindible a la hora de hablar de redes neuronales es la función de pérdida o función de error. Ya hemos comentado que el objetivo de la red neuronal es ajustar los pesos asociados a sus conexiones de forma que el resultado final obtenido sea el resultado deseado. Para medir la diferencia entre el resultado obtenido y el

resultado deseado establecemos una función de pérdida. El valor obtenido por esta función será utilizado para realizar un ajuste en los pesos del modelo en un proceso conocido como “backpropagation”. La información de la función de pérdida se transmite desde la última capa de la red hacia la capa inicial. De nuevo, la función de perdida utilizada dependerá del objetivo a desempeñar y de los resultados experimentales. Existen funciones específicas para regresión como la función error absoluto medio (MAE) y la función error cuadrático medio (MSE). También hay funciones específicas para tareas de clasificación como la función de entropía cruzada de categorías (Categorical CrossEntropy) o la función de entropía cruzada binaria (Binary CrossEntropy).

El proceso iterativo de propagar la información de la capa de entrada a través de los nodos, procesar esta información mediante combinaciones lineales de las señales y la aplicación de funciones de activación, evaluar la función de pérdida y realizar backpropagation para ajustar los pesos del modelo se conoce como entrenamiento de una red neuronal. Según avanza este proceso iterativo, el objetivo de la red es minimizar el valor de la función de pérdida.

Es importante tener en cuenta que una red neuronal puede sobreaprender los datos de entrada, incurriendo en un fenómeno conocido como “overfitting” (al igual que otros modelos dentro del campo del ML). El “overfitting” o sobreajuste se produce cuando la red neuronal es capaz de clasificar los datos de entrenamiento con una precisión muy alta, pero sin embargo al intentar clasificar nuevos datos diferentes a los de entrenamiento, la red no consigue generalizar correctamente y su desempeño es mucho menor. Para esto existen procesos de regularización que están diseñados para evitar que se produzca sobreajuste en el modelo.

2.2.2. Redes neuronales convolucionales

Las **convolutional neural networks** (CNN) o redes neuronales convolucionales [24, 25] son un tipo de algoritmo de DL, concretamente una subclase de las redes neuronales. Estas redes están diseñadas para trabajar con imágenes como dato de entrada¹. Son capaces de detectar las regiones características capaces de diferenciar las imágenes que reciben como input a partir de los pesos de la red. El preprocesado necesario en este tipo de redes es mucho menor que en el de otros modelos, ya que son capaces de aprender y desarrollar filtros capaces de detectar las características discriminatorias de una imagen.

¹ Generalmente es así, pero existen arquitecturas de CNN específicas para modelos 3D y otro tipo de datos.

La arquitectura de este tipo de redes es análoga al patrón de conexión entre neuronas de las redes neuronales clásicas. Las CNN son capaces de capturar las dependencias temporales y espaciales presentes en una imagen mediante la aplicación de filtros relevantes que la red aprende en el proceso de entrenamiento. El objetivo de las CNN es reducir las imágenes de forma que se facilite el proceso de aprendizaje sin perder características discriminatorias de las mismas. Esto favorece a que este tipo de redes sean escalables a conjuntos de datos masivos.

Capa de convolución

El operador de **convolución** es la base de las CNN (ver Figura 2.3). Una convolución es un operador matemático que transforma dos funciones f y g en una tercera función que representa la magnitud en la que se superponen f y una versión invertida y trasladada de g . En el caso de las CNN, estas funciones f y g se tratan de una imagen (I) y un kernel o filtro (K) respectivamente. El kernel se desplaza por la imagen calculando la convolución entre la sección de la imagen correspondiente y el kernel (ver Figura 2.4). El kernel se desplaza hacia la derecha con determinado valor de salto (“stride”) hasta que se completa la primera fila. Después el kernel comienza a desplazarse de la misma manera por las siguientes filas comenzando desde la izquierda. Este proceso continua hasta que el kernel pasa por toda la imagen. En el caso de imágenes con más de un canal, el kernel tiene la misma profundidad que la imagen y el resultado de la convolución es una imagen con un único canal que se compone de la suma de la convolución de cada canal con el kernel.

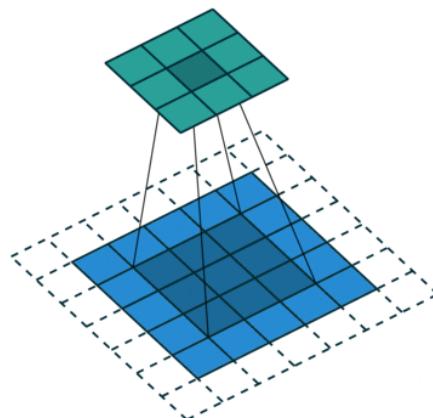


Figura 2.3: Operación de convolución [26].

El objetivo de la capa de convolución es extraer características de alto nivel presentes en la imagen. Generalmente la primera capa convolucional es

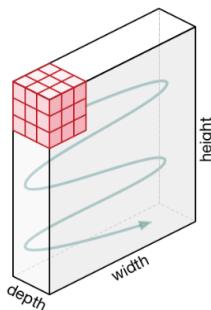


Figura 2.4: Movimiento del kernel sobre la imagen [26].

responsible de extraer características de bajo nivel [26] como colores, bordes, gradiente de orientación, etc. Según se van añadiendo capas de convolución, la arquitectura se adapta para extraer también características de alto nivel. Esta combinación de varias capas convolucionales, proporciona a la red un entendimiento completo de las imágenes del conjunto de datos.

Hay dos posibles resultados tras aplicar la convolución a una imagen, esta puede o reducir sus dimensiones o mantener las mismas. Este resultado se controla mediante la aplicación de “padding” o relleno. El padding consiste en aumentar las dimensiones de la imagen, introduciendo información redundante o llenando el nuevo espacio con ceros, de forma que la imagen resultado de la convolución no pierda tamaño.

Capa de pooling

Al igual que la capa de convolución, la capa de “pooling” o agrupación se encarga de reducir la dimensión espacial de las convoluciones realizadas. La reducción de la dimensionalidad contribuye a reducir el coste computacional necesario para procesar los datos de entrada. Además la capa de pooling sirve para extraer las características dominantes que son invariantes a rotaciones y a la posición, manteniendo la efectividad en el proceso de entrenamiento de la red.

Los dos tipos de pooling más utilizados son: **max pooling** (agrupación máxima) y **average pooling** (agrupación media). El max pooling consiste en conservar el valor máximo de la sección de la imagen cubierta por el kernel. Además aplicar max pooling contribuye a suprimir ruido presente en la imagen. Descarta las activaciones provocadas por ruido mientras produce una reducción de la dimensionalidad. El average pooling por otro lado, consiste en obtener la media de los valores de la sección de la imagen cubierta por el kernel. El average pooling suprime el ruido presente en la imagen simplemente consiguiendo una reducción de la dimensionalidad. Un ejemplo

de ambos tipos de pooling puede verse en la Figura 2.5.

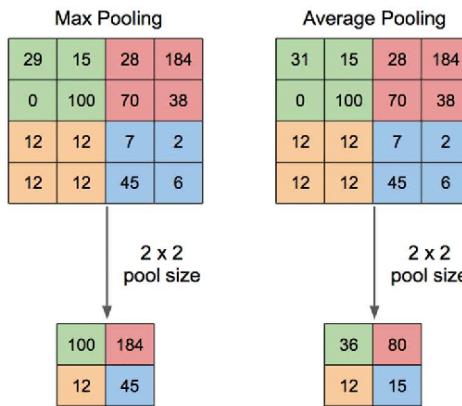


Figura 2.5: Tipos de pooling [27].

La composición de capas de convolución y pooling forman la base de las CNN. Aumentando y reduciendo su número, puede controlarse la extracción de características de mayor o menor nivel en función de la complejidad de la imagen de entrada. Sin embargo, un mayor número de capas conllevará un mayor coste computacional en el procesado de los datos.

Capa densa o totalmente conectada

Una vez terminado el proceso de aplicación de convolución y pooling, los datos procesados se reducen a un único vector de características que se utiliza como entrada para una red neuronal clásica que se encarga de realizar la clasificación, regresión o la tarea que se desee (ver Figura 2.6). Podemos entender las CNN como una red neuronal formada por dos bloques principales: uno de convolución y pooling, que actúa como extractor automático de características y otro bloque formado por una red neuronal clásica, que utiliza las características extraídas para cumplir con el objetivo de la red.

En las CNN, las capas en las que los nodos están totalmente conectados, en una arquitectura análoga a una red neuronal clásica, se las conoce como **capas densas, totalmente conectadas** o “**fully connected**”. Esta es una forma sencilla de conseguir que el modelo aprenda combinaciones no lineales de las características obtenidas en las capas de convolución y pooling.

La salida de esta capa, es el resultado final de la CNN, que será evaluado por una función de pérdida. Se llevará a cabo un proceso de backpropagation que actualizará los pesos de las conexiones entre nodos así como los valores de los filtros utilizados en las convoluciones, con el objetivo de minimizar el valor de la función de pérdida utilizada. El proceso de entrenamiento de una

CNN es análogo al entrenamiento de una red neuronal clásica.

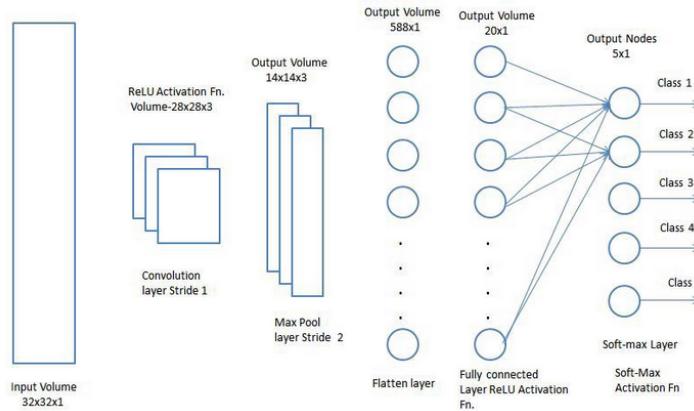


Figura 2.6: Arquitectura de CNN en la que pueden apreciarse las capas de convolución, pooling y totalmente conectada o densa [26].

Batch normalization y dropout

Como hemos comentado en la sección 2.2.1, durante el proceso de entrenamiento de una CNN puede producirse un fenómeno conocido como “overfitting”. Este fenómeno impide a la CNN generalizar correctamente la información aprendida, por lo que el desempeño de la red fuera de los datos de entrenamiento se reduce drásticamente. Para evitar el overfitting, existen estrategias que pueden aplicarse en las redes neuronales clásicas y en las CNN. Aquí nos centraremos en dos de ellas, batch normalization y dropout.

El **batch normalization** (normalización de subconjunto) es un método algorítmico que permite que el entrenamiento de la CNN sea más rápido y más estable. En la práctica consiste en una capa que se encarga de normalizar los vectores de activación de las capas ocultas de la red. Para esto utiliza la media y la varianza de cada “batch” de datos. El concepto de batch en las CNN hace referencia al número de datos que se introducen a la red antes de que se produzca el proceso de backpropagation y la actualización de los pesos de la red. Batch normalization tiene un efecto regularizador en la red, por lo que contribuye a prevenir el overfitting.

El **dropout** (abandono) es una técnica de regularización en redes neuronales. Consiste en “apagar” nodos aleatorios (tanto de las capas ocultas como de las visibles) durante el proceso de entrenamiento (ver Figura 2.7). Esto se consigue omitiendo el valor de los nodos en función de una probabilidad concreta. La aplicación del dropout consigue que la red entrenada sea más robusta [28], ya que permite que la red aprenda correctamente evitan-

do la posibilidad de que varios nodos compensen los errores de otros en el proceso de entrenamiento.

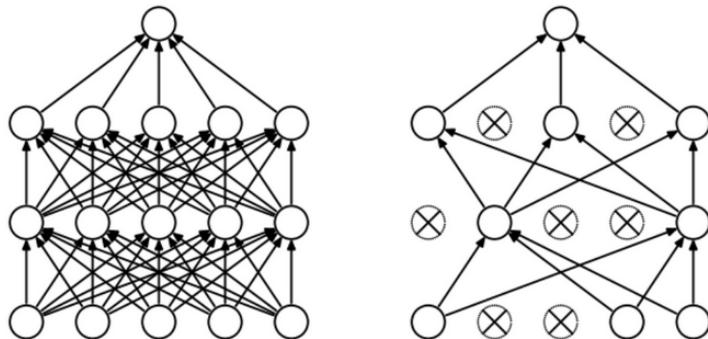


Figura 2.7: Ejemplo de aplicación de dropout. **Izquierda:** Red completa. **Derecha:** Red tras la aplicación de dropout [29].

2.3. Tipos de representación de datos en 3D

La aplicación de técnicas de DL en visión por computador ha estado generalmente dirigida al uso de datos 2D. En este campo se han obtenido numerosos resultados en tareas de clasificación, reconocimiento, segmentación, etc. La principal ventaja de los modelos DL es su capacidad para aprender progresivamente características jerárquicas de los datos proporcionados como entrada [30]. Sin embargo estos modelos requieren una gran cantidad de datos para su entrenamiento. Debido a esto, el uso de modelos DL en datos 3D ha estado limitado. En los últimos años, debido al avance de la tecnología en escaneo 3D, al abaratamiento y mejor accesibilidad de la misma, la cantidad de datos 3D disponibles para entrenamiento ha aumentado notablemente. Ejemplos como la popularidad del Kinect o el desarrollo de los sensores LIDAR han contribuido a eso.

Los algoritmos utilizados en modelos DL orientados a datos 3D deben afrontar principalmente dos desafíos: la representación de estos modelos 3D y la estructura de red utilizada. En esta sección abordaremos el primero de estos desafíos analizando las representaciones existentes de datos 3D, sus ventajas y desventajas y su posible aplicación en la estimación del PB.

Según la literatura consultada, existen varias formas de categorizar los tipos de datos 3D. Gezawa et al. [13] distinguen en su estudio 5 grandes categorías: datos en bruto, superficies, sólidos, estructuras de alto nivel y múltiples vistas. Esta distinción se basa en la forma de adquisición de los datos y la estructura interna de los mismos. Por otro lado Ahmed et al. [12]

dividen los tipos de datos 3D en euclídeos y no euclídeos. Los datos euclídeos cuentan con una estructura de “rejilla” subyacente que permite una parametrización global de los datos y establecer un sistema de coordenadas común. Los datos no euclídeos no cuentan con esta estructura subyacente, por lo que no existe parametrización global. En la Tabla 2.1 vemos un representación esquemática de estas categorías y las diferentes formas de representar los datos 3D que pertenecen a cada una de ellas.

	Euclídeo	No Euclídeo
Datos en bruto	RGB-D [31]	Nubes de puntos [32]
	Proyecciones 3D [33]	
Sólidos	Octrees [34]	
	Vóxeles [34]	
Superficies		Malla 3D [35]
Estructuras de alto nivel	Descriptores 3D [36]	Grafos [35]
Datos de múltiples vistas	Múltiples vistas [37]	

Tabla 2.1: Diferentes clasificaciones para los distintos tipos de datos 3D.

2.3.1. Datos en bruto

Dentro de la categoría de datos en bruto se agrupan representaciones que utilizan datos directamente obtenidos del positivo de captura obtenido, aplicando transformaciones simples o directamente no aplicando ninguna transformación. Podemos encontrar los datos RGB-D, las proyecciones 3D y las nubes de puntos.

Datos RGB-D

Los datos RGB-D representan la información de los datos en 2.5 dimensiones. Están formados por una imagen clásica de 2 dimensiones (RGB) a la que se añade un mapa de profundidad (D) que aporta el valor de profundidad captado por el sensor para cada píxel. Es una representación simple pero efectiva.

La principal ventaja de esta representación es la sencillez en su obtención (ya que los dispositivos utilizados, como Kinect, tienen gran disponibilidad) y su facilidad de aplicación a modelos DL preparados para 2D existentes. Su principal desventaja es que, dado que la representación RGB-D no contiene la totalidad de la geometría del objeto 3D, los modelos utilizados no podrán aprender la geometría completa de dicho objeto, por lo que solo se podrán

inferir algunas propiedades 3D en base a la profundidad. Esto ha motivado a los investigadores a explorar otras técnicas de representación de datos.

Proyecciones 3D

Proyectar puntos 3D en un espacio 2D es otra forma de representar de datos 3D. La proyección encapsula información clave del objeto 3D original. Comúnmente las proyecciones más utilizadas son las proyección cilíndricas y esféricas. Estas últimas poseen la ventaja de proporcionar invarianza a la rotación en el eje principal de la proyección.

Su principal ventaja es que dado que la proyección contiene información clave del objeto 3D original, es posible utilizar modelos de DL clásicos para trabajar con datos 3D. La principales desventajas son que la aplicación directa de esta representación en modelos DL en 2D requiere un mayor ajuste fino y que al proyectar datos 3D en un espacio 2D las propiedades geométricas de la forma del objeto 3D original se pierden. Por esto, muchos investigadores han tratado de utilizar varias proyecciones en conjunto o combinar esta representación con otras para compensar esta falta de información. Pese a sus desventajas, la utilización de este tipo de representación 3D en modelos DL ha probado ser eficaz a la hora de aprender formas 3D.

Nubes de puntos

Una nube de puntos no es más que un conjunto no estructurado de puntos 3D que aproximan la geometría de objetos 3D en el que cada punto está representado por tres coordenadas. Aunque las nubes de puntos pueden obtenerse fácilmente (utilizando dispositivos como Kinect o sensores LIDAR) su procesamiento puede ser un desafío debido a la falta de información de conectividad entre los puntos que forman la nube. También puede darse el caso en que debido a las condiciones de adquisición de los datos, ciertos puntos no se representen correctamente, las nubes representen información incompleta o se introduzca ruido en los datos, provocando pérdidas de información.

Si podemos pasar por alto estos inconvenientes, las nubes de puntos presentan varias ventajas. Estas proveen una representación expresiva, homogénea y compacta de la geometría de la superficie 3D con mucha menor complejidad respecto a otras representaciones como mallas 3D o grafos. Su principal desventaja a la hora de procesar los datos es la estructura desordenada y no uniforme que presentan, debido a los errores derivados de la adquisición de los datos. Esto ha llevado a los investigadores a desarrollar métodos de aprendizaje que sean invariantes al orden de los puntos dentro de la nube.

2.3.2. Sólidos

En esta categoría encontramos representaciones de los datos 3D que simplemente proporcionan información de control del espacio ocupado por un objeto determinado. Generalmente dicha información es binaria, lo que significa que un punto del espacio puede estar ocupado o no. Dentro de esta categoría encontramos representaciones con véxeles y octrees.

Vóxeles

Los véxeles modelan los datos 3D describiendo como el objeto 3D se distribuye en el espacio tridimensional. La información que proporciona el punto de vista también puede ser codificada clasificando los véxeles ocupados en visibles u ocluidos.

La principal ventaja es que proporciona información volumétrica de la totalidad del objeto 3D que se representa. Sin embargo, esta representación presenta una gran desventaja: requiere una enorme cantidad de memoria, ya que se deben almacenar todos los véxeles independientemente de si son visibles o no. Por esto, la representación en véxeles no es adecuada para representar objetos 3D con resoluciones altas. Cuanto mayor sea la resolución, mayor será el almacenamiento necesario.

Octrees

Como una forma más eficiente de representación de la información volumétrica de objetos 3D, surgen los octrees (árboles octales). Un octree es simplemente una estructura de representación de información espacial basada en véxeles de tamaño variable. Los octrees se construyen como árboles de los que cada nodo cuelgan ocho nodos diferentes [38] que dividen el espacio 3D en cubos que pueden estar tanto dentro como fuera del objeto.

Sus principales ventajas son la eficiencia en la utilización de la memoria y su capacidad para generar véxeles de alta resolución [39]. Por otro lado, su principal desventaja es que presenta una incapacidad para mantener la geometría de algunos objetos 3D como la suavidad de la superficie, por ejemplo.

2.3.3. Superficies

Las superficies de polígonos se suelen utilizar en la representación de los límites de los objetos 3D que rodean la parte interior del objeto. El conjunto de estos polígonos suele almacenarse para la descripción del objeto, lo que tiene la ventaja de la simplicidad y la rapidez de la representación de la superficie y de la visualización del objeto dado que todas superficies

se pueden caracterizar con ecuaciones lineales. Dentro de esta categoría solo encontramos las mallas 3D.

Mallas 3D

Las mallas 3D son una de las representaciones más populares de objetos 3D. Consisten en una combinación de vértices, aristas y caras que se utilizan generalmente se utilizan gráficos por ordenador para renderizado y almacenamiento de modelos 3D. Los vértices se asocian con una lista de conectividad que describen como estos se interconectan entre ellos. La complejidad e irregularidad de estas mallas provocan que no hayan sido utilizadas con técnicas DL orientadas a modelos 3D hasta ahora.

Su principal ventaja es que se trata de una de las representaciones de datos 3D más importantes y potentes dentro de su campo. Esto ha propiciado su amplio uso dentro del campo de los gráficos por ordenador. Su principal desventaja es la ya comentada complejidad e irregularidad de las mallas 3D. El estudio de la aplicación de estas mallas al campo del DL continua avanzando a día de hoy, por lo que los modelos actuales son bastante recientes además de computacionalmente costosos.

2.3.4. Estructuras de alto nivel

En problemas de clasificación de objetos 3D, a veces es necesaria una representación concisa y detallada. Dentro de las estructuras de alto nivel, podemos encontrar descriptores 3D de alto nivel y grafos.

Descriptores 3D

Los descriptores de forma son representaciones simplificadas de objetos 3D que describen características geométricas o topológicas de los mismo. Estos descriptores pueden obtenerse a partir de elementos del objeto 3D tales como la geometría, topología, superficie, textura o cualquier otra, siendo posible también utilizar una combinación de todas ellas [40, 41]. Un descriptor de forma puede entender como una “firma” de la forma del objeto 3D para facilitar el procesamiento y computación así como facilitar la comparación entre diferentes objetos. Estos descriptores normalmente se combinan con un modelo basado en el aprendizaje para extraer características jerárquicas discriminatorias para representar el objeto 3D de mejor forma. Los descriptores 3D se clasifican en 2 categorías: descriptores locales y globales. Los descriptores globales proporcionan información del objeto 3D en su totalidad, mientras que los locales proveen una representación de una zona concreta del modelo.

Como principales ventajas encontramos que los descriptores 3D pueden codificar mucha información relativa a un modelo 3D de forma simple. Por esto han sido ampliamente utilizados en el campo del DL. Sin embargo, su principal desventaja es que un descriptor demasiado simple puede conllevar una pérdida de las propiedades del modelo que esta representando.

Además revisando la literatura, la mayoría de aplicaciones de los descriptores 3D se centran en el paradigma del aprendizaje no supervisado. Esto se debe a que los modelos de aprendizaje supervisado aprenden abstracciones jerárquicas de los datos en bruto. Dado que los descriptores 3D son en si mismos una abstracción, su aplicación en modelos de aprendizaje supervisado podría derivar en no obtener características representativas. El proceso de aprendizaje de abstracciones a partir de abstracciones puede llevar a una pérdida de las propiedades actuales de los modelos 3D representados si el descriptor 3D utilizado es demasiado simple o abstracto. Sin embargo en algunos casos, los descriptores pueden proveer información lo suficientemente completa como para que la operación de convolución pueda aprender características jerárquicas de los datos de entrada [42].

Grafos

La representación por grafos de un objeto 3D recoge la esencia de la geometría del objeto enlazando diferentes partes del modelo mediante un grafo. Las mallas 3D son extensiones de datos estructurados en forma de grafos, en los que los nodos del grafo se utilizan como los vértices de la malla y las aristas representan las conexiones entre los vértices [43]. Los primeros acercamientos a redes neuronales convoluciones para grafos, Graph Convolutional Neural Networks (GNN), aplican principalmente dos operadores de convolución distintos: métodos de filtrado espectral y métodos de filtrado espacial. Aunque ambos métodos son equivalentes aunque se fundamentan en principios matemáticos diferentes, por lo que presentan varias diferencias.

Los métodos de filtrado espectral se fundamentan en la descomposición en vectores y valores propios de la matriz laplaciana de un grafo [44]. Esto se conoce como descomposición espectral de un grafo [45]. Esta descomposición se utiliza para definir un operador de convolución. Los métodos de filtrado espectral son robustos, fiables y tienen un amplio recorrido en la bibliografía. Sin embargo son métodos matemáticos complejos que conllevan costes computacionales más elevados. Además son dependientes de la base del grafo utilizado lo que hace que la generalización de estos modelos sea aún objeto de estudio.

Por otra parte, los métodos de filtrado espacial se basan en que dentro de un modelo de GNN, el aprendizaje de las características que definen a

un grafo depende del vecindario de cada vértice. De forma análoga a las CNN clásicas, un función no lineal se aplica a todos los nodos del grafo. La elección de esta varía en función de la tarea. Los métodos de filtrado espacial son menos complejos que los de filtrado espectral, pero no por ello menos potentes. La principal diferencia radica en que los métodos espaciales combinan los vectores de características de los vecindarios de vértices basándose en la topología del grafo y considerando su estructura espacial.

2.3.5. Datos de múltiples vistas

Los datos 3D pueden representarse como una combinación de múltiples imágenes 2D generadas como diferentes puntos de vista del objeto 3D [46].

Esta representación permite el aprendizaje de múltiples conjuntos de características para reducir posibles efectos de ruido, ocultación, estado incompleto de los datos y cambios de iluminación generados por el proceso de captura de los datos. Además, permite trabajar con entradas de alta resolución, pues pueden aplicarse modelos DL basados en imágenes 2D. El aprendizaje de datos 3D a partir de múltiples imágenes 2D del objeto, aspira a aprender una función capaz de modelar cada vista por separado para después optimizar conjuntamente todas las funciones aprendidas de forma que pueda representarse el objeto 3D en su totalidad además de generalizar la representación de otros objetos 3D. Sin embargo, su mayor inconveniente es establecer el número de vistas suficientes para representar el objeto 3D. Un número de vistas demasiado bajo puede producir que no se capturen las propiedades del objeto 3D en su totalidad y además podría conllevar un problema de overfitting. Por otro lado, un número de vistas muy elevado podría incrementar innecesariamente el coste computacional. Además el uso de este tipo de representación no es capaz de preservar las características geométricas intrínsecas de los modelos 3D.

A modo de resumen, en la Figura 2.8 vemos varios ejemplos de las diferentes formas de representación de los datos 3D presentadas en esta sección.

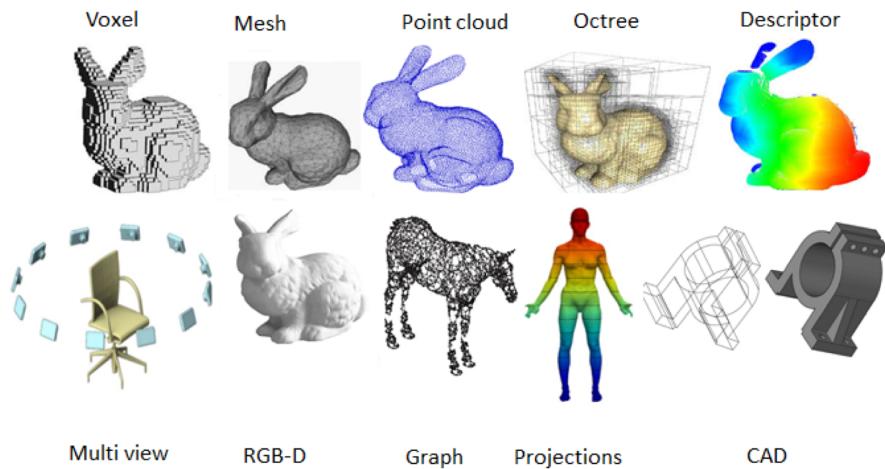


Figura 2.8: Diferentes formas de representación de los datos 3D [13].

Capítulo 3

Estado del arte

3.1. Estimación del perfil biológico

En el contexto de la AF y más concretamente en el desempeño de la tarea de la estimación del PB, existen dos escenarios posibles. Esta puede realizarse tanto para individuos fallecidos, a través de sus restos óseos, en casos de identificación de cadáveres; como para individuos vivos, en casos de personas desaparecidas o menores no acompañados que solicitan asilo fuera de su lugar de origen.

La estimación de la edad es un tema que ha ido tomando relevancia a lo largo del tiempo. En la Figura 3.1 podemos ver como el número de publicaciones relacionadas con esta tarea, enmarcada dentro de la AF, ha ido aumento año tras año hasta superar más de 100 publicaciones en los últimos años. Sin embargo, en dicha figura también podemos ver como el número de publicaciones relacionadas con la estimación de la edad y la AF que hagan mención a la inteligencia artificial o al aprendizaje automático no es un tema tan popular. Los métodos automáticos (métodos que hacen usos de técnicas de inteligencia artificial) han comenzado a surgir desde mediados de la década y actualmente no se supera la veintena de publicaciones sobre el tema. Por esto, se remarca la importancia de este proyecto, aplicar un nuevo enfoque a la estimación de la edad a partir de restos óseos de forma automática.

Para el desarrollo de esta sección, nos centraremos en el escenario de la estimación de la edad en individuos fallecidos a partir de sus restos óseos. Analizaremos el estado del arte de las técnicas actuales tanto manuales como automáticas. Además centraremos nuestra atención en los métodos que estiman la edad mediante el uso de sínfisis púbicas.

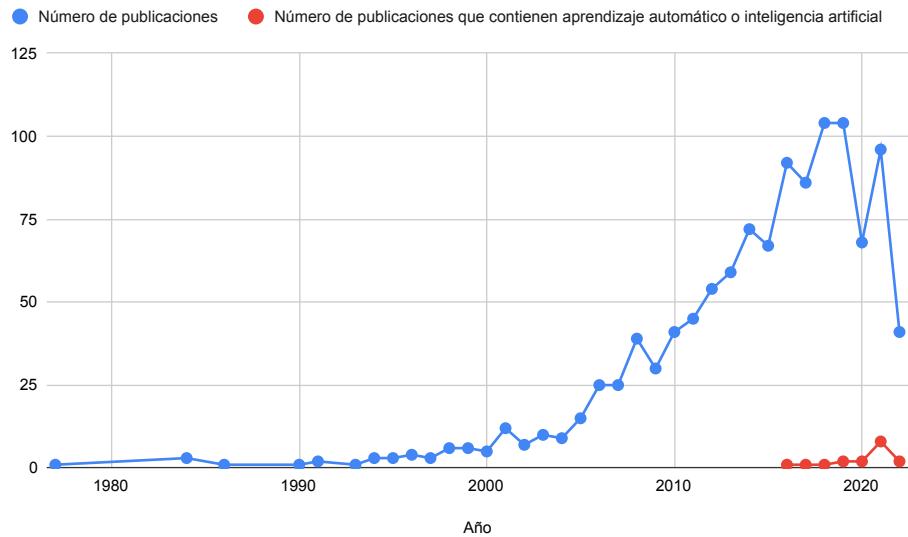


Figura 3.1: Número de publicaciones relacionadas con la estimación de la edad y la AF por año. **Azul**: cualquier publicación relacionada con este campo. **Rojo**: publicaciones que contienen referencias a aprendizaje automático o inteligencia artificial

3.1.1. Estimación de la edad a partir de restos óseos

Realizar esta estimación sobre restos humanos de individuos adultos es una tarea compleja. En la AF existen diferentes métodos aplicables en base a la evolución del proceso degenerativo normal de los restos [11]. Las regiones anatómicas más útiles para llevar a cabo la estimación son aquellas que no se ven afectadas por factores externos como el trabajo o la actividad física. Los métodos basados en el uso de sínfisis púricas [8, 10, 47], la superficie auricular del ilium [48] y el extremo esternal de la cuarta costilla [49] han mostrado una buena precisión en la estimación de la edad. El análisis del desarrollo y proceso degenerativo del hueso púbico es el método más extendido y esta reconocido como una de las alternativas más fiables [50, 51].

3.1.2. Estimación de la edad a partir de la sínfisis púrica

Los primeros estudios para la estimación de la edad (edad de muerte) mediante el análisis de restos de las sínfisis del pubis fueron los realizados por T. W. Todd en 1921 [8]. Todd analizó 360 esqueletos de varones caucásicos en un rango de entre 18 y 60 años para establecer 10 rangos de desarrollo y progreso degenerativo. Para establecer estas fases, Todd definió una descripción morfológica de la sínfisis púrica para cada edad.

Aunque existen críticas al método diseñado por Todd, aún 100 años después de la publicación original de [8], las características y criterios morfológicos descritos son utilizados a día de hoy como metodología recomendada.

Debemos distinguir dos tipos de enfoques en los métodos de estimación de la edad basados en el uso de las sínfisis pélvicas. Los métodos que proporcionan una estimación de la edad de la muerte en un intervalo o fase (métodos basados en fases) y los métodos que proveen una estimación de un valor específico para la edad de la muerte de un individuo (métodos numéricos). En función de como estos métodos estimen la edad, podemos diferenciar métodos que estiman un valor para cada rasgo o componente por separado y posteriormente combinan las diferentes estimaciones (métodos basados en componentes) o métodos que realizan una estimación de forma global.

Actualmente uno de los métodos más recomendados para la estimación de la edad a partir de sínfisis pélvicas es el propuesto por Suchey y Brooks [10]. Este método es una adaptación de la propuesta de Todd. Suchey y Brooks reducen el número de fases a 6 y modifican los rangos de edad, ampliándolos y superponiéndolos entre ellos. Sin embargo, la forma de aplicar el método seguía siendo prácticamente la misma que la propuesta por Todd: la evaluación subjetiva de las características morfológicas de la sínfisis pélvica y la asignación manual de una fase de edad de la muerte por parte de un antropólogo forense.

Como alternativa a los métodos basados en fases, Gilbert y McKern [9] fueron posiblemente los primeros en considerar un análisis independiente de cada uno de los componentes. Su propuesta se basaba en un sencillo sistema de regresión, que utilizaba una variable continua para la edad de la muerte, obtenida mediante la suma del valor asociado a cada componente por separado (p.ej. las características de la sínfisis pélvica). Este valor asociado a las componentes es asignado por un antropólogo forense tras una inspección visual. Aunque el método de Gilbert-McKern no funcionó correctamente, se convirtió en el punto de partida del criticismo hacia los métodos basados en fases por parte de muchos autores en las décadas siguientes.

A pesar de las adaptaciones propuestas sobre el método de Todd, este método y sus variantes como la de Suchey y Brooks y la de Gilbert y McKern se aplican de forma rutinaria de forma similar a la de la propuesta original. Esto significa que la valoración de los rasgos morfológicos de la sínfisis pélvica y su correspondencia con el estándar de cada fase es una valoración subjetiva y depende de la experiencia del antropólogo forense. Por lo tanto, la principal desventaja de estos métodos es la subjetividad presente tanto en la descripción morfológica de las sínfisis pélvicas como en la elección de la fase de desarrollo y proceso degenerativo a la que pertenece

[51]. Esta subjetividad se deriva de la falta de sistematización del método, el cual ofrece descripciones demasiado genéricas que dificultan la discriminación entre fases. En la Tabla 3.1 podemos apreciar un ejemplo de modelos 3D pertenecientes a cada una de las fases descritas por Todd.

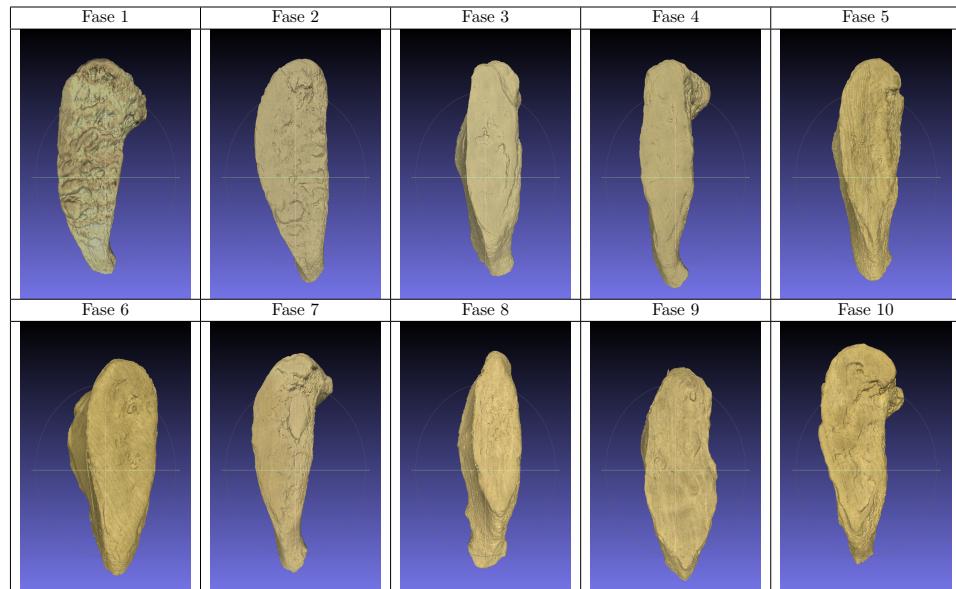


Tabla 3.1: Ejemplos de modelos 3D de sínfisis púlicas pertenecientes a cada una de las fases de Todd.

Los métodos de estimación de la edad a partir de restos óseos deberían ofrecer el mismo resultado independientemente del experto que lo aplique. Sin embargo, numerosos estudios han demostrado que la efectividad de estos métodos dependen casi exclusivamente de la experiencia del antropólogo forense [52, 53]. Ante esta situación, es necesario utilizar métodos que reduzcan la subjetividad de los procesos de estimación de la edad. El uso de técnicas de IA como algoritmos de ML y DL o inteligencia artificial explicable (XAI) en la AF puede beneficiar este área ampliando los métodos existentes y extendiendo su aplicación desde expertos antropólogos a practicantes.

Diferentes métodos por ordenador se han considerado con el propósito de objetivizar (y automatizar en algunos casos) la estimación de la edad de la sínfisis del pubis. Los principales avances han ido encaminados a sustituir el estudio directo del hueso por el estudio de modelos 3D digitalizados, ya sea utilizando nuevos criterios o los criterios utilizados en los métodos originales [54].

Dentro de los métodos basados en fases encontramos varias propuestas

presentadas por equipos de investigación de la Universidad de Granada.

En [55], Villar et al. modelan el problema como una tarea de clasificación de ML clásico con 10 fases independientes de edad. Los principales rasgos morfológicos del hueso descritos por Todd fueron analizados y modelados como nueve variables lingüísticas basadas en el conocimiento de expertos forenses. Dos antropólogos forenses distintos etiquetaron manualmente un pequeño conjunto de sínfisis púbicas de 74 individuos. Se obtuvieron bases de reglas compactas compuestas por entre 17 y 20 reglas utilizando árboles de decisión difusos. Estos reportaron un error medio absoluto (MAE) ordinal de 1.07 en todo el conjunto y 1.68 en el conjunto de test mediante validación cruzada en 10 iteraciones. Sin embargo las reglas derivadas no cubrían todas las fases, por lo que el proceso de validación no era totalmente fiable debido a las limitaciones del conjunto de datos (pequeño y no equilibrado) y el enfoque de ML seguido.

En [56], Gámez-Granados et al. diseñan un sistema explicable y basado en reglas para estimar la edad de la muerte a partir de sínfisis púbicas evocando el conocimiento del método de Todd y su procedimiento. Este método considera nueve características del hueso púbico basadas en las descripciones de Todd y extraídas por antropólogos forenses. Dado que el conjunto de datos utilizado estaba fuertemente desbalanceado, consideran el de sobre-muestreo de datos y un algoritmo de aprendizaje de reglas de clasificación ordinal (NSLVOrd [57]). El modelo entrenado aporta un error cuadrático medio (RMSE) de 12,34 años y un MAE de 10,38 años en una muestra con 892 huesos del pubis, considerando los valores intermedios de los rangos de edad de las fases de Todd. Hasta el momento es el mejor resultado del estado del arte en cuanto a métodos semi-automáticos basados en fases.

Por otro lado dentro de los métodos que estiman directamente la edad de la muerte, encontramos diversos autores que consideran diferentes técnicas de regresión y técnicas de inteligencia artificial para desarrollar nuevos métodos.

En [58], Slice y Algee-Hewitt introducen SHA, un índice para cuantificar la variación de la morfología de la superficie de la sínfisis púbica asociada al envejecimiento en el método de Todd. Se calcula a partir de un análisis de componentes principales (PCA) de los modelos 3D escaneados por láser. SHA se utiliza para diseñar un modelo de regresión lineal evaluado en un estudio que considera 41 esqueletos de varones americanos. [58] obtiene un rendimiento similar a Suchey-Brooks [10], reportando RMSE de 17,15 años.

Más adelante en [59], Stoyanova et al. diseñan otro modelo lineal que estima la edad a partir de una única variable extraída automáticamente

a partir de los modelos 3D de las sínfisis del pubis, la energía de flexión (bending energy). Este enfoque se prueba con 44 esqueletos escaneados de varones blancos documentados y 12 moldes en representación de las fases de Suchey-Brooks [10] en el rango de entre 16 y 100 años, reportando un RMSE de aproximadamente 19 años. En [60] amplían su anterior trabajo extrayendo nuevas características en base a la curvatura del modelo 3D e integrándolas todas en dos modelos de regresión multivariante. Este trabajo se evalúa en un conjunto de 93 modelos 3D de varones blancos entre 16 y 90 años (68 reales y 25 moldes) obteniendo un RMSE de entre 13,7 y 16,5.

En [61], Koterova et al. proponen nueve métodos de regresión automática para la estimación de la edad. En su propuesta, combinan tres rasgos de la sínfisis pública con cuatro de la superficie auricular del ilium. Estos métodos se validan en un conjunto de datos de 941 individuos multiétnicos de adultos entre 19 y 100 años. Se consideraron enfoques de regresión clásica, K-NN vecinos cercanos, modelos Bayesianos, arboles de regresión y redes neuronales. La regresión multilineal fue el mejor de los predictores, aportando un RMSE de 12,1 años y un MAE de 9,7 años.

En [51], Dudzik y Langley se centran en la incertidumbre que surge de las definiciones de los rasgos óseos y la dificultad de los antropólogos para entenderlas. Se analizan las características óseas en las etapas previas a los cambios degenerativos avanzados en 237 individuos americanos. La regresión logística multinomial y los arboles de decisión se aplican sobre una muestra de 47 individuos jóvenes en el rango de edad de entre 18 y 40 años, utilizando cinco rasgos de la sínfisis pública. Reportan un 94% de precisión, pero considerando solo 3 fases distintas.

En [62], Bermejo et al. proponen un nuevo enfoque semi-automático basado en componentes a partir de los rasgos morfológicos de las sínfisis púbicas propuestos por Todd. Este método se basa en la obtención de en una expresión matemática que combina los valores de los distintos componentes, proporcionados por el experto forense, en un valor numérico para la edad de la muerte. Se consideran métodos de ML explicables, haciendo uso de la programación genética para derivar dichas fórmulas a partir de los datos de la base de datos pública disponible en el Laboratorio de Antropología Física de la Universidad de Granada. Se sigue un enfoque de ciencia de datos guiado por la teoría [63] que combina la experiencia del antropólogo forense y las descripciones explicables derivadas de los datos. Se proponen dos enfoques, uno aplica técnicas de programación genética (GP) y el otro combinando esto con algoritmos genéticos (GA-P). Ambos enfoques aportan resultados de RMSE 10,81 y MAE 8,55.

Finalmente en [64], Castillo et al. proponen un nuevo método basado en

fases y en componentes diseñado utilizando métodos de ML. Con el objetivo de manejar tanto la variación interindividual como la subjetividad de los sistemas de puntuación visual, su marco considera 16 rasgos diferentes de la síntesis pública, incluyendo uno novedoso que no ha sido utilizado previamente (microranuras). Todos ellos sólo presentan dos valores categóricos (presente o ausente) para facilitar la tarea de etiquetado al antropólogo forense. Se aplica un método de selección de características sobre cada uno de las 5 fases de edad de la muerte analizadas para seleccionar las mejores características considerando para ello tres algoritmos diferentes de ML (ZeroR, Naïve Bayes y Random Forest). Los autores muestran resultados muy prometedores para algunas fases de edad así como una discriminación entre características útiles e inútiles.

En la Tabla 3.2 podemos ver un resumen de los resultados presentes en el estado del arte introducidos en esta sección.

Método	Tipo	N muestras	Rango de edad	RMSE	MAE
Slice y Algee-Hewitt [58]	N, CS	41	19 - 96	17,15	-
Stoyanova et al. [59]	N, CS	56	16 - 100	19	-
Stoyanova et al. [60]	N, CS	93	16 - 90	13,7 - 16,5	-
Koterova et al. [61]	N, CS	941	19 - 100	12,1	9,7
Gámez-Granados et al. [56]	PB, G	892	18 - 60	13,19	10,38
Gámez-Granados et al. [56]	PB, G	960	18 - 60	14,61	11,62
Bermejo et al [62] (GP)	N, CS	1152	18-82	10,82	8,56
Bermejo et al [62] (GA-P)	N, CS	1152	18-82	10,81	8,55

Tabla 3.2: Resumen de los resultados presentes en el estado del arte de los métodos de estimación de la edad. **Tipo:** PB, basado en fases; N, numérico; CS, basado en componentes; G, global. **MAE** y **RMSE** indican el error medido en años.

3.2. Representación de datos 3D

En la Sección 2.3 se han introducido todas las opciones para representar datos 3D utilizadas actualmente en el paradigma del DL. En esta sección haremos un análisis de cuáles son las más adecuadas para la utilización en la estimación de la edad mediante el uso de síntesis públicas.

En primer lugar descartamos la representación RGB-D. Dado que los modelos de los que se dispone son modelos 3D completos especificados como una lista de vértices y aristas, se produciría una pérdida de información geométrica al utilizar esta representación. Descartaremos también el uso de representaciones volumétricas (vóxeles y octrees). Este tipo de representaciones se centran en proporcionar información sobre como el modelo 3D ocupa

el espacio. Los vóxeles requieren un uso elevado de memoria para representar datos en alta resolución. Los octrees por otro lado, aunque no tienen el problema de usar demasiada memoria, no pueden capturar correctamente propiedades del modelo 3D como la morfología de la superficie. Dado que esta propiedad puede proporcionar información fundamental de la que se podrán extraer características discriminatorias, la representación con vóxeles y octrees de los modelos 3D conllevaría grandes pérdidas de información.

Por tanto las posibles opciones se reducen a: proyecciones 3D, nubes de puntos, mallas 3D, descriptores 3D, grafos y múltiples vistas. Descartamos los descriptores 3D debido a que los resultados obtenidos dependerán de la capacidad del descriptor escogido de codificar correctamente la información geométrica de los modelos 3D. Dado que los modelos 3D empleados poseen un alto nivel de detalle, sería necesario utilizar un descriptor muy potente. Además, dado que las características que permiten estimar la edad de las síntesis pública se encuentran bien definidas [8], se ha decidido optar por un modelo que no dependa de una representación codificada. De esta forma el modelo podrá extraer las características directamente de los datos y estas podrán compararse con las definidas por Todd.

Descartamos también las nubes de puntos dado que estas son similares a la representación de grafos y mallas pero eliminando el componente de conectividad de los puntos. Por esto, si utilizamos esta representación se estaría eliminando información útil necesaria para el proceso de estimación de la edad. Finalmente descartamos la representación en múltiples vistas, pues aunque estas codifican información de modelos 3D, el objetivo de este proyecto es trabajar directamente con los datos 3D como entrada del modelo.

Descartadas estas opciones solo nos quedan las proyecciones 3D, la representación en grafos y la representación en mallas 3D. Estas últimas son muy similares entre si, dado que una malla 3D puede verse también como un grafo no dirigido $G = (V, E)$ tal que V está formada por los vértices de la malla y E por las aristas. Se barajó el uso de modelos de geometric deep learning (GDL), que utilizan mallas 3D como entrada de datos. Sin embargo, estos tienen grandes inconvenientes. Los ejemplos disponibles en la literatura actual trabajan con mallas 3D formadas por un número limitado de polígonos. Dado que los datos disponibles para entrenamiento, cuentan con una media de 1.000.000 de polígonos¹. Para poder procesar modelos con una definición tan alta, sería necesario calcular una malla simplificada para cada modelo 3D, con menor número de polígonos. Esto conllevaría una perdida de información geométrica del modelo, pero de no realizarse, sería imposible entrenar el modelo con el hardware disponible. Por ello, la representación

¹ La media de polígonos de los datos de entrenamiento es 1.001.507,08

en mallas 3D/grafos queda descartada.

Finalmente, queda como última opción la representación haciendo uso de proyecciones 3D. Esta representación puede ser muy versátil, ya que su aplicación es compatible con el uso de modelos y arquitecturas de DL clásico que han mostrado ser muy competitivas a la hora de resolver tareas de clasificación y regresión. Procedamos entonces a estudiar y analizar las representaciones basadas en proyecciones 3D presentes en la literatura.

3.2.1. Representaciones basadas en proyecciones 3D

Zhu et al. [65] proponen uno de los primeros intentos de aprender de datos 3D proyectándolos sobre superficies 2D. El proceso propuesto comienza con un preprocesamiento de datos en el que se aplica la traslación, el escalado y la normalización de la postura a cada modelo 3D. A continuación, se aplican varias proyecciones 2D a cada modelo 3D procesado para alimentar una pila de máquinas de Boltzmann restringidas (RBM) para extraer las características de las diferentes proyecciones. Los experimentos muestran que este sistema se comporta mejor que las técnicas basadas en descriptores globales.

Shi et al. [66] proponen DeepPano. DeepPano se refiere a la extracción de vistas panorámicas 2D de objetos 3D empleando una proyección cilíndrica alrededor del eje principal del objeto 3D. Para entrenar el modelo se utiliza una arquitectura CNN clásica 2D. El modelo propuesto se prueba en tareas de reconocimiento y recuperación de objetos en 3D donde demuestra su eficacia en comparación con los modelos anteriores.

Sinha et al. en [67] proponen imágenes geométricas en las que los objetos 3D se proyectan en una cuadrícula 2D para poder emplear las CNN clásicas 2D. El método propuesto crea una parametrización planificadora para los objetos 3D utilizando una parametrización auténtica (que conserva el área) en un dominio esférico para aprender las superficies de las formas 3D. A continuación, las imágenes geométricas construidas se utilizan como entradas de una arquitectura clásica de CNN para aprender las características geométricas de los objetos 3D. Los resultados muestran que las imágenes geométricas pueden producir resultados comparables con el estado del arte.

Cao et al. [68] proponen utilizar una proyección de dominio esférico para proyectar objetos 3D alrededor de su baricentro produciendo un conjunto de parches cilíndricos. El modelo propuesto utiliza los parches cilíndricos proyectados como entrada para CNNs pre-entrenadas. También se utilizan dos proyecciones complementarias para capturar mejor las características 3D. El modelo propuesto se utiliza para la tarea de clasificación de objetos 3D y

se prueba en múltiples conjuntos de datos, produciendo resultados comparables a los de los métodos anteriores.

Sfikas et al. [69] proponen representar objetos 3D como vistas panorámicas extraídas de objetos 3D con poses normalizadas. Inicialmente, los objetos 3D son preprocesados para normalizar sus poses utilizando el marco de “Pose Normalization of 3D Models via Reflective Symmetry on Panoramic Views (SymPan)” [70]. A continuación, se extraen las vistas panorámicas para combinarlas y alimentar la CNN para realizar tareas de clasificación y recuperación de pose. Una extensión de este modelo se introduce en [71] donde se utiliza un conjunto de CNNs para el proceso de aprendizaje. Esta extensión produce muy buenos resultados en los conjuntos de datos utilizados.

En la Tabla 3.3 podemos ver un resumen de los resultados presentes en el estado del arte introducidos en esta sección. Los resultados han sido probados en el conjunto de datos ModelNet [72], un conjunto a gran escala de modelos 3D de objetos cotidianos (en su versión de 10 y 40 clases), para resolver un problema de clasificación.

Método	Modelo DL	Dataset	Precisión
Shi et al. [66]	CNN	ModelNet10	85,45 %
		ModelNet40	77,63 %
Sinha et al. [67]	CNN	ModelNet10	88,4 %
		ModelNet40	83,9 %
Cao et al. [68]	CNN	ModelNet40	94,24 %
Sfikas et al. [69]	CNN	ModelNet10	91,1 %
		ModelNet40	90,7 %
Sfikas et al. [71]	CNN Ensemble	ModelNet10	96,85 %
		ModelNet40	95,56 %

Tabla 3.3: Resumen de los resultados presentes en el estado del arte de los métodos de representación de datos 3D mediante proyecciones.

Capítulo 4

Métodos propuestos

4.1. Representación PANORAMA

Como vimos en la Tabla 3.3, el mejor método presente en el estado del arte para la representación de modelos 3D mediante proyecciones es el propuesto por Sfikas et al. [71]. Por esto, se ha elegido esta representación como forma de representación de datos 3D para el desarrollo de este proyecto.

La representación PANORAMA presentada en 2010 en el artículo **PANORAMA: A 3D Shape Descriptor Based on Panoramic Views for Unsupervised 3D Object Retrieval** [73] destaca como una de las representaciones 3D basadas en proyecciones que mejores resultados aporta en tareas de clasificación/reconocimiento [12, 69, 71].

No fue hasta el año 2017 que se presentó un artículo [69] en el que esta representación se utilizó como forma de representación de datos 3D para su posterior uso en CNN. En 2018 se publicó un artículo que expandió el planteamiento anterior [71]. A continuación se detallará en qué consiste la representación PANORAMA y su uso en el campo de las CNN.

4.1.1. Propuesta original de PANORAMA

La representación PANORAMA está formada por un conjunto de varias vistas panorámicas. Para obtener cada una de ellas, proyectamos el modelo en la superficie lateral de un cilindro de radio R y altura $H = 2R$, centrado en el origen con uno de sus ejes paralelo a uno de los ejes de coordenadas (ver Figura 4.1).

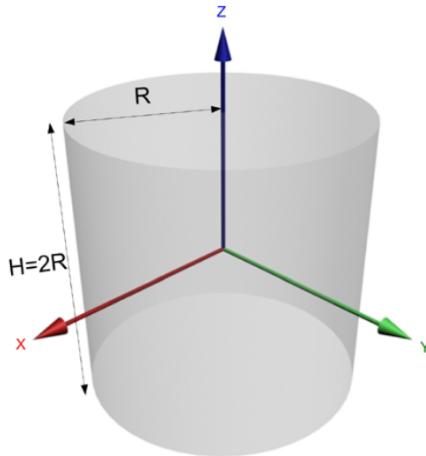


Figura 4.1: Cilindro utilizado para adquirir la proyección de un modelo 3D [73].

Establecemos $R = 2 \cdot d_{max}$ ¹, siendo d_{max} la distancia máxima de la superficie del modelo hasta su centroide.

Una vez definido el cilindro que contendrá y sobre el que se proyectará el modelo 3D, es necesario parametrizar su superficie lateral haciendo uso del conjunto de puntos $s(\varphi_u, y_v)$ donde $\varphi \in [0, 2\pi]$ es el ángulo en el plano xy , $y \in [0, H]$ y muestreamos las coordenadas φ e y a razón de $2B$ y B , respectivamente (fijamos $B = 180$)². La dimensión de φ se muestrea al doble de frecuencia para contabilizar la diferencia en longitud entre el perímetro de la superficie lateral del cilindro y su altura. Aunque el perímetro de esta superficie es $2\pi \approx 3$ veces la altura del cilindro, la frecuencia de muestreo se fija a $2B$ dado que se ha comprobado experimentalmente que produce el mejor resultado en términos de efectividad-eficiencia.

En resumen, establecemos un conjunto de puntos $s(\varphi_u, y_v)$ donde $\varphi_u = u * 2\pi/(2B)$, $u \in [0, 2B - 1]$ y $y_v = v \cdot H/B$, $v \in [0, B - 1]$ (ver Figura 4.2). Una vez definido $s(\varphi_u, y_v)$, hay que calcular un valor para cada punto. Este cálculo se realiza de forma iterativa para $v = 0, 1, \dots, B$ considerando un conjunto de puntos coplanares de s para cada valor de v (cada conjunto de puntos tiene origen en el centro del cilindro c_v , ver Figura 4.3).

¹ La propuesta original establece $R = 3 \cdot d_{media}$ donde d_{media} es la distancia media de la superficie del modelo hasta su centroide. Este valor se calcula originalmente aplicando un método de normalización de pose diferente, por lo que se han reflejado los cambios propuestos en [69].

² La propuesta original fija $B = 64$

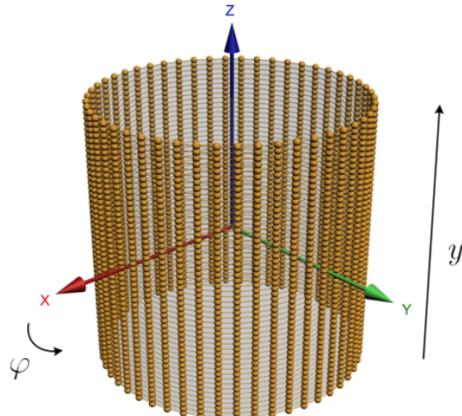


Figura 4.2: Discretización de la superficie lateral del cilindro de proyección (puntos en naranja) al conjunto de puntos $s(\varphi_u, y_v)$ [73].

Las proyecciones cilíndricas capturan dos características de la superficie del modelo 3D:

- La posición de la superficie del modelo en el espacio 3D. Nos referiremos a esta proyección como **Spatial Distribution Map (SDM)**.
- La orientación de la superficie del modelo. Nos referiremos a esta proyección como **Normals' Deviation Map (NDM)**.

Para capturar estas características se utilizan dos proyecciones cilíndricas distintas, que denotaremos como los conjuntos de puntos s_1 y s_2 . Por defecto, el valor inicial de cada punto es 0. Estas proyecciones también pueden verse como mapas de características.

Para calcular la posición de la superficie del modelo 3D, para cada sección transversal a la altura y_v , calculamos la distancia desde c_v hasta la intersección de la superficie del modelo con cada uno de los rayos φ_u en cada dirección (ver Figura 4.3). Definamos $pos(\varphi_u, y_v)$ como la distancia de c_v al punto más lejano consecuencia de la intersección del rayo con origen c_v y dirección φ_u con la superficie del modelo, por tanto:

$$s_1(\varphi_u, y_v) = pos(\varphi_u, y_v) \quad (4.1)$$

El valor de cada punto de s_1 pertenece al intervalo $[0, R]$.

Para capturar la orientación de la superficie del modelo, para cada sección transversal a la altura y_v , calculamos la intersección de la superficie del modelo con el rayo con origen c_v y dirección φ_u . Una vez calculada la intersección, calculamos el ángulo entre el rayo y el vector normal del

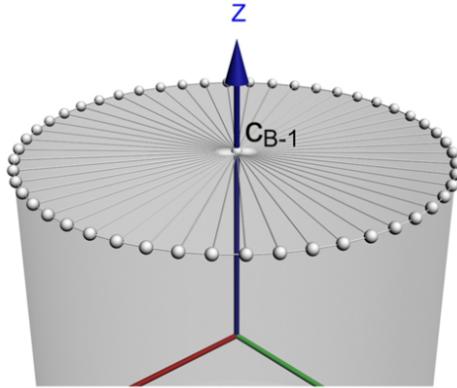


Figura 4.3: Sección transversal más alta del cilindro junto con los rayos correspondientes que emanan del centro de la sección transversal c_{B-1} , ($v = B - 1$) [73].

triangulo intersecado. Definimos $\text{ang}(\varphi_u, y_v)$ como el coseno del ángulo entre el rayo con origen c_v y dirección φ_u y el vector normal del triángulo al que pertenece el punto más lejano consecuencia de las intersección del rayo, por tanto:

$$s_2(\varphi_u, y_v) = |\cos(\text{ang}(\varphi_u, y_v))|^n \quad (4.2)$$

Tomamos la potencia enésima de $|\cos(\text{ang}(\varphi_u, y_v))|^n$ donde $n \geq 2$ debido a que así aumentamos el contraste de la proyección obtenida. Esto se ha demostrado experimentalmente como una mejora en la capacidad de discriminación de la superficie del modelo en la proyección. Los mejores resultados los aporta $n \in [4, 6]$, aumentando el contraste de forma proporcional a n . Además calculamos el valor absoluto del coseno para no tener que prestar atención a la orientación de los triángulos en la superficie del modelo. No es necesario aumentar el contraste en los valores de s_1 ya que esta proyección produce valores menos discriminatorios [73].

En la Figura 4.4 se puede ver un ejemplo de las proyecciones que hemos definido previamente.

Una de las limitaciones de estas representaciones, es que no son capaces de representar la concavidad de los modelos 3D. Para superar esta limitación, se calcula cada una de las proyecciones para un cilindro alineado con cada uno de los tres ejes de coordenadas X, Y y Z. Un ejemplo de esto puede verse en la Figura 4.5.

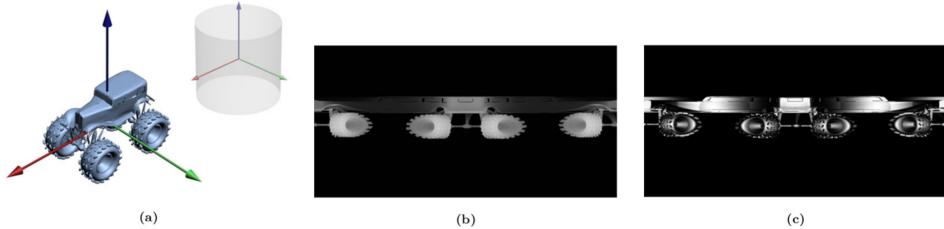


Figura 4.4: (a) Modelo 3D. (b) SDM, la proyección cilíndrica desplegada que captura la posición de la superficie modelo 3D. (c) NDM, la proyección cilíndrica desplegada que captura la orientación de la superficie modelo 3D[73].

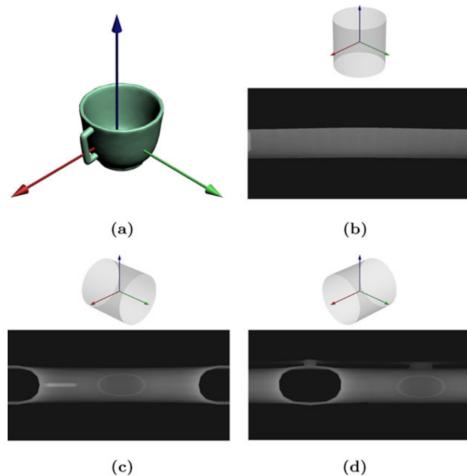


Figura 4.5: (a) Modelo 3D de una taza. (b)-(d) las correspondientes proyecciones cilíndricas $s_{1,t}(\varphi_u, y_v)$ utilizando tres cilindros alineados cada uno con los ejes de coordenadas X, Y y Z respectivamente[73].

4.1.2. Propuesta de adaptación

Una vez revisado el método de representación PANORAMA, veamos como puede aplicarse al campo de las redes neuronales convolucionales. Primero nos centraremos en describir las dos propuestas de aplicación de PANORAMA presentadas por el equipo desarrollador de la representación.

Para poder entrenar una CNN utilizando la representación PANORAMA, en [69] se propone un esquema aumentado que se basa en las representaciones panorámicas producidas por cada eje de coordenadas. Por cada eje, se computa tanto las representaciones SDM como NDM del modelo 3D. Cada mapa de características se amplia añadiendo la mitad de la imagen al final de la misma. De esta forma evitamos que haya lagunas en la represen-

tación. En total se computan seis vistas panorámicas: $NDM(X)$, $NDM(Y)$, $NDM(Z)$, $SDM(X)$, $SDM(Y)$ y $SDM(Z)$. Estas seis vistas forman la entrada de la red neural convolucional. El tamaño total de las vistas del modelo 3D es $1,5 \cdot 360 = 540$ píxeles de ancho y $180 \cdot 6 = 1080$ píxeles de alto (ver Figura 4.6).

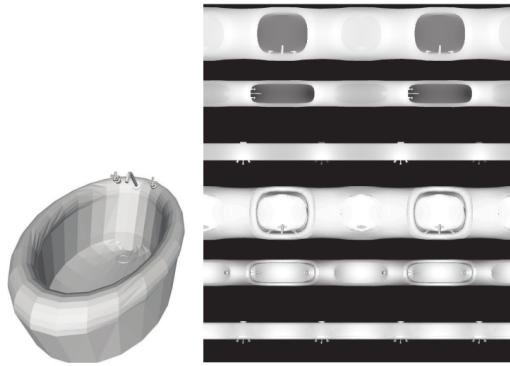


Figura 4.6: Ejemplo de vista panorámica aumentada de un modelo 3D. El orden de las representaciones apiladas es el siguiente: $NDM(X)$, $NDM(Y)$, $NDM(Z)$, $SDM(X)$, $SDM(Y)$ y $SDM(Z)$ [69].

El segundo enfoque, presentado en [71] es el siguiente. Se computan las representaciones SDM y NDM para cada uno de los ejes principales del modelo 3D. Además se computa el gradiente de la representación NDM . A esta nueva representación la llamaremos Gradient NDM ($GNDM$). De igual forma que en el enfoque anterior, la representación PANORAMA se extiende añadiendo la mitad de la imagen al final de la misma. Por cada modelo 3D, se obtiene un total de tres vistas panorámicas, una por cada eje de coordenadas. Cada vista, se apila en una imagen RGB de tres canales de información de la siguiente forma: $NDM(X) - SDM(X) - GNDM(X)$, $NDM(Y) - SDM(Y) - GNDM(Y)$, $NDM(Z) - SDM(Z) - GNDM(Z)$. Estas tres imágenes forman la entrada de la CNN. El tamaño total de las vistas del modelo 3D es $1,5 \cdot 720 = 1080$ píxeles de ancho y $360 \cdot 3 = 1080$ píxeles de alto (ver Figura 4.7).

Vistas las propuestas originales, para el desarrollo de este proyecto se proponen dos representaciones distintas. La primera de ellas aplica directamente el enfoque presentado en [71]. La entrada de la red neuronal convolucional serán tres imágenes RGB que codifican la información de cada uno de los mapas de características (SDM , NDM y $GNDM$) en un canal (ver Figura 4.8). La segunda utiliza un enfoque híbrido entre [69] y [71]. Se generan tres imágenes distintas (de un solo canal), una por cada mapa de características. Cada una de las imágenes contiene la representación de los

tres ejes concatenados de forma vertical (ver Figura 4.9).

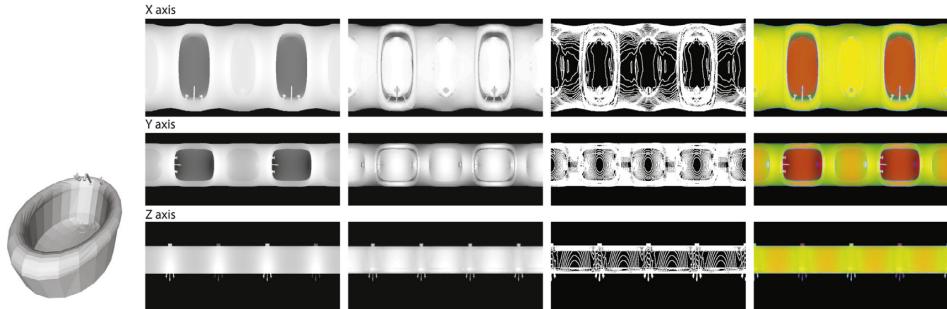


Figura 4.7: Ejemplo de imagen RGB (SDM-NDM-GNDM). De izquierda a derecha: Modelo 3D, representación SDM, representación NDM, representación GNDM, [71].

Ambas propuestas de representación generan tres imágenes por cada modelo 3D del conjunto de datos. Estas imágenes conformarán la entrada de tres CNN idénticas que procesarán las imágenes por separado. En la Sección 4.2 se comentará con detalle la composición de estas CNN.

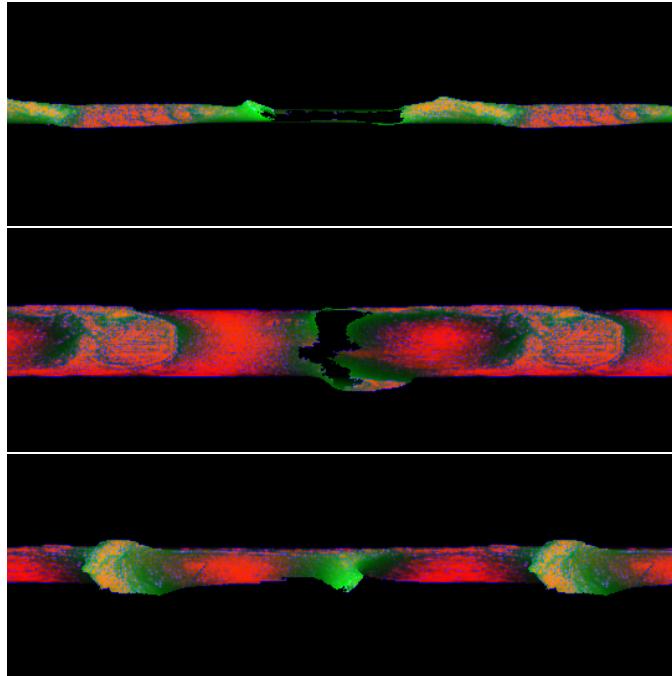


Figura 4.8: Primera propuesta de representación aplicada a una muestra del conjunto de datos.

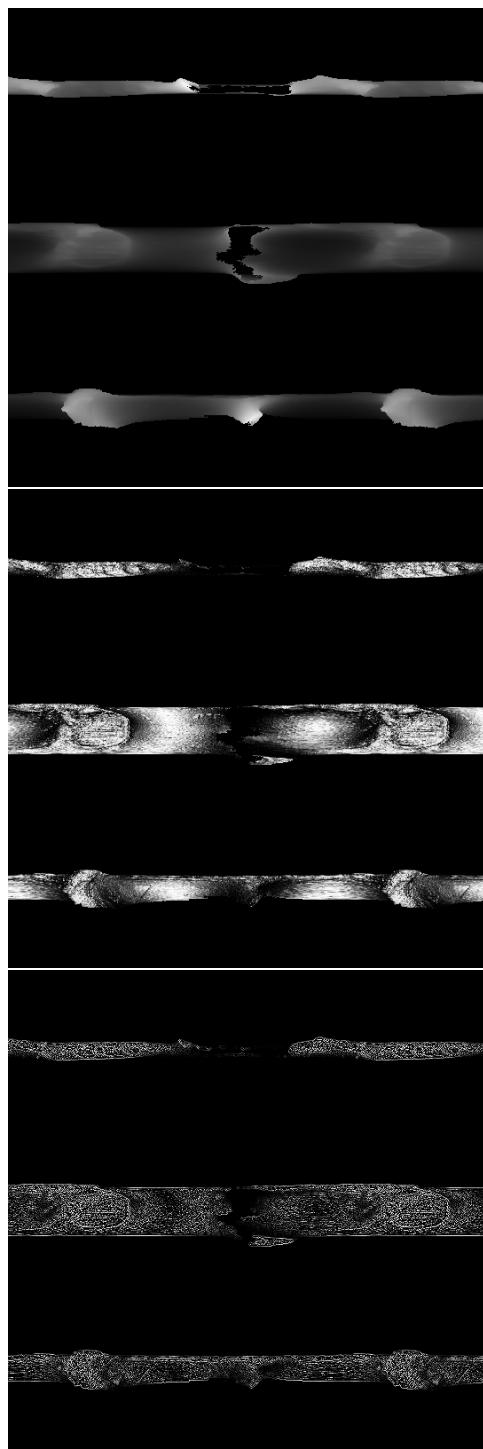


Figura 4.9: Segunda propuesta de representación aplicada a una muestra del conjunto de datos.

4.2. Arquitecturas de red

Una vez revisada la representación 2D de los datos que se utilizará, pasemos a describir las CNN que recibirán estos datos como entrada.

En este proyecto se definen 2 arquitecturas de red diferentes. Una de ellas aplica directamente la propuesta de los desarrolladores de PANORAMA y es aplicable para las dos representaciones presentadas. Mientras tanto, la otra es una arquitectura de red que se basa en aprovechar el concepto de “transfer learning” o transferencia de conocimiento, que consiste en utilizar un modelo de red entrenado previamente a resolver un problema parecido para adaptarlo al problema actual.

Panorama-CNN

Esta propuesta de CNN es una aplicación directa de la propuesta de CNN presentada en [69, 71]. La red esta compuesta por un modulo convolucional que se repite varias veces y una parte final que se encarga de realizar la regresión. El módulo convolucional esta formado por una capa convolucional, una capa de activación ReLU y una capa de max pooling. Este módulo se repite tres veces con un tamaño de 64, 256 y 1024 mapas de activación (profundidad). El tamaño de los filtros es 5, 5 y 3 respectivamente y se aplica un zero-padding de 2 a la entrada de cada capa convolucional para evitar que el resultado reduzca su tamaño. La capa de max pooling esta configurada con un tamaño de salto de 2x2, lo que implica que la imagen se reduce a la mitad en cada salida del bloque convolucional. Posterior a los tres módulos convolucionales, se conectan dos capas totalmente conectadas con un tamaño de 100 neuronas, una capa de dropout para ayudar a reducir el overfitting y finalmente una capa totalmente conectada de tamaño 1, que es la salida final del modelo. Todas las capas totalmente conectadas tienen como función de activación la función ReLU. En las Figuras 4.10 y 4.11 podemos ver un esquema de esta arquitectura de red.

Definimos tres redes Panorama-CNN y a cada una se le proporciona una imagen como entrada. Esta entrada pueden ser o bien las imágenes RGB descritas en la primera propuesta (Figura 4.8) o bien los mapas de características por separado descritos en la segunda propuesta (Figura 4.9). La salida de estas tres redes se combina haciendo la media aritmética de las tres estimaciones de edad y esa es la estimación final del modelo.

Resnet-CNN

Esta propuesta de CNN se basa en el uso del “transfer learning”. En la Figura 4.12 puede verse de forma esquemática en que consiste este método.

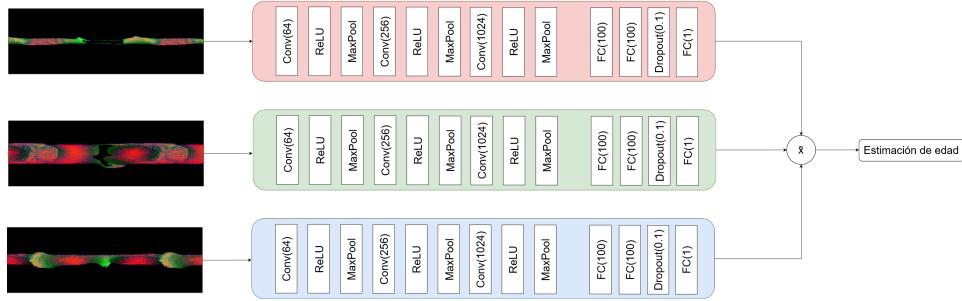


Figura 4.10: Red Panorama-CNN con entrada de imágenes RGB.

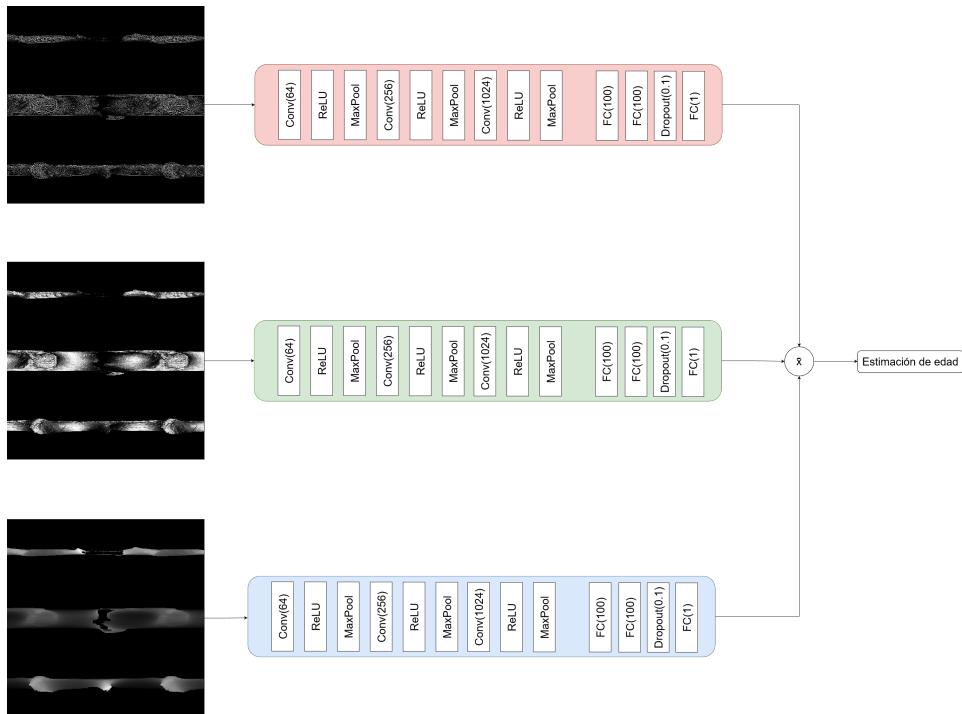


Figura 4.11: Red Panorama-CNN con entrada de mapas de características.

Como base de la red se utiliza una arquitectura predefinida y pre-entrenada para resolver otro problema distinto. Sin embargo, puede utilizarse este conocimiento previo que ha adquirido la red como punto de partida para comenzar un nuevo entrenamiento específico al problema de estimación de la edad. Este proceso mejora los resultados obtenidos en problemas que sufren de falta de datos de entrenamiento [74], como es el caso de este problema.

La red ResNet (Residual Network) es una arquitectura de CNN basada en aprendizaje residual. Fue presentada por Kaiming He et al. en [75]. Esta red originalmente fue propuesta como una alternativa de CNN pro-

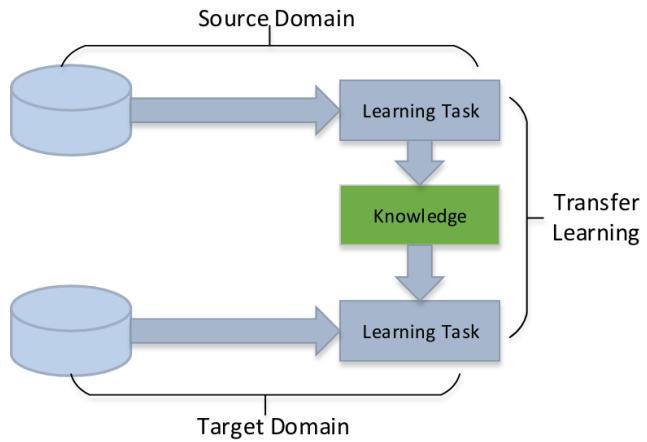


Figura 4.12: Esquema del funcionamiento de la técnica de transfer learning.

funda para el reconocimiento de imágenes. Se propone una arquitectura de red basada en la filosofía de las redes VGG [76], que está compuesta por bloques convoluciones con filtro de tamaño 3x3 (en su mayoría). Esta arquitectura sigue dos principios de diseño: para el mismo tamaño del mapa de características de salida, las capas tienen el mismo número de filtros y si el tamaño del mapa de características se reduce a la mitad, el número de filtros se duplica para preservar la complejidad temporal. Además a esta propuesta se añade conexiones directas entre capas para convertir a la red en una red neuronal residual (ver Figura 4.14).

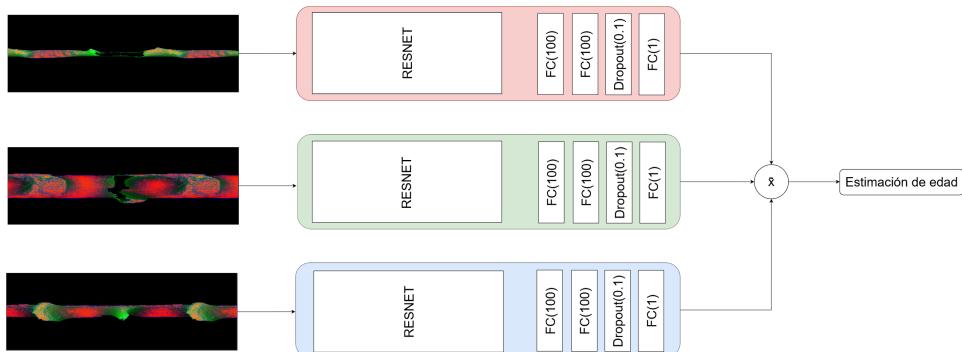


Figura 4.13: Red Resnet-CNN.

Concretamente la versión utilizada es ResNet50, que no es más que la arquitectura de ResNet descrita formada por 50 capas de parámetros. Esta red ha sido previamente entrenada para la clasificación de Imagenet [77]. Éste es un conjunto de datos de imágenes pertenecientes a 1000 clases distintas. Dado

que ResNet estaba originalmente pensada para clasificación, para poder adaptarla a un problema de regresión, se ha eliminado la última capa softmax, encargada de la clasificación, y se ha sustituido por el bloque totalmente conectado descrito en la red Panorama-CNN encargado de realizar la regresión. En la Figura 4.13 podemos ver un esquema de esta arquitectura de red.

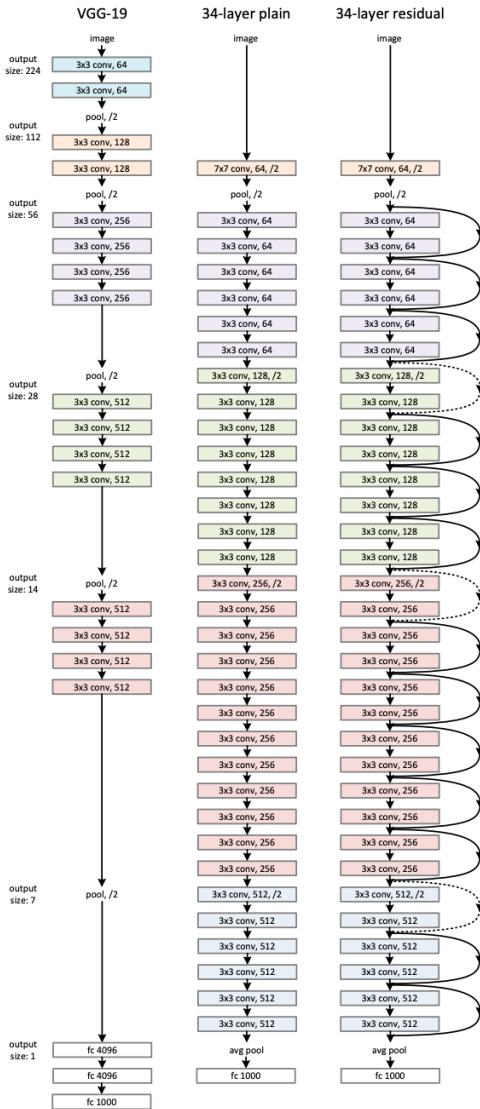


Figura 4.14: Ejemplo de la arquitectura ResNet34 para Imagenet. **Izquierda:** VGG-19 [76]. **Centro:** Red plana de 34 capas. **Derecha:** ResNet34, red residual de 34 capas. [75]

4.3. Deep imbalanced regression

El problema de la estimación de la edad a partir de modelos 3D de las síntesis púberas es un problema claramente desbalanceado (ver Figura 6.2). Esta situación no es exclusiva de nuestro problema, si no que es común e inherente a muchos problemas de ML y DL que trabajan con datos obtenidos directamente de la vida real, especialmente en contextos de diagnóstico médico. Generalmente en problemas relacionados con este campo, existen muchos ejemplos negativos y pocos ejemplos positivos (pensemos en un problema de detección de cáncer mediante radiografías, por ejemplo).

Los problemas de regresión mediante DL en el contexto de datos no balanceados se conocen como problemas de **deep imbalanced regression** (DIR). Estos problemas pueden definirse como el aprendizaje de objetivos continuos a partir de datos que provienen del mundo real, tratando con la potencial falta de datos para ciertos valores objetivos, y generalizando a un conjunto de test que está balanceado sobre el rango completo de valores objetivo [78]. Esta definición es análoga al problema del desbalanceo de clases [79] pero centrado en el espectro continuo.

En [78], Yuzhe Yang et al. proponen varios métodos para reducir el problema de desbalanceo de los datos: label distribution smoothing (LDS) y feature distribution smoothing (FDS). Para el desarrollo de este proyecto se ha escogido LDS como método de aplicación para reducir el desbalanceo presente en los datos.

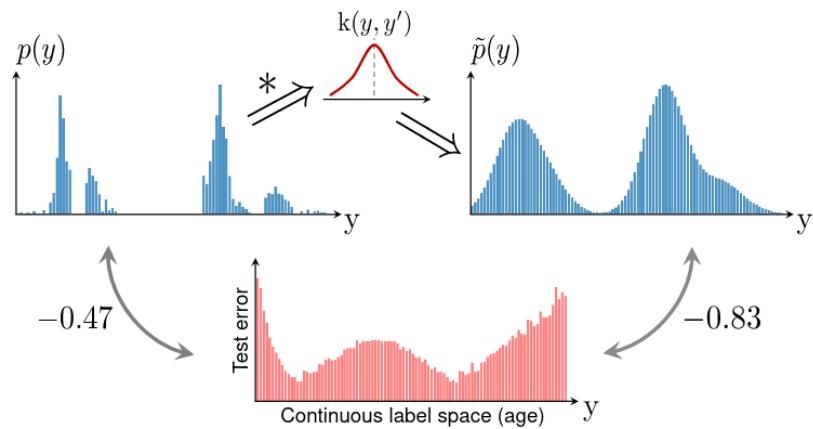


Figura 4.15: Esquema de aplicación de LDS [78].

La técnica LDS consiste en convolucionar un kernel simétrico con la densidad de etiquetas empírica para estimar la distribución de densidad de etiquetas efectiva que tiene en cuenta la continuidad de las etiquetas. (ver Figura 4.15). Una vez obtenida esta distribución de densidad efectiva de etiquetas, las muestras del conjunto de datos son ponderadas en base al inverso de su valor en dicha distribución. En la Sección 6.1.2 se muestra cómo se aplica este método en el conjunto de datos.

Capítulo 5

Planificación e implementación

5.1. Planificación del proyecto

Para el desarrollo de este proyecto, se planteó una temporización como la que puede verse en la Tabla 5.1. El trabajo de fin de grado esta dotado con una carga lectiva de 12 créditos ECTS, por lo que aplicando una conversión de 1 ECTS = 25 horas de trabajo, el desarrollo de un TFG esta planteado para tener una duración de 300 horas. Se cuenta con un total de 20 semanas, la duración aproximada de un cuatrimestre. Por tanto quedan unas 15 horas por semana, lo que hace un total de 3 horas al día teniendo en cuenta solo días laborables.

Fase	Duración semanas - horas	Febrero	Marzo	Abril	Mayo	Junio
Revisión bibliográfica	3 - 45	14 21 28	07 14 21 28	04 11 18 25	02 09 16 23 30	06 13 20 27
Implementación de la representación 3D	5 - 75					
Diseño de modelos	2 - 30					
Entrenamiento de modelos	10 - 150					

Tabla 5.1: Planificación temporal inicial del proyecto.

Finalmente la planificación inicial sufrió una serie de retrasos, alargando el tiempo dedicado a la revisión bibliográfica y retrasando las siguientes tareas, aunque el proceso de entrenamiento de los modelos conllevo menos tiempo del esperado. La planificación final del proyecto puede verse en la Tabla 5.2.

Fase	Duración semanas - horas	Febrero			Marzo			Abril			Mayo			Junio							
		14	21	28	07	14	21	28	04	11	18	25	02	09	16	23	30	06	13	20	27
Revisión bibliográfica	6 - 90																				
Implementación de la representación 3D	5 - 75																				
Diseño de modelos	2 - 30																				
Entrenamiento de modelos	7 - 105																				

Tabla 5.2: Planificación temporal final del proyecto.

Respecto a los costes del proyecto, en la Tabla 5.3 puede consultarse la estimación de costes realizada. Para esta estimación se ha supuesto un coste por hora para un investigador senior o responsable de I+D en una empresa de base tecnológica de 35 euros/hora. Además se ha añadido una categoría de coste de los materiales donde se incluyen los gastos derivados de acceder a artículos de pago, desgaste de componentes hardware, adquisición de dispositivos de almacenamiento masivos, etc.

Fecha de inicio	14/02/2022
Fecha de fin	03/07/2022
Duración en días (laborables - totales)	100 - 140
Coste del desarrollo	10.500,00 €
Coste de los materiales	3.000,00 €
Coste total	13.500,00 €

Tabla 5.3: Costes del proyecto.

5.2. Detalles técnicos de la implementación

5.2.1. Implementación de la representación PANORAMA

La implementación del preprocesador para obtener la representación PANORAMA de un modelo 3D se ha desarrollado en C++. Inicialmente el desarrollo comenzó en Python pero la primera versión del programa tomaba demasiado tiempo para generar una sola proyección panorámica. Dado que PANORAMA se basa en computar varias proyecciones por cada modelo 3D, el tiempo necesario para generar una proyección se convirtió en un factor clave. El desarrollo pasó a C++ y se consiguió una mejora de velocidad sustancial.

Toda la estructura del código se basa en la clase **Mesh3D**. En la Figura 5.4 puede consultarse el diagrama de esta clase. Las principales funciones desarrolladas son las siguientes:

- **calculate_panorama()**: Es la función principal. Computa una mapa de características (SDM o NDM+GNDM) para un eje de coordenadas

dados como parámetros. Para mejorar el rendimiento, dado que es necesario realizar raytracing para poder calcular las colisiones con los rayos, se ha definido la función `filter_faces()` que realiza un filtrado de las caras del modelo 3D según la orientación y la altura del rayo para ahorrar costes de computación en cálculo de rayos. Para calcular la colisión entre un rayo y una cara del modelo se utiliza el algoritmo de intersección Möller–Trumbore [80] que se implementa en la función `RayIntersectsTriangle()`. En la Figura 5.1 puede verse el grafo de llamadas a esta función.

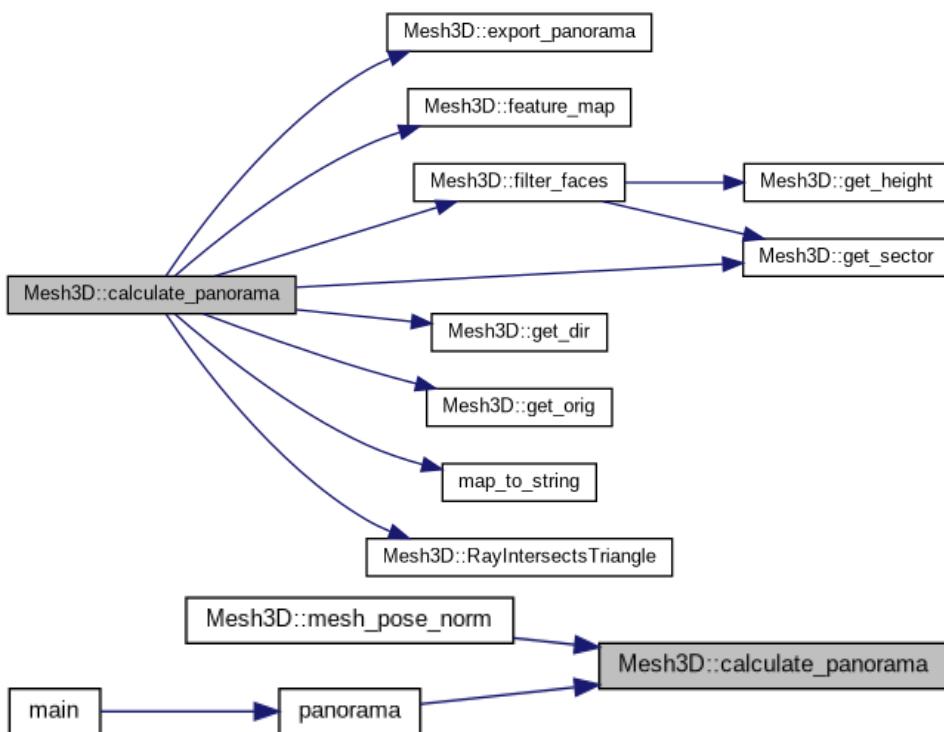


Figura 5.1: Grafo de llamadas de `calculate_panorama()`.

- `combine_panorama()`: Genera la primera de las propuestas de representación PANORAMA utilizada en este proyecto (Figura 4.8) y la almacena en una ruta de salida proporcionada como parámetro. En la Figura 5.2 puede verse el grafo de llamadas a esta función.

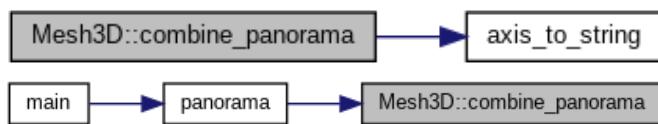


Figura 5.2: Grafo de llamadas de `combine_panorama()`.

- `concat_panorama()`: Genera la segunda de las propuestas de representación PANORAMA utilizada en este proyecto (Figura 4.9) y la almacena en una ruta de salida proporcionada como parámetro. En la Figura 5.3 puede verse el grafo de llamadas a esta función.

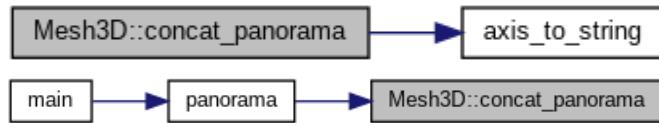


Figura 5.3: Grafo de llamadas de `concat_panorama()`.

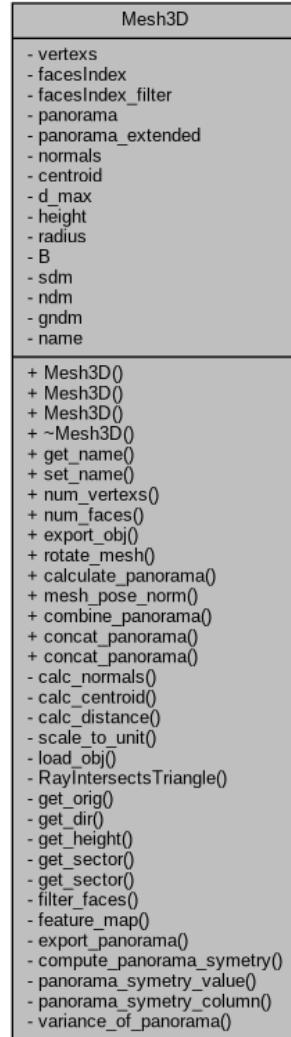


Figura 5.4: Diagrama de la clase Mesh3D.

Toda la documentación sobre la implementación de PANORAMA está disponible en: https://alexmnz1ms.github.io/panorama_extended/. Gracias al software de documentación de Doxygen y a la funcionalidad de Github Pages, la documentación es totalmente accesible de forma pública. Además el código también se encuentra públicamente accesible en: https://github.com/Alexmnz1ms/panorama_extended.

5.2.2. Entrenamiento de los modelos

Para el entrenamiento de los modelos se ha utilizado Python como lenguaje de programación por las potentes bibliotecas de ML y DL disponibles. Los modelos se han definido construido y entrenado haciendo uso de las bibliotecas de TensorFlow(v.2.9.0 gpu), una biblioteca desarrollada por Google para aplicar métodos de ML en conjunto con Keras(v.2.9.0), una biblioteca específica para redes neuronales capaz de operar sobre TensorFlow. Se ha utilizado la versión de Tensorflow gpu, que permite la aplicación de optimización GPU para acelerar el entrenamiento de los modelos junto con las bibliotecas de CUDA Toolkit (v.11.3.1) y CuDNN (v.8.2.1) para gestionar los núcleos CUDA de las gráficas NVIDIA. Como bibliotecas auxiliares se ha utilizado Numpy(v.1.22.3) para el uso de vectores, Pandas (v.1.4.2) para poder trabajar con los datos utilizando DataFrames, Scikit Learn (v.1.0.2) para funciones auxiliares de ML, OpenCV (v.4.5.5.64) para trabajar con las imágenes de entrada de las CNN y finalmente Scipy (v.1.7.3) para la aplicación del operador matemáticos de convolución. Todo esto se ha gestionado utilizando Anaconda (v.2021-05.1) como gestor de entornos.

Todo el código utilizado para entrenar los modelos esta disponible en: https://github.com/Alexmnz1ms/Age_estimation_from_3D_models.

Todos los modelos han sido entrenados en un entorno Arch Linux x86_64 Linux 5.17.9-arch1-1, junto con el siguiente hardware:

- **CPU:** AMD Ryzen 5 3600 6-Core Processor 12x 3.6GHz
- **RAM:** 16GB DDR4
- **GPU:** NVIDIA GeForce RTX 3060 12GB GDDR6

Capítulo 6

Experimentación

6.1. Conjunto de datos

El problema de la estimación de la edad a partir de modelos 3D de sínfisis púbicas abordado en este TFG y los resultados obtenidos en él, dependerán en gran medida del conjunto de datos disponible. Para llevar a cabo esta tarea, utilizaremos el conjunto de modelos 3D de las sínfisis púbicas escaneados manualmente por personal del laboratorio de Antropología Física del Departamento de Medicina Legal, Toxicología y Antropología Física de la Universidad de Granada.

Este conjunto de datos cuenta con 565 parejas de sínfisis púbicas humanas comprendidas en un rango de entre 15 y 82 años. Cada pareja está formada por dos modelos 3D, uno para el lado derecho de la sínfisis y otro para el lado izquierdo (salvo excepciones de muestras que solo cuentan con uno de sus lados). En la práctica contamos con un total de 1104 modelos 3D. Podemos ver en la Figura 6.1 como el número de muestras del lado izquierdo y del lado derecho están perfectamente balanceados.

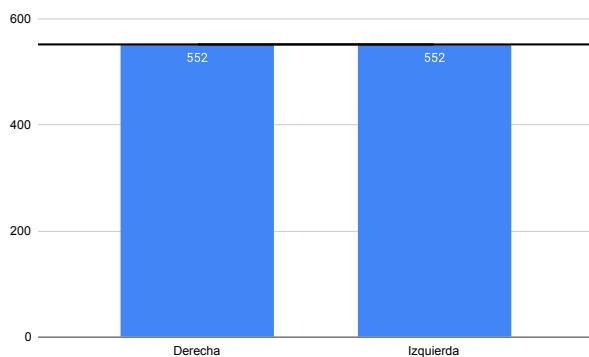


Figura 6.1: Número de muestras Izquierda y Derecha.

Uno de los aspectos más importantes a analizar es la distribución de la edad de las muestras, que se muestra en la Figura 6.2. Como podemos ver, los datos están fuertemente desbalanceados. Las muestras de entre menos de 18 años y más de 65 no superan los 10 ejemplos, mientras que las muestras contenidas entre 18 y 65 presentan números muy dispares. Este desbalanceo puede notarse mucho más si observamos la distribución de las muestras en el rango de fases de edad propuestas por Todd (ver Figura 6.3).

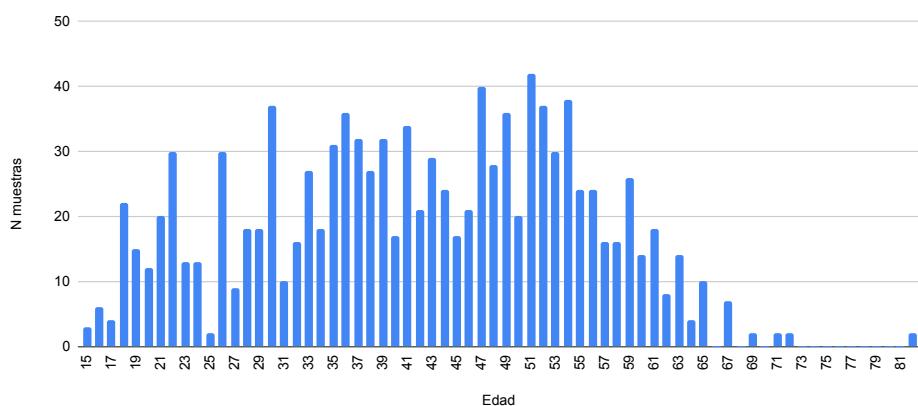


Figura 6.2: Histograma de la edad de las muestras de síntesis públicas.

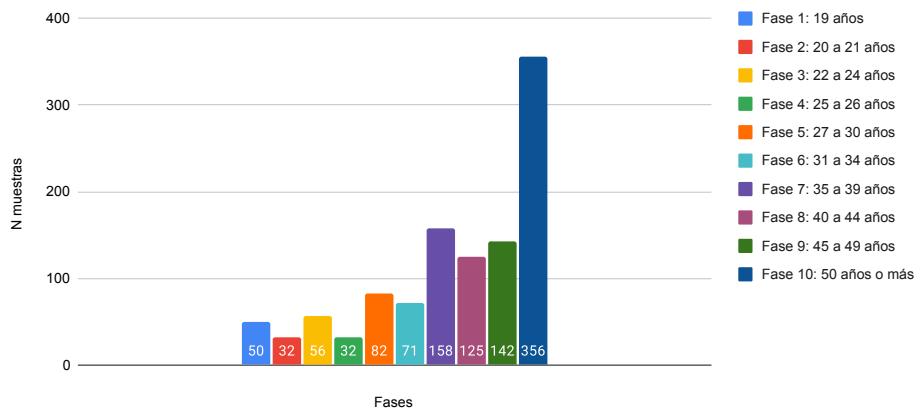


Figura 6.3: Histograma de la edad de las muestras de síntesis públicas por rangos de Todd.

No hay un número equitativo de muestras para cada edad, siendo que los huesos correspondientes a niños o ancianos son los que menor porcentaje del total representan. Dentro del grupo de los adultos (entre 19 y 64 años), encontramos que apenas hay muestras de individuos de 25 años, mientras

que, el mayor número de muestras pertenecen a individuos de 52 años.

Además el total de muestras, 1104 modelos 3D, puede parecer grande pero queda bastante mermado si lo comparamos con otros conjuntos de datos disponibles para tareas de reconocimiento de imágenes. Estos conjuntos cuentan con decenas de miles de ejemplos de entrenamiento para que un modelo pueda aprender a realizar una tarea de clasificación o de regresión (p.ej. estimación de la edad a partir de fotografías humanas).

Dado que estamos enfocando este problema desde el paradigma del DL, los datos empleados determinarán directamente el resultado final de estimación que se obtendrá. Es necesario emplear estrategias que permitan paliar la falta de datos y el desbalanceo presente en los mismos, de forma que el modelo final entrenado sea capaz de clasificar correctamente muestras de cualquier edad, aunque el conjunto de datos de entrenamiento no tuviera un elevado número de muestras correspondientes a dicha edad. Para conseguir esto se han seguido dos estrategias: aplicación de técnicas de aumento de datos y aplicación de LDS para balancear las etiquetas.

6.1.1. Aumento de datos

La primera de las estrategias adoptadas es la aplicación de “data augmentation” o aumento de datos. Esta es una técnica muy común en problemas de DL y consiste en que para cada muestra del conjunto de datos se generan nuevas muestras aplicando pequeñas transformaciones como rotaciones, traslaciones, cambio de brillo, etc.

Para este problema se ha aplicado un enfoque distinto. Generalmente existen dos maneras de aplicar aumento de datos. Si los datos con los que se trabaja se tratan de imágenes, se aplican transformaciones a los datos de entrada para generar nuevas muestras. Si, por otro lado, trabajamos directamente con vectores de características, existen algoritmos como SMOTE [81] que pueden generar nuevos ejemplos sintéticos a partir de datos reales para aumentar los datos de entrenamiento. La estrategia de utilizar SMOTE no es aplicable en nuestro caso, porque dependeríamos de un extracto de características y al trabajar con datos reales (huesos de personas reales) podríamos estar introduciendo un error derivado de crear ejemplos no realistas. Por otro lado la estrategia de generar nuevas muestras modificando las imágenes sería una alternativa viable, pero no estaríamos introduciendo nueva información de calidad al modelo, simplemente estaríamos realimentando la red con los mismos modelos.

La estrategia que hemos adoptado se basa en realizar el aumento de datos directamente sobre los modelos 3D. Para cada modelo 3D, que tiene una pose por defecto, se generan tres proyecciones PANORAMA según lo especificado en el Capítulo 4. Cada modelo se rotará un ángulo aleatorio en el eje Y de forma que la cara principal del hueso permanezca perfectamente visible en las representaciones. Estas rotaciones aleatorias se realizan un total de tres veces. En la Tabla 6.1 podemos ver un ejemplo de la aplicación de estas rotaciones.

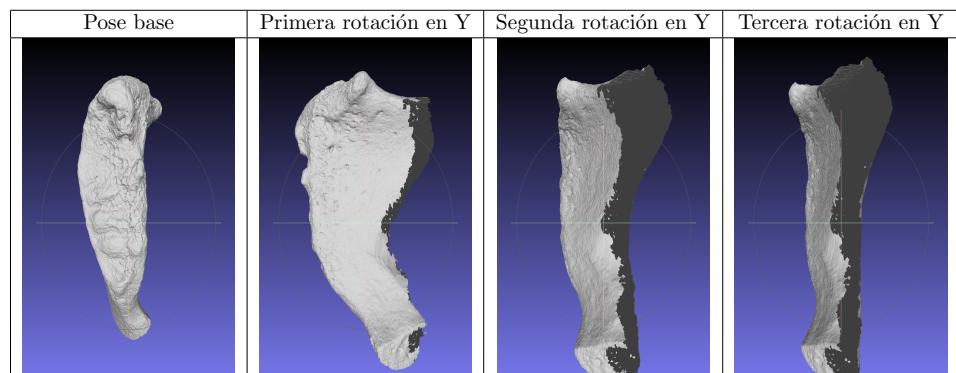


Tabla 6.1: Rotaciones en el eje Y de una muestra del conjunto de datos.

Para cada una de estas rotaciones y en la pose original, se rotará el modelo un ángulo aleatorio en sentido positivo y negativo en los ejes X (inclinación) y Z (cabeceo). Aplicando estas rotaciones se obtendrán 5 proyecciones por eje contando las cuatro rotaciones aleatorias en X y Z y la pose sin aplicar rotación. En la Tabla 6.2 podemos apreciar un ejemplo de estas rotaciones. Como estas rotaciones se realizan cuatro veces (pose base más tres rotaciones en Y) se generan un total de 60 imágenes por cada modelo entre los tres ejes (20 para cada eje). En la Figura 6.4 puede verse un ejemplo de esta estrategia de aumento de datos aplicado a una muestra del conjunto de datos.

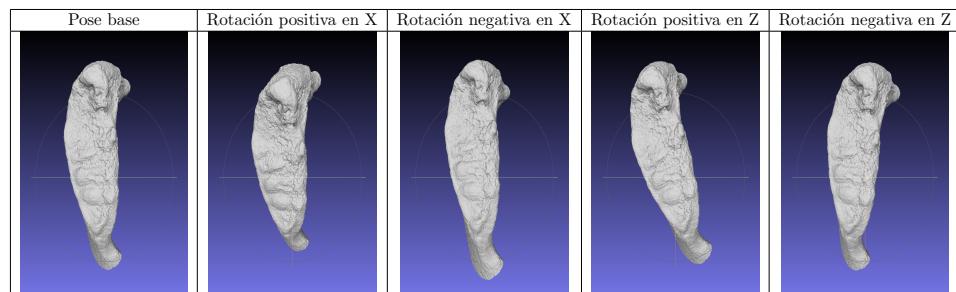


Tabla 6.2: Rotaciones en los ejes X y Z de una muestra del conjunto de datos.

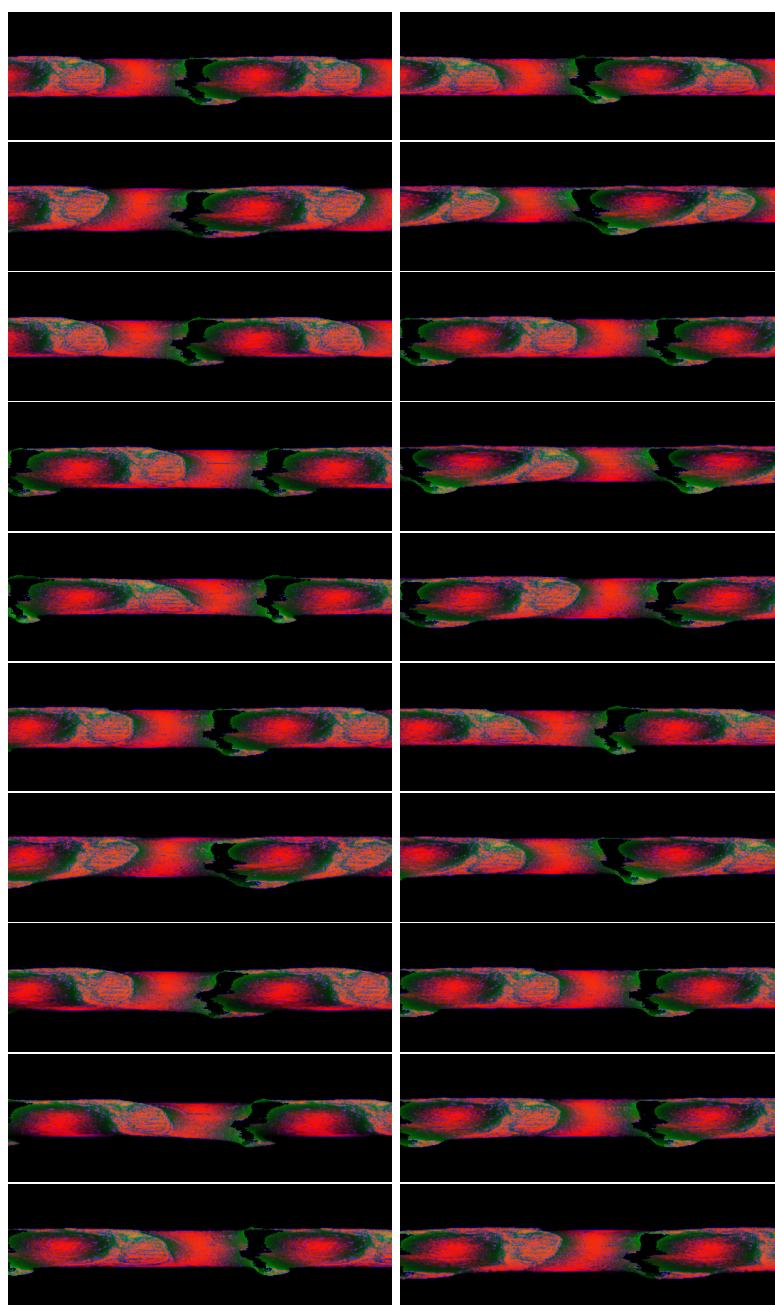


Figura 6.4: Proyecciones en el eje Y de una muestra del conjunto de datos después de aplicar la técnica de aumento de datos.

6.1.2. Balanceo de datos

La segunda estrategia que seguiremos será aplicar LDS al conjunto de datos. Como ya comentamos en la sección 4.3, LDS es una técnica aplicable en problemas de regresión donde el conjunto de datos está desbalanceado.

Para aplicar esta técnica, una vez obtenida la distribución de densidad de etiquetas (Figura 6.2) esta se convoluciona con un Kernel Gaussiano de tamaño 5 y $\sigma = 2$. Esto nos genera una nueva distribución de densidad de etiquetas en base a la continuidad de las mismas. El enfoque de este método se basa en que dos muestras que se encuentren en clases cercanas (muestras de 24, 25 y 26 años) tendrán características similares, por lo que la importancia de las mismas puede repartirse. Entendemos por importante una muestra que no es representativa, es decir, que es única o poco frecuente en el conjunto de datos. La distribución que obtenemos tras aplicar esta técnica puede verse en la Figura 6.5.

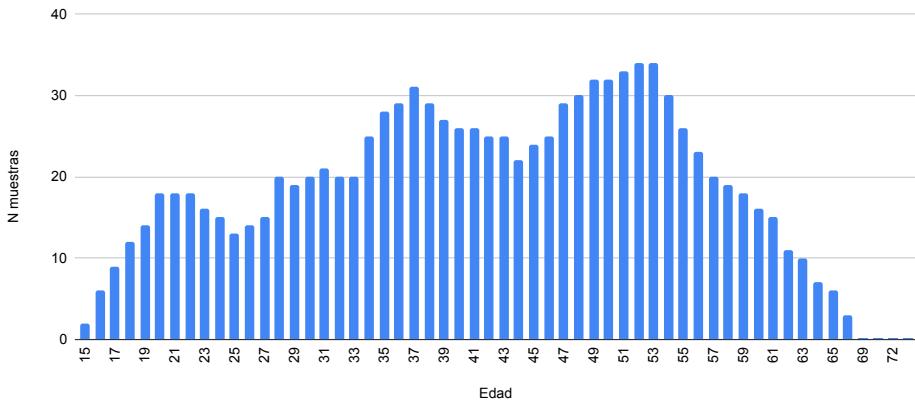


Figura 6.5: Histograma del conjunto de datos una vez aplicado LDS.

Vemos que la distribución de densidad se ha suavizado considerablemente. Se puede apreciar como en el caso de las muestras de 24, 25 y 26 años ya no existe un hueco en la distribución. Esto se entiende como que, dado que hay muchas muestras de 24 y 26 años, la falta de muestras de 25 años no es un problema grave. Sin embargo, vemos como ha desaparecido en la gráfica el valor para muestras de 82 años. Esto implica que las muestras de 82 años tienen una gran importancia porque su ausencia no puede compensarse de ninguna manera.

En base a esta distribución de densidad, se calcula el peso que tiene cada muestra en el conjunto de datos como la inversa de la distribución suavizada. Una muestra de 82 años tendrá el máximo peso (1), mientras que otra de 25 no recibirá un peso elevado a pesar de su escasez en el conjunto de datos.

Estos pesos se utilizarán en el proceso de entrenamiento para poder guiar al modelo de forma que dé mayor importancia a las muestras más escasas y que no puedan ser compensadas.

6.2. Métricas y protocolo de validación experimental

6.2.1. Métricas de calidad consideradas

Como métricas para medir la efectividad del entrenamiento del modelo se proponen las siguientes:

- *Mean Absolute Error (MAE)*: Desviación media en valor absoluto entre los valores predecidos y los reales.

$$MAE = \frac{1}{N} \sum_{i=1}^N |(\hat{Y}_i - Y_i)|$$

- *Root Mean Square Error (RMSE)*: Raíz cuadrada de la desviación media de las diferencias al cuadrado.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{Y}_i - Y_i)^2}$$

Se proponen estas métricas dado que son las métricas de referencia presentes en el estado del arte. De esta forma se podrán comparar los resultados obtenidos con los del estado del arte de forma directa.

6.2.2. Hiperparámetros utilizados

Para el entrenamiento de los modelos se han definido los siguientes hiperparámetros:

- Épocas: 100. Se ha establecido el límite en 100 épocas para que el tiempo de entrenamiento máximo no fuera demasiado elevado y el modelo pudiera converger a buenos resultados
- Tamaño de Batch: 32. Se ha establecido este tamaño de batch dado que un tamaño mayor provocaba el llenado completo de la memoria gráfica disponible. Se utiliza una potencia de dos para aprovechar las optimizaciones de la gráfica utilizada para el entrenamiento.
- Dropout: 10 %. Probabilidad de desactivar neuronas durante el entrenamiento.
- Función de perdida: MAE. Se ha establecido MAE como función de pérdida tanto para los conjuntos de train como validación.

- Early stopping: 20 épocas. El “early stopping” es una técnica que consiste en detener el entrenamiento si el modelo no ha obtenido una mejora en una métrica específica en un número concreto de épocas. Se ha establecido un 20 % de las épocas máximas como límite. La métrica para determinar la detención del entrenamiento ha sido el MAE obtenido en el conjunto de validación.
- Optimizador: Adam. Este es un método de gradiente descendente estocástico que se basa en la estimación adaptativa de momentos de primer y segundo orden. Este método es eficiente desde el punto de vista computacional, tiene pocos requisitos de memoria, es invariable al reescalado diagonal de los gradientes y se adapta bien a los problemas que son grandes en términos de datos o parámetros [82].
- Se ha establecido además un sistema de checkpoints. Este sistema guarda siempre el mejor modelo en términos del MAE obtenido en el conjunto de validación, de forma que el modelo resultado obtenido después del entrenamiento es siempre el mejor.

6.2.3. Protocolo de entrenamiento

Para el entrenamiento de todos los modelos (Resnet-CNN y Panorama-CNN en sus dos variantes) se ha realizado una normalización y estandarización previa de las imágenes de entrada. Esto es que se han normalizado las imágenes de entrada para que queden en el intervalo [0, 1] y se han estandarizado para que todo el conjunto tenga media 0 y derivación estándar 1 (en función de los datos de entrenamiento). Además las imágenes se han escalado al 20 % de su tamaño. Las imágenes RGB se han escalado para obtener unas dimensiones de 36x108 y los mapas de características se han escalado a 108x108.

El conjunto de datos se ha separado en un conjunto de train y test en una proporción de 90 % y 10 % respectivamente. Además el conjunto de train se ha dividido en dos conjuntos, uno de entrenamiento (train) y otro de validación en una proporción de 75 % y 25 % (ver Figura 6.6). Estos conjuntos se han obtenido separando los modelos en parejas de forma que para una muestra, tanto su lado izquierdo como derecho permanecen juntos en el mismo conjunto. Esto es así para no introducir ningún tipo de información en test proveniente del conjunto de train y porque en el caso de aplicación a nuevos datos reales, las síntesis púberes siempre van en parejas, por lo que no tendría sentido separarlas. Podemos ver la distribución de densidad resultante después de realizar la división de los datos en la Figura 6.7.

Para obtener resultados fiables se han realizado 5 entrenamientos distintos para cada modelo. Para la red Resnet-CNN el proceso de entrenamiento

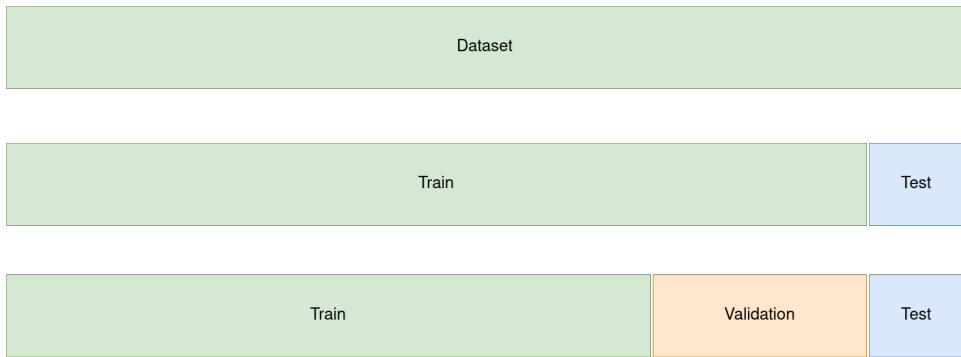


Figura 6.6: Esquema de la división del conjunto de datos (dataset) en los conjuntos de entrenamiento (train), validación (validation) y prueba (test)

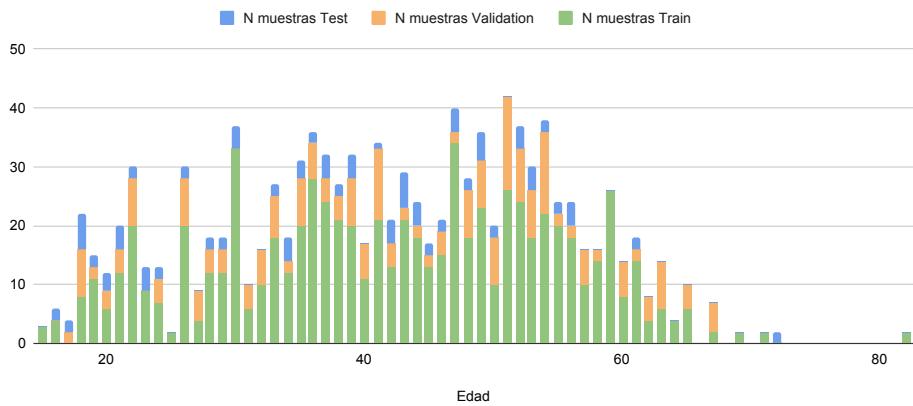


Figura 6.7: Histograma de la edad de las muestras de síntesis púbicas divididas en conjuntos de train, validation y test.

se ha dividido en dos. Una etapa de transferencia de conocimiento y otra de ajuste fino. En la primera fase, se bloquean todas las capas de la red ResNet, por lo que solo se actualizan los pesos de las capas del bloque de regresión. Esto se hace para que el modelo no destruya el conocimiento previo de la red y se aprenda a estimar la edad a partir de las características obtenidas por ResNet entrenada con Imagenet. Una vez completado este proceso, se realiza un “fine tuning” o ajuste fino. Este proceso consiste en desbloquear todas las capas del modelo y entrenar el mismo de forma completa, reduciendo la tasa de aprendizaje para ajustar la extracción de características al nuevo conjunto de datos. De esta manera, el modelo comienza el entrenamiento desde un punto mucho más favorable y converge a mejores resultados, ya que parte de una red que ya fue previamente entrenada. Esta estrategia de aplicar transferencia de conocimiento y ajuste fino en el proceso de entrenamiento de modelos de ML es muy utilizada [83].

6.2.4. Evolución del entrenamiento de los modelos

En esta sección veremos las gráficas que muestran el entrenamiento de los modelos Resnet-CNN y Panorama-CNN. Veremos primero el entrenamiento de Resnet-CNN.

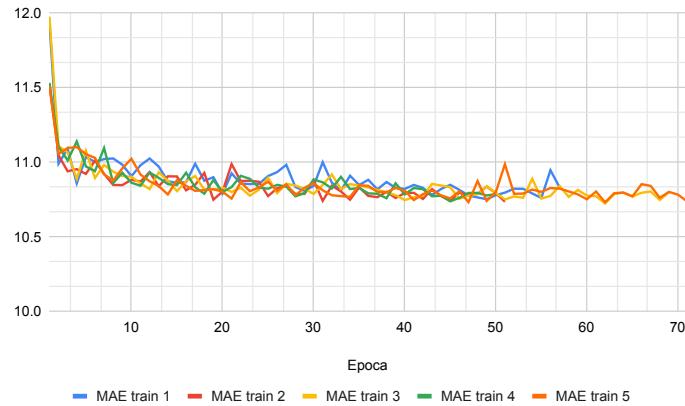


Figura 6.8: Evolución del MAE en el proceso de entrenamiento del conjunto de entrenamiento en Resnet-CNN durante la etapa de transfer learning.

Vemos en la Figura 6.8 la evolución del MAE en el conjunto de train en la etapa de transferencia de conocimiento. Podemos ver como en el conjunto de las 5 ejecuciones el modelo apenas converge a un valor de MAE inferior a 10,5.

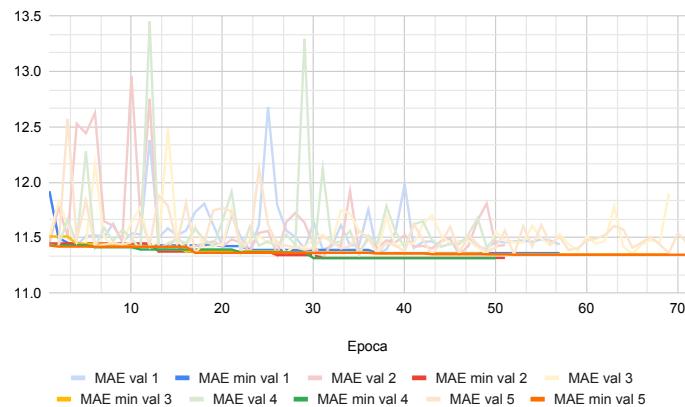


Figura 6.9: Evolución del MAE en el proceso de entrenamiento del conjunto de validación en Resnet-CNN durante la etapa de transfer learning.

En la Figura 6.9 vemos como este comportamiento se repite en el conjunto de validación. En las gráficas del conjunto de validación se muestran en color oscuro los mínimos obtenidos (el mejor modelo hasta el momento) y en color claro el valor del MAE por épocas.

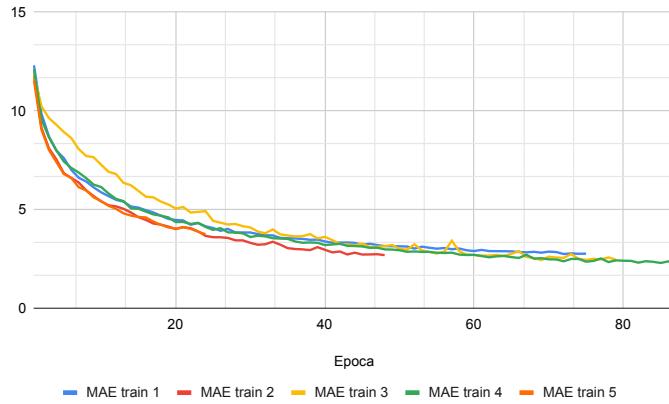


Figura 6.10: Evolución del MAE en el proceso de entrenamiento del conjunto de entrenamiento en Resnet-CNN durante la etapa de fine tuning.

Sin embargo, en la Figura 6.10 se puede apreciar como en la etapa de ajuste fino, el modelo converge a valores cercanos a 3. Esto es señal de que el modelo está aprendiendo correctamente.

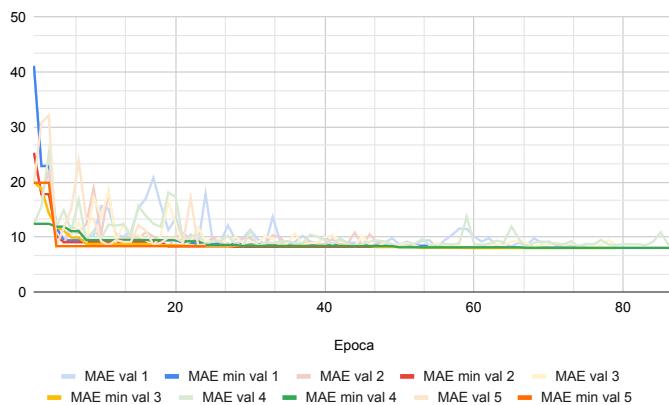


Figura 6.11: Evolución del MAE en el proceso de entrenamiento del conjunto de validación en Resnet-CNN durante la etapa de fine tuning.

En la Figura 6.11 vemos como ahora el MAE para el conjunto de validación converge de forma lenta, pero continua, alcanzando valores entre 8 y 10.

Ahora veremos el progreso del entrenamiento del modelo Panorama-CNN tanto para imágenes RGB como para mapas de características.

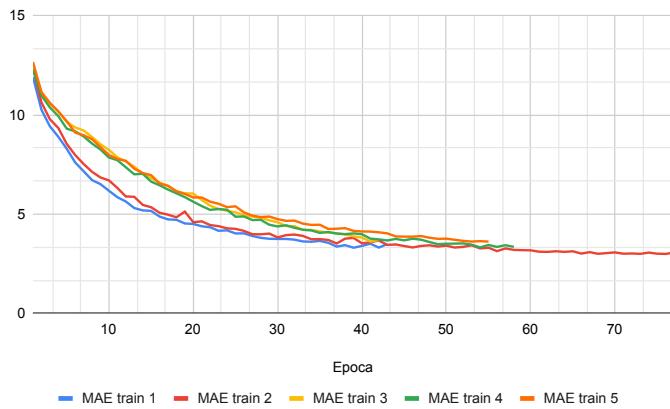


Figura 6.12: Evolución del MAE en el proceso de entrenamiento del conjunto de entrenamiento en Panorama-CNN con imágenes RGB.

En la Figura 6.12 vemos como el modelo converge correctamente a valores de MAE menores que 5 para en conjunto de train formado por imágenes RGB.

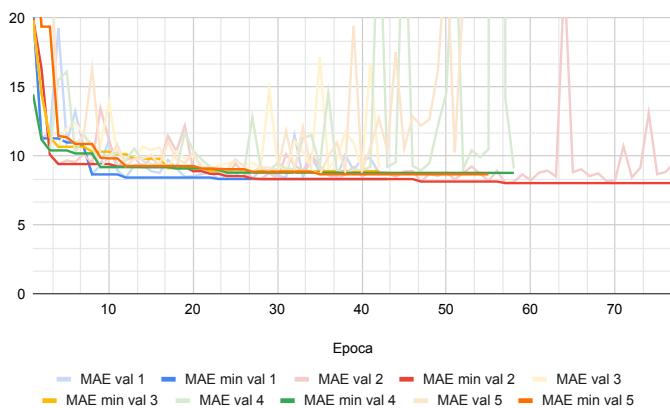


Figura 6.13: Evolución del MAE en el proceso de entrenamiento del conjunto de validación en Panorama-CNN con imágenes RGB.

En la Figura 6.13 vemos como el modelo también converge para conjunto de validación. En las gráficas de color claro pueden apreciarse picos máximos de MAE. Es por la presencia de estos picos que se aplica el sistema de checkpoints que guarda el mejor modelo encontrado hasta el momento. Si fuera de otro modo, de una época a la siguiente podríamos perder todo el proceso de convergencia conseguido.

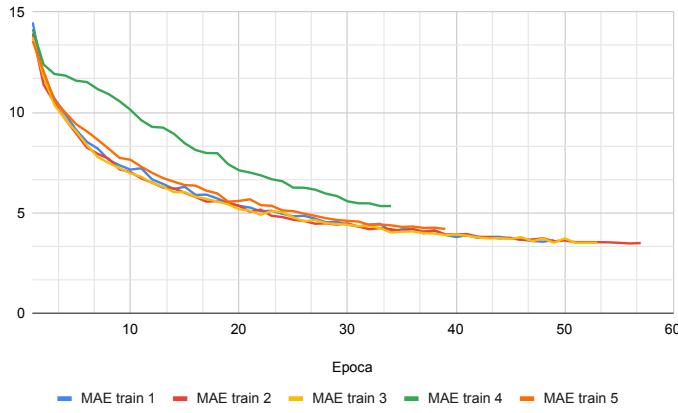


Figura 6.14: Evolución del MAE en el proceso de entrenamiento del conjunto de entrenamiento en Panorama-CNN con mapas de características.

En la Figura 6.14 vemos como el modelo Panorama-CNN para mapas de características también converge a valores de MAE por debajo de 5 en cuatro de sus ejecuciones. Dado que los modelos de CNN son modelos estocásticos, podemos atribuir este fenómeno a que esa iteración en concreto no ha podido converger a un mejor valor, ya que las otras cuatro iteraciones corroboran que el modelo es capaz de extraer información de las imágenes y aprender de ella de forma correcta.

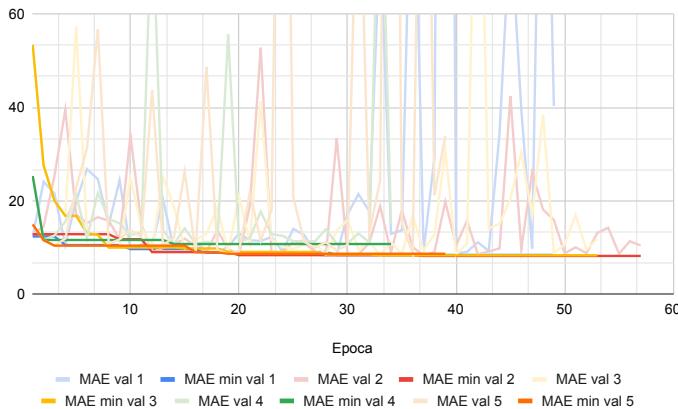


Figura 6.15: Evolución del MAE en el proceso de entrenamiento del conjunto de validación en Panorama-CNN con mapas de características.

Finalmente vemos en la Figura 6.15 el proceso de convergencia de Panorama-CNN en el conjunto de validación de mapas de características.

6.3. Resultados

Una vez vista la evolución del entrenamiento de los modelos Resnet-CNN y Panorama-CNN, en esta sección se presentan los resultados finales obtenidos para el problema de estimación de la edad a partir de modelos 3D de síntesis púbicas.

Es importante comentar que los modelos se han probado para dos versiones del conjunto de test. Generalmente, en problemas de ML donde se hace uso de alguna técnica de aumento de datos, esta nunca se aplica en el conjunto de test. Sin embargo, la técnica de aumento de datos utilizada en este proyecto no introduce nueva información en el conjunto de test. Los modelos 3D que se incluyen en este conjunto son distintos a los incluidos en train, por lo que podemos decir que son “nuevos” para los modelos entrenados. Dado que solo estamos generando vistas adicionales de los modelos 3D, es de interés comprobar la eficacia de los modelos en un conjunto de test más amplio. De esta forma se podrá constatar de mejor manera la eficacia del modelo entrenado. Aún así, se incluyen ambos enfoques para no perder de vista que nos encontramos ante un problema de ML.

6.3.1. Resultados totales

Red	Datos	N muestras	MAE avg	MAE std	MAE min	RMSE avg	RMSE std	RMSE min
Resnet-CNN	RGB	2200	8,33	0,64	7,92	11,18	0,69	10,73
Panorama-CNN	RGB	2200	9,16	0,36	8,75	11,70	0,51	11,15
Panorama-CNN	FM	2200	9,61	0,94	8,76	12,38	1,00	11,38

Tabla 6.3: Resultados RMSE y MAE totales para un conjunto de test con aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.

Red	Datos	N muestras	MAE avg	MAE std	MAE min	RMSE avg	RMSE std	RMSE min
Resnet-CNN	RGB	110	8,55	0,86	7,95	11,89	0,92	11,10
Panorama-CNN	RGB	110	8,72	0,44	8,18	11,09	0,44	10,70
Panorama-CNN	FM	110	13,03	2,20	10,57	15,32	2,09	12,94

Tabla 6.4: Resultados RMSE y MAE totales para un conjunto de test sin aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.

6.3.2. Resultados divididos en niños/adultos/ancianos

Red	Datos	Rango	N muestras	MAE avg	MAE std	MAE min	RMSE avg	RMSE std	RMSE min
Resnet-CNN	RGB	Total	2200	8,33	0,64	7,92	11,18	0,69	10,73
		15 - 18	200	9,60	3,73	5,96	12,98	3,64	9,70
		19 - 64	1960	7,70	0,39	7,39	9,96	0,48	9,51
		65 - 90	40	32,48	3,32	28,47	32,43	3,30	28,56
Panorama-CNN	RGB	Total	2200	9,16	0,36	8,75	11,70	0,51	11,15
		15 - 18	200	12,73	1,94	10,76	15,35	2,23	12,86
		19 - 64	1960	8,40	0,27	8,11	10,55	0,42	10,12
		65 - 90	40	28,89	3,01	24,39	29,51	3,09	24,88
Panorama-CNN	FM	Total	2200	9,61	0,94	8,76	12,38	1,00	11,38
		15 - 18	200	12,89	4,54	9,39	14,40	4,16	11,15
		19 - 64	1960	8,74	0,68	8,14	11,11	0,81	10,31
		65 - 90	40	35,68	2,80	32,76	35,94	2,83	32,89

Tabla 6.5: Resultados RMSE y MAE para niños/adultos/ancianos totales para un conjunto de test con aumento de datos. **Datos:** RBG, imágenes RGB; FM, mapas de características.

Red	Datos	Rango	N muestras	MAE avg	MAE std	MAE min	RMSE avg	RMSE std	RMSE min
Resnet-CNN	RGB	Total	110	8,55	0,86	7,95	11,89	0,92	11,10
		15 - 18	10	9,93	5,23	4,69	11,36	4,79	5,37
		19 - 64	98	7,66	0,48	7,22	9,74	0,68	9,26
		65 - 90	2	45,31	1,83	43,19	32,35	3,02	29,16
Panorama-CNN	RGB	Total	110	8,72	0,44	8,18	11,09	0,44	10,70
		15 - 18	10	9,26	2,19	5,65	11,48	2,22	8,03
		19 - 64	98	8,23	0,47	7,78	10,27	0,58	9,78
		65 - 90	2	29,63	2,92	24,79	30,09	2,84	25,29
Panorama-CNN	FM	Total	110	13,03	2,20	10,57	15,32	2,09	12,94
		15 - 18	10	15,93	5,02	11,26	17,05	4,44	12,87
		19 - 64	98	12,36	2,43	9,60	14,53	2,43	11,97
		65 - 90	2	31,13	4,83	25,06	31,19	4,79	25,18

Tabla 6.6: Resultados RMSE y MAE para niños/adultos/ancianos totales para un conjunto de test sin aumento de datos. **Datos:** RBG, imágenes RGB; FM, mapas de características.

6.3.3. Resultados divididos en los rangos de Todd

Red	Datos	Rango	N muestras	MAE avg	MAE std	MAE min	RMSE avg	RMSE std	RMSE min
Resnet-CNN	RGB	Total	2200	8,33	0,64	7,92	11,18	0,69	10,73
		0 - 19	240	10,94	4,15	7,60	14,76	3,87	10,41
		20 - 21	140	9,51	3,20	7,30	12,59	2,43	10,25
		22 - 24	160	10,65	1,83	7,47	12,55	1,77	9,46
		25 - 26	40	4,54	1,47	2,51	4,82	1,69	2,58
		27 - 30	160	6,93	0,98	5,85	9,02	0,96	8,02
		31 - 34	120	4,45	1,21	3,52	5,93	1,57	4,67
		35 - 39	300	7,23	0,60	6,57	8,53	0,66	7,65
		40 - 44	300	6,36	0,34	6,01	7,65	0,41	7,08
		45 - 49	300	6,50	1,07	5,33	8,25	1,48	6,73
		50 - 90	440	10,93	1,22	9,29	14,28	1,31	12,67
Panorama-CNN	RGB	Total	2200	9,16	0,36	8,75	11,70	0,51	11,15
		0 - 19	240	13,29	1,94	10,53	15,97	1,98	13,78
		20 - 21	140	14,05	3,01	10,00	15,95	2,71	12,87
		22 - 24	160	11,28	2,47	8,53	13,35	2,14	10,91
		25 - 26	40	4,21	1,12	2,53	5,27	1,23	3,23
		27 - 30	160	9,78	1,85	6,94	11,61	1,65	9,12
		31 - 34	120	6,59	0,91	5,57	8,08	1,21	6,79
		35 - 39	300	7,99	0,64	7,12	9,35	0,51	8,68
		40 - 44	300	5,12	0,26	4,77	6,42	0,26	6,08
		45 - 49	300	6,13	0,67	5,04	7,45	0,78	6,21
		50 - 90	440	11,14	1,56	9,41	14,01	1,70	12,27
Panorama-CNN	FM	Total	2200	9,61	0,94	8,76	12,38	1,00	11,38
		0 - 19	240	12,86	4,54	9,37	14,30	4,17	11,06
		20 - 21	140	16,48	1,96	14,26	17,84	1,81	15,63
		22 - 24	160	11,15	1,67	9,60	13,52	1,29	12,05
		25 - 26	40	5,93	3,07	3,57	7,32	3,32	4,84
		27 - 30	160	11,08	3,00	8,23	13,70	3,16	10,68
		31 - 34	120	11,22	1,21	9,62	13,43	1,36	11,50
		35 - 39	300	7,89	0,95	6,51	9,45	1,15	7,85
		40 - 44	300	5,08	1,07	3,91	6,52	1,28	5,28
		45 - 49	300	5,62	0,37	5,00	6,96	0,54	6,12
		50 - 90	440	11,43	0,90	10,29	15,19	1,12	13,88

Tabla 6.7: Resultados RMSE y MAE para cada fase de Todd totales para un conjunto de test con aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.

Red	Datos	Rango	N muestras	MAE avg	MAE std	MAE min	RMSE avg	RMSE std	RMSE min
Resnet-CNN	RGB	Total	110	8,55	0,86	7,95	11,89	0,92	11,10
		0 - 19	12	10,78	5,46	5,53	13,59	5,56	7,20
		20 - 21	7	9,86	3,91	6,83	11,81	3,45	9,63
		22 - 24	8	9,40	3,03	4,63	10,59	2,94	5,47
		25 - 26	2	9,80	0,92	9,26	4,43	1,41	2,23
		27 - 30	8	6,33	1,36	4,92	9,86	1,48	8,55
		31 - 34	6	4,19	1,46	2,99	4,12	2,08	1,76
		35 - 39	15	6,78	0,69	5,82	8,36	0,97	7,25
		40 - 44	15	6,39	0,63	5,95	7,54	0,77	7,08
		45 - 49	15	7,23	1,60	5,18	8,89	2,18	6,44
		50 - 90	22	12,09	1,30	10,28	13,90	1,36	12,37
Panorama-CNN	RGB	Total	110	8,72	0,44	8,18	11,09	0,44	10,70
		0 - 19	12	9,57	2,98	5,28	12,00	3,14	7,48
		20 - 21	7	14,76	4,08	8,86	16,40	3,83	11,76
		22 - 24	8	8,34	3,81	5,32	10,11	3,21	7,82
		25 - 26	2	2,26	0,97	1,20	2,61	1,34	1,26
		27 - 30	8	9,57	2,14	6,30	10,80	1,98	8,06
		31 - 34	6	6,79	2,00	4,72	7,95	2,50	5,38
		35 - 39	15	7,99	0,49	7,16	9,54	0,40	9,22
		40 - 44	15	5,03	0,53	4,28	6,38	0,87	5,46
		45 - 49	15	6,96	1,36	5,34	8,12	1,56	6,06
		50 - 90	22	11,47	1,87	8,90	14,17	1,87	11,81
Panorama-CNN	FM	Total	110	13,03	2,20	10,57	15,32	2,09	12,94
		0 - 19	12	16,23	5,05	11,50	17,36	4,51	13,23
		20 - 21	7	22,80	3,81	18,39	23,51	3,57	18,96
		22 - 24	8	16,32	3,48	11,64	17,28	2,91	13,53
		25 - 26	2	12,86	5,54	4,38	13,13	5,22	5,20
		27 - 30	8	18,96	2,96	16,36	20,00	2,83	17,10
		31 - 34	6	14,93	4,16	9,80	15,92	4,05	11,04
		35 - 39	15	12,49	4,73	5,59	13,95	4,78	6,85
		40 - 44	15	10,04	3,84	5,38	11,45	4,03	6,65
		45 - 49	15	9,02	3,24	5,96	10,37	3,39	7,09
		50 - 90	22	9,44	2,60	6,52	12,69	2,55	9,51

Tabla 6.8: Resultados RMSE y MAE para cada fase de Todd totales para un conjunto de test sin aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.

6.4. Análisis de los resultados

Una vez presentados los resultados, es el momento de realizar un análisis de los mismos. Como podemos ver en la Tabla 6.3 el modelo que alcanza mejores resultados es Resnet-CNN con un MAE de 8,33 y RMSE de 11,18 para el conjunto de test con aumento de datos. Le siguen Panorama-CNN para imágenes RGB con un MAE de 9,16 y RMSE de 11,70 y finalmente Panorama-CNN para mapas de características con un MAE de 9,61 y RMSE de 12,38. Si vemos estos resultados desglosados por rango niños/adultos/ancianos (Tabla 6.5), vemos cómo el mejor ajuste se realiza para el grupo de adultos. Esto tiene sentido ya que, al fin y al cabo, este rango posee el mayor número de muestras. Destacan los buenos resultados obtenidos en el grupo de niños: que pese a ser menos numeroso, los modelos Resnet-CNN y Panorama-CNN obtienen un error de entre 9 y 12 años. Desgraciadamente el ajuste es bastante deficiente para el grupo de los ancianos, obteniéndose un error de entre 28 y 35 años para el conjunto de test con aumento de datos. Si nos centramos en los resultados divididos según los rangos de edad propuestos por Todd (Tabla 6.7), vemos como el mejor ajuste se obtiene en las fases 4 a la 9. Esto de nuevo tiene sentido ya que estas fases cuentan con el mayor número de muestras. Sin embargo en base a esta división podemos ver como para las fases 2 y 3 pese a ser también bastante numerosas, no se consigue un ajuste tan bueno, aunque este se mantiene en valores de 9 - 10 años de MAE.

Centrándonos ahora en los resultados obtenidos para el conjunto de test sin aumento de datos, en general puede notarse como estos son ligeramente peores que los resultados para el mismo conjunto con aumento de datos. Esto de nuevo tiene sentido. En el conjunto de test con aumento de datos, el modelo está clasificando 20 veces la misma muestra, por lo que puede cometer algún error en una proyección y compensarlo con la estimación de otra. Sin embargo, en el conjunto sin aumento de datos, los modelos se evalúan una sola vez, eliminando la posibilidad de que un error se compense. Un ejemplo de esta compensación lo vemos en las Tablas 6.3 y 6.4 en los resultados de Panorama-CNN. Los resultados para el conjunto ampliado empeoran en el caso de las imágenes RGB y mejoran en el caso de los mapas de características. Vemos en la Tabla 6.4 como el mejor modelo sigue siendo Resnet-CNN con un MAE de 8,55 y RMSE de 11,89. Aunque esto es en base al MAE medio, si nos fijamos en el RMSE medio, el mejor modelo es Panorama-CNN para imágenes RGB. Dado que las métricas mostradas son una media entre las 5 ejecuciones realizadas y observando también que la derivación estándar de ambos modelos es muy baja, podemos concluir que realmente ambos modelos están a la par en cuanto a resultados para el conjunto de test sin aumentar. Para los resultados divididos por rangos (Tablas 6.6 y 6.8), observamos la misma tendencia descrita en los resultados para el conjunto de test con aumento de datos.

De estos resultados extraemos además de las dos propuestas de representación presentadas en este proyecto (imágenes RGB y mapas de características), la mejor ha sido la propuesta de imágenes RGB (utilizada por ambas propuestas).

6.5. Comparación con el estado del arte

Finalmente, veamos cómo son los resultados en comparación con el estado del arte. En la Tabla 6.9 vemos una comparativa entre los resultados obtenidos por los modelos Resnet-CNN y Panorama-CNN, y los resultados presentes en el estado del arte. Esta tabla esta estructurada diferenciando los métodos que realizan una estimación de la edad en fases o intervalos de edad (PB) y métodos que estiman un valor numérico para la edad de la muerte (N). Además se añade un nivel más de segmentación de los resultados, diferenciando entre métodos basados en componentes (CS) y métodos globales (G). Cada método de los presentes en el estado del arte cuenta con un enfoque y diseño experimental distinto, por lo que la comparativa que se muestra es aproximada.

Método	Tipo	N muestras	Rango de edad	RMSE	MAE
Slice y Algee-Hewitt [58]	N, CS	41	19 - 96	17,15	-
Stoyanova et al. [59]	N, CS	56	16 - 100	19	-
Stoyanova et al. [60]	N, CS	93	16 - 90	13,7 - 16,5	-
Koterova et al. [61]	N, CS	941	19 - 100	12,1	9,7
Gámez-Granados et al. [56]	PB, G	892	18 - 60	13,19	10,38
Gámez-Granados et al. [56]	PB, G	960	18 - 60	14,61	11,62
Bermejo et al [62] (GP)	N, CS	1152	18-82	10,82	8,56
Bermejo et al [62] (GA-P)	N, CS	1152	18-82	10,81	8,55
Resnet-CNN (Aumento)	N, G	1104	15 - 82	11,18	8,33
Resnet-CNN	N, G	1104	15 - 82	11,89	8,55
Panorama-CNN (RGB) (Aumento)	N, G	1104	15 - 82	11,7	9,16
Panorama-CNN (RGB)	N, G	1104	15 - 82	11,09	8,72
Panorama-CNN (Mapas) (Aumento)	N, G	1104	15 - 82	12,38	9,61
Panorama-CNN (Mapas)	N, G	1104	15 - 82	15,32	13,03

Tabla 6.9: Comparación entre los resultados obtenidos y los resultados del estado del arte. **Tipo:** PB, basado en fases; N, numérico; CS, basado en componentes. **MAE** y **RMSE** indican el error medido en años.

Nuestro modelo se enmarca dentro de la categoría de métodos numéricos, ya que estimamos directamente un valor numérico para la edad de la muerte a partir de las características morfológicas globales extraídas automáticamente a partir de los modelos 3D. Comparándolo con otros métodos de estimación numéricos, vemos cómo nuestro modelo iguala a los mejores resultados del estado del arte, los aportados por Bermejo et al [62]. Nuestra mejor estimación obtenida por los modelos de Resnet-CNN y Panorama-

CNN es de un MAE de 8,33 y un RMSE de 11,18 para el conjunto de test aumentado y un MAE de 8,55 y un RMSE de 11,09 para el conjunto de test sin aumento, mientras que [62] aportan en su mejor resultado un MAE de 8,55 y un RMSE de 10,81. Cabe recalcar las diferencias fundamentales de ambos modelos, pues en la propuesta de [62] plantean un modelo de apoyo al antropólogo forense que depende de la extracción subjetiva de características de un experto, mientras que nuestro modelo extrae estas características a partir de los modelos 3D de forma automática. Esto nos sugiere la posibilidad que las características aprendidas por nuestro modelo son, al menos, tan buenas como las utilizadas actualmente por los antropólogos forenses. Comparando nuestro método con métodos basados en fases del estado del arte, vemos como mejoramos los mejores resultados aportados por [56] con un RMSE de 14,61 y un MAE de 11,68. Por tanto vemos como nuestra propuesta queda enmarcada a la altura de otros métodos del estado del arte y que presenta unos resultados de estimación muy competitivos, comparables a los métodos que cuentan con la intervención de un experto antropólogo forense.

Capítulo 7

Conclusiones

A lo largo de este TFG, establecemos el problema de la estimación del PB dentro del marco de la AF como uno de los principales retos dentro de este campo y constatamos la importancia de dotar a los antropólogos forenses con sistemas de apoyo que ayuden a reducir los errores cometidos por el experto humano en la toma de decisiones en el proceso de estimación de PB, en concreto de la estimación de la edad a partir de la sínfisis púbica.

En base a los resultados obtenidos, el problema de **estimar la edad mediante la aplicación de técnicas de DL a modelos 3D de la sínfisis púbica** ha sido abordado y resuelto con éxito. Nos encontramos ante un problema complejo, que a priori presentaba grandes retos, dado el limitado número de modelos 3D disponibles para el entrenamiento y el desbalanceo del conjunto de datos. Ha sido necesario encontrar e implementar una representación de datos 3D para uso en modelos de DL y además ha sido necesario diseñar e implementar modelos capaces de hacer un buen uso de esta representación.

Todos los objetivos definidos en este TFG se han cumplido en su totalidad. Se ha realizado un estudio pormenorizado y sistematizado del estado del arte de la representación de modelos 3D y se ha elegido una representación adecuada para los modelos 3D a nuestra disposición. Se han desarrollado las herramientas necesarias para implementar esta representación y, además, estas están disponibles de forma libre a disposición de cualquiera en Github (https://github.com/Alexmnzlms/Age_estimation_from_3D_models). Se han diseñado modelos profundos convolucionales que han probado ser eficaces a la hora de resolver la tarea de estimar la edad. Se han aplicado estrategias para paliar la falta de datos y el desbalanceo de clases, y estas han probado ser estrategias eficaces. Y, para terminar, se ha comprobado que los modelos desarrollados, en combinación con la representación escogida, han sido capaces de aprender de forma automática las características representativas de los restos óseos.

Finalmente se han obtenido resultados que igualan o incluso mejoran a los presentes en el estado del arte, por lo que este trabajo se presenta como una opción muy competitiva dentro del campo de la estimación de la edad a partir de las sínfisis púbicas. Igualamos al mejor resultado del estado del arte, aportado por [62], con un MAE de 8,55. Respecto al RMSE, quedamos por encima de [62] por una diferencia menor a 0,5 años. La principal diferencia entre ambos métodos, es que mientras que [62] realiza una estimación basada en componentes e involucra la participación de un experto forense en la extracción de características, nuestra propuesta extrae las características morfológicas de las sínfisis púbicas de forma automática a partir del modelo 3D en su totalidad. Constatamos que las características extraídas por nuestro modelo a partir de los modelos 3D, a priori, parecen ser de la misma calidad que las características morfológicas de las sínfisis púbicas utilizadas en los métodos propuestos en el estado del arte. La mejor estimación obtenida por nuestro modelo, se enmarca dentro del rango de individuos adultos (entre 19 y 64 años) con una media de un MAE de 7,7 y un RMSE de 9,96 en el conjunto de test aumentado (ver Tabla 6.5) y un MAE de 7,66 y un RMSE de 9,74 en el conjunto de test no aumentado (ver Tabla 6.6). Dentro de este rango, vemos como las mejores estimaciones obtenidas para cada una de las fases propuestas por Todd (fases 4 a 9) se encuentran por debajo de 10 años de error tanto en MAE como en RMSE. Seguido de esta, la segunda mejor estimación se da en el grupo de los niños. Nuestro modelo obtiene un MAE de 9,6 y un RMSE de 12,98 en el conjunto de test aumentado y un MAE de 9,26 y un RMSE de 11,36 en el conjunto de test no aumentado. Destaca la calidad del ajuste dado la escasez de datos pertenecientes a este grupo. Para terminar, la peor estimación se da en el rango de los ancianos, dado que este es el grupo más reducido de todos. Nuestro modelo no obtiene estimaciones con un error inferior a 25 años. Esto puede estar motivado por la falta de datos y por la dificultad de estimar y extraer características morfológicas discriminatorias a partir de muestras de individuos de avanzada edad.

Nuestro modelo presenta la principal ventaja de ser automático, es decir, no requiere la presencia de un experto antropólogo para realizar la estimación. Una vez obtenida la representación 3D de las sínfisis púbicas, los métodos propuestos pueden ser aplicados de forma directa y efectiva. A su vez, dado que aplica técnicas y algoritmos de DL clásicos, es perfectamente escalable y aplicable a nuevos conjuntos de datos. Sin embargo, esta propuesta no está exenta de desventajas, ya que no se extrae ninguna información de utilidad para el antropólogo forense más allá de la estimación de la edad de la muestra. Aunque este no era el objetivo del enfoque propuesto, varios de los métodos presentes en el estado del arte pueden servir para validar la utilidad de utilizar ciertas regiones de la sínfisis para la estimación u otras

en base a los resultados.

7.1. Trabajos futuros

Aunque este proyecto haya concluido, existen varias líneas que seguir para mejorar el desempeño del sistema en el futuro y comprender de mejor manera el proceso de aprendizaje que ha realizado.

En primer lugar, una expansión de este proyecto podría ser estudiar nuevas formas de combinar los outputs de los modelos más allá de realizar una media aritmética (combinación de las salidas de los clasificadores, weak learners, en un ensemble de clasificadores [84, 85]). El problema podría enfocarse incluso como un reto que el modelo debe resolver, en el sentido de que el propio modelo podría aprender de forma empírica cuál de las imágenes empleadas como input aporta más información y beneficia más al aprendizaje.

Otra línea de trabajo podría ser dotar a las redes convolucionales consideradas de una componente explicable [86, 87, 88], por ejemplo, estudiando las regiones de activación de los modelos. Para una red CNN cualquiera, es posible visualizar sus mapas de activación para identificar aquellas regiones de la imagen que han tenido un mayor peso a la hora de realizar una predicción determinada [89]. Además como el input de las redes profundas son proyecciones, estos mapas de activación podrían proyectarse sobre el modelo 3D de forma inversa a la proyección actual, de forma que se obtuviera un mapa 3D de las zonas del hueso que están siendo utilizadas para realizar la discriminación. Esto permitiría a los antropólogos forenses y expertos comparar estas regiones de activación con otras regiones características propuestas en la literatura (por ejemplo las regiones características propuestas por Todd) de forma que se pudiera corroborar la eficacia de estas regiones o incluso descubrir nuevas regiones características.

Bibliografía

- [1] Douglas H Ubelaker. «Forensic anthropology: methodology and diversity of applications». En: *Biological anthropology of the human skeleton* (2008), págs. 41-69.
- [2] Steven N Byers. *Introduction to forensic anthropology*. Taylor & Francis, 2016.
- [3] Gloria Vallejo y Antonio Alonso. «La identificación genética en grandes catástrofes: avances científicos y normativos en España». En: *Revista española de medicina legal* 35.1 (2009), págs. 19-27.
- [4] Diego Casallas y Juliana Padilla Piedrahita. «Antropología forense en el conflicto armado en el contexto latinoamericano: Estudio comparativo Argentina, Guatemala, Perú y Colombia». En: *Maguaré* 18 (2004), pág. 12.
- [5] Pablo Mesejo y col. «A survey on artificial intelligence techniques for biomedical image analysis in skeleton-based forensic human identification». En: *Applied sciences* 10.14 (2020), pág. 4703.
- [6] Kazuro Hanihara y Takao Suzuki. «Estimation of age from the pubic symphysis by means of multiple regression analysis». En: *American journal of physical anthropology* 48.2 (1978), págs. 233-239.
- [7] Salina Hisham y col. «Quantification of pubic symphysis metamorphosis based on the analysis of clinical MDCT scans in a contemporary Malaysian population». En: *Journal of forensic sciences* 64.6 (2019), págs. 1803-1811.
- [8] T Wingate Todd. «Age changes in the pubic bone». En: *American journal of physical anthropology* 4.1 (1921), págs. 1-70.
- [9] B Miles Gilbert y Thomas W McKern. «A method for aging the female os pubis». En: *American journal of physical anthropology* 38.1 (1973), págs. 31-38.
- [10] Sheilagh Brooks y Judy M Suchey. «Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods». En: *Human evolution* 5.3 (1990), págs. 227-238.

- [11] Douglas H Ubelaker, Quinnlan R Cordero y Nadia F Linton. «Recent research in forensic anthropology». En: *European journal of anatomy* 24.3 (2020), págs. 221-227.
- [12] Eman Ahmed y col. «A survey on deep learning advances on different 3D data representations». En: *arXiv preprint arXiv:1808.01462* (2018).
- [13] Abubakar Sulaiman Gezawa y col. «A review on deep learning approaches for 3d data representations in retrieval and classifications». En: *IEEE access* 8 (2020), págs. 57566-57593.
- [14] Christopher M. Bishop. *Pattern recognition and machine learning*. Vol. 4. 4. Springer, 2006.
- [15] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.
- [16] Yaser S Abu-Mostafa, HT Lin y M Magdon-Ismail. *Learning from data: a short course*. AMLBook, 2012.
- [17] Yann LeCun, Yoshua Bengio y Geoffrey Hinton. «Deep learning». En: *Nature* 521.7553 (2015), págs. 436-444.
- [18] Jürgen Schmidhuber. «Deep learning in neural networks: An overview». En: *Neural networks* 61 (2015), págs. 85-117.
- [19] Ian Goodfellow, Yoshua Bengio y Aaron Courville. *Deep learning*. MIT press, 2016.
- [20] Yoshua Bengio, Aaron Courville y Pascal Vincent. «Representation learning: A review and new perspectives». En: *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013), págs. 1798-1828.
- [21] IBM Cloud Education. *What is machine learning?* Jul. de 2020. URL: <https://www.ibm.com/cloud/learn/machine-learning> (visitado 21-06-2022).
- [22] Christopher M. Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995.
- [23] José David Villanueva García y col. *Redes neuronales desde cero (i) - introducción*. Oct. de 2020. URL: https://www.iartificial.net/redes-neuronales-desde-cero-i-introduccion/#Redes_neuronales (visitado 21-06-2022).
- [24] Yann LeCun y col. «Backpropagation applied to handwritten zip code recognition». En: *Neural computation* 1.4 (1989), págs. 541-551.
- [25] Yann LeCun y col. «Gradient-based learning applied to document recognition». En: *Proceedings of the IEEE* 86.11 (1998), págs. 2278-2324.
- [26] Sumit Saha. *A comprehensive guide to Convolutional Neural Networks - the eli5 way*. Dic. de 2018. URL: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53> (visitado 21-06-2022).

- [27] Muhamad Yani y col. «Application of transfer learning using convolutional neural network method for early detection of terry's nail». En: *Journal of physics: conference series*. Vol. 1201. 1. 2019, pág. 012052.
- [28] Daniel Nelson. *Neural net dropout: Dealing with overfitting*. Ene. de 2018. URL: <https://www.datascience.us/neural-net-dropout-dealing-overfitting/> (visitado 22-06-2022).
- [29] Amar Budhiraja. *Learning less to learn better - dropout in (deep) machine learning*. Mar. de 2018. URL: <https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5> (visitado 22-06-2022).
- [30] Alex Krizhevsky, Ilya Sutskever y Geoffrey E Hinton. «Imagenet classification with deep convolutional neural networks». En: *Advances in neural information processing systems* 25 (2012).
- [31] Kevin Lai y col. «A large-scale hierarchical multi-view rgb-d object dataset». En: *2011 IEEE international conference on robotics and automation*. 2011, págs. 1817-1824.
- [32] Facundo Mémoli y Guillermo Sapiro. «A theoretical and computational framework for isometry invariant recognition of point cloud data». En: *Foundations of Computational Mathematics* 5.3 (2005), págs. 313-347.
- [33] Haggai Maron y col. «Convolutional neural networks on surfaces via seamless toric covers.» En: *ACM Trans. Graph.* 36.4 (2017), págs. 71-1.
- [34] Samuel Hornus Sylvain Lefebvre y Fabrice Neyret. *GPU Gems 2*. 2005. URL: <https://developer.nvidia.com/gpugems/gpugems2/copyright> (visitado 02-07-2022).
- [35] Michael M Bronstein y col. «Geometric deep learning: Grids, groups, graphs, geodesics, and gauges». En: *arXiv preprint arXiv:2104.13478* (2021).
- [36] Alexander Bronstein y Michael Bronstein. *Dense Descriptors for Images and Surfaces*. 2007. URL: <http://vision.mas.ecp.fr/Personnel/iasonas/descriptors.html> (visitado 02-07-2022).
- [37] Hang Su y col. «Multi-view convolutional neural networks for 3d shape recognition». En: *Proceedings of the IEEE international conference on computer vision*. 2015, págs. 945-953.
- [38] Maxim Tatarchenko, Alexey Dosovitskiy y Thomas Brox. «Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs». En: *Proceedings of the IEEE international conference on computer vision*. 2017, págs. 2088-2096.

- [39] Gernot Riegler, Ali Osman Ulusoy y Andreas Geiger. «Octnet: Learning deep 3d representations at high resolutions». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, págs. 3577-3586.
- [40] Lisha Zhang y col. «Survey on 3D shape descriptors». En: *Fundaçao para a Ciencia ea Tecnologia, Lisboa, Portugal, Tech. Rep. Technical Report, DecorAR (FCT POSC/EIA/59938/2004)* 3 (2007).
- [41] Ismail Khalid Kazmi, Lihua You y Jian Jun Zhang. «A survey of 2D and 3D shape descriptors». En: *International conference computer graphics, imaging and visualization*. 2013, págs. 1-10.
- [42] Zhizhong Han y col. «Mesh convolutional restricted Boltzmann machines for unsupervised learning of features with structure preservation on 3-D meshes». En: *IEEE transactions on neural networks and learning systems* 28.10 (2016), págs. 2268-2281.
- [43] Matthias Fey y col. «Splinecnn: Fast geometric deep learning with continuous b-spline kernels». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, págs. 869-877.
- [44] Flawnson Tong. *Graph Convolutional Networks for Geometric Deep Learning*. Mayo de 2019. URL: <https://towardsdatascience.com/graph-convolutional-networks-for-geometric-deep-learning-1faf17dee008> (visitado 19-03-2022).
- [45] David K Hammond, Pierre Vandergheynst y Rémi Gribonval. «Wavelets on graphs via spectral graph theory». En: *Applied and Computational Harmonic Analysis* 30.2 (2011), págs. 129-150.
- [46] Jing Zhao y col. «Multi-view learning overview: Recent progress and new challenges». En: *Information fusion* 38 (2017), págs. 43-54.
- [47] Kristen M Hartnett. «Analysis of age-at-death estimation using data from a new, modern autopsy sample—part I: pubic bone». En: *Journal of forensic sciences* 55.5 (2010), págs. 1145-1151.
- [48] Joanne L Buckberry y Andrew T Chamberlain. «Age estimation from the auricular surface of the ilium: a revised method». En: *American journal of physical anthropology* 119.3 (2002), págs. 231-239.
- [49] M Yaşar İşcan, Susan R Loth y Ronald K Wright. «Metamorphosis at the sternal rib end: a new method to estimate age at death in white males». En: *American journal of physical anthropology* 65.2 (1984), págs. 147-156.
- [50] Eugénia Cunha y col. «The problem of aging human remains and living individuals: a review». En: *Forensic science international* 193.1-3 (2009), págs. 1-13.

- [51] Beatrix Dudzik y Natalie R Langley. «Estimating age from the pubic symphysis: A new component-based system». En: *Forensic science international* 257 (2015), págs. 98-105.
- [52] Aurore Schmitt y col. «Variability of the pattern of aging on the human skeleton: evidence from bone indicators and implications on age at death estimation». En: *Journal of forensic science* 47.6 (2002), págs. 1203-1209.
- [53] Erin H Kimmerle, Debra A Prince y Gregory E Berg. «Inter-observer variation in methodologies involving the pubic symphysis, sternal ribs, and teeth». En: *Journal of forensic sciences* 53.3 (2008), págs. 594-600.
- [54] Fabrice Dedouit y col. «Virtual autopsy and forensic identification-practical application: a report of one case». En: *Journal of forensic sciences* 52.4 (2007), págs. 960-964.
- [55] Pedro Villar y col. «A first approach to a fuzzy classification system for age estimation based on the pubic bone». En: *International conference on fuzzy systems*. 2017, págs. 1-6.
- [56] J. C. Gámez y col. «Automating the decision making process of todd's age estimation method from the pubic symphysis with explainable machine learning». En: *Sometido a Information sciences* (2021).
- [57] Juan Carlos Gámez y col. «Ordinal classification based on the sequential covering strategy». En: *International journal of approximate reasoning* 76 (2016), págs. 96-110.
- [58] Dennis E Slice y Bridget FB Algee-Hewitt. «Modeling bone surface morphology: a fully quantitative method for age-at-death estimation using the pubic symphysis». En: *Journal of forensic sciences* 60.4 (2015), págs. 835-843.
- [59] Detelina Stoyanova, Bridget FB Algee-Hewitt y Dennis E Slice. «An enhanced computational method for age-at-death estimation based on the pubic symphysis using 3 D laser scans and thin plate splines». En: *American journal of physical anthropology* 158.3 (2015), págs. 431-440.
- [60] Detelina K Stoyanova y col. «A computational framework for age-at-death estimation from the skeleton: Surface and outline analysis of 3D laser scans of the adult pubic symphysis». En: *Journal of forensic sciences* 62.6 (2017), págs. 1434-1444.
- [61] Anežka Kotěrová y col. «Age estimation of adult human remains from hip bones using advanced methods». En: *Forensic science international* 287 (2018), págs. 163-175.
- [62] Enrique Bermejo y col. «Design of New Component-Scoring Age-at-death Estimation Methods from the Pubic Symphysis Using Genetic Programming-based Symbolic Regression». En: *Artículo en preparación* (2022).

- [63] Anuj Karpatne y col. «Theory-guided data science: A new paradigm for scientific discovery from data». En: *IEEE Transactions on knowledge and data engineering* 29.10 (2017), págs. 2318-2331.
- [64] Andrés Castillo y col. «preliminary insight into a new method for age-at-death estimation from the pubic symphysis». En: *International journal of legal medicine* 135.3 (2021), págs. 929-937.
- [65] Zhuotun Zhu y col. «Deep learning representation using autoencoder for 3D shape retrieval». En: *Neurocomputing* 204 (2016), págs. 41-50.
- [66] Baoguang Shi y col. «Deeppano: Deep panoramic representation for 3-d shape recognition». En: *IEEE Signal processing letters* 22.12 (2015), págs. 2339-2343.
- [67] Ayan Sinha, Jing Bai y Karthik Ramani. «Deep learning 3D shape surfaces using geometry images». En: *European conference on computer vision*. 2016, págs. 223-240.
- [68] Zhangjie Cao, Qixing Huang y Ramani Karthik. «3d object classification via spherical projections». En: *2017 international conference on 3D Vision (3DV)*. 2017, págs. 566-574.
- [69] Konstantinos Sfikas, Theoharis Theoharis y Ioannis Pratikakis. «Exploiting the PANORAMA Representation for Convolutional Neural Network Classification and Retrieval.» En: *Eurographics workshop on 3D object retrieval*. 2017.
- [70] Konstantinos Sfikas, Theoharis Theoharis y Ioannis Pratikakis. «Pose normalization of 3D models via reflective symmetry on panoramic views». En: *The visual computer* 30.11 (2014), págs. 1261-1274.
- [71] Konstantinos Sfikas, Ioannis Pratikakis y Theoharis Theoharis. «Ensemble of PANORAMA-based convolutional neural networks for 3D model classification and retrieval». En: *Computers graphics* 71 (2018), págs. 208-218.
- [72] Zhirong Wu y col. «3d shapenets: A deep representation for volumetric shapes». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, págs. 1912-1920.
- [73] Panagiotis Papadakis y col. «Panorama: A 3d shape descriptor based on panoramic views for unsupervised 3d object retrieval». En: *International journal of computer vision* 89.2 (2010), págs. 177-192.
- [74] Chuanqi Tan y col. «A survey on deep transfer learning». En: *International conference on artificial neural networks*. 2018, págs. 270-279.
- [75] Kaiming He y col. «Deep residual learning for image recognition». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, págs. 770-778.

- [76] Karen Simonyan y Andrew Zisserman. «Very Deep Convolutional Networks for Large-Scale Image Recognition». En: *3rd International Conference on Learning Representations*. 2015.
- [77] Olga Russakovsky y col. «Imagenet large scale visual recognition challenge». En: *International journal of computer vision* 115.3 (2015), págs. 211-252.
- [78] Yuzhe Yang y col. «Delving into Deep Imbalanced Regression». En: *International conference on machine learning*. 2021.
- [79] Ziwei Liu y col. «Large-scale long-tailed recognition in an open world». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2019, págs. 2537-2546.
- [80] Tomas Möller y Ben Trumbore. «Fast, Minimum Storage Ray-Triangle Intersection». En: *Journal of graphics tools* 2.1 (1997), págs. 21-28.
- [81] Nitesh V Chawla y col. «SMOTE: synthetic minority over-sampling technique». En: *Journal of artificial intelligence research* 16 (2002), págs. 321-357.
- [82] Diederik P Kingma y Jimmy Ba. «Adam: A method for stochastic optimization». En: *arXiv preprint arXiv:1412.6980* (2014).
- [83] *Keras Documentation: Transfer Learning Fine-tuning*. URL: https://keras.io/guides/transfer_learning/ (visitado 26-06-2022).
- [84] Yue Cao y col. «Ensemble deep learning in bioinformatics». En: *Nature machine intelligence* 2.9 (2020), págs. 500-508.
- [85] Mudasir A Ganaie, Minghui Hu y col. «Ensemble deep learning: A review». En: *arXiv preprint arXiv:2104.02395* (2021).
- [86] Alejandro Barredo Arrieta y col. «Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI». En: *Information fusion* 58 (2020), págs. 82-115.
- [87] Plamen Angelov y Eduardo Soares. «Towards explainable deep neural networks (xDNN)». En: *Neural networks* 130 (2020), págs. 185-194.
- [88] Jaegul Choo y Shixia Liu. «Visual analytics for explainable deep learning». En: *IEEE computer graphics and applications* 38.4 (2018), págs. 84-92.
- [89] Ramprasaath R Selvaraju y col. «Grad-cam: Visual explanations from deep networks via gradient-based localization». En: *International conference on computer vision*. 2017, págs. 618-626.

Apéndice A

Tablas de resultados por ejecución

A.1. Resultados divididos en niños/adultos/ancianos

Red	Datos	Rango	N muestras	MAE Ej. 1	MAE Ej. 2	MAE Ej. 3	MAE Ej. 4	MAE Ej. 5
Resnet-CNN	RGB	Total	2200	7,9314	8,3572	7,9955	7,9212	9,4296
		15 - 18	200	8,7837	10,2417	5,9555	7,3788	15,6252
		19 - 64	1960	7,4253	7,6140	7,7365	7,3946	8,3521
		65 - 90	40	28,4655	35,3538	30,8857	36,4374	31,2462
Panorama-CNN	RGB	Total	2200	8,9679	8,7470	9,0275	9,5631	9,5110
		15 - 18	200	12,5989	10,7610	10,9665	15,2166	14,1037
		19 - 64	1960	8,1831	8,1074	8,3479	8,6835	8,6625
		65 - 90	40	29,2660	30,0180	32,6305	24,3943	28,1207
Panorama-CNN	FM	Total	2200	8,9952	8,7591	9,6103	11,1693	9,5134
		15 - 18	200	10,4945	9,3871	14,9872	19,9334	9,6707
		19 - 64	1960	8,2038	8,1445	8,5892	9,7888	8,9849
		65 - 90	40	40,2778	35,7361	32,7574	34,9959	34,6228

Tabla A.1: Resultados MAE para niños/adultos/ancianos por ejecución para un conjunto de test con aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.

Red	Datos	Rango	N muestras	RMSE Ej. 1	RMSE Ej. 2	RMSE Ej. 3	RMSE Ej. 4	RMSE Ej. 5
Resnet-CNN	RGB	Total	2200	10,8660	11,0961	10,8158	10,7275	12,3852
		15 - 18	200	13,3955	13,2536	9,7012	9,8715	18,6543
		19 - 64	1960	9,8041	9,6963	10,0466	9,5148	10,7415
		65 - 90	40	28,5566	34,9130	30,7633	36,6846	31,2323
Panorama-CNN	RGB	Total	2200	11,2525	11,1452	11,7146	12,3443	12,0306
		15 - 18	200	14,5320	12,8638	14,2962	18,6303	16,4228
		19 - 64	1960	10,1236	10,1724	10,5069	11,0904	10,8572
		65 - 90	40	29,7684	30,5620	33,4142	24,8788	28,9102
Panorama-CNN	FM	Total	2200	11,8915	11,3773	12,1722	14,0285	12,4515
		15 - 18	200	11,9300	11,1503	15,9120	21,0476	11,9587
		19 - 64	1960	10,5176	10,3080	10,8813	12,2539	11,6006
		65 - 90	40	40,5639	35,9341	32,8874	35,4269	34,9012

Tabla A.2: Resultados RMSE para niños/adultos/ancianos por ejecución para un conjunto de test con aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.

Red	Datos	Rango	N muestras	MAE Ej. 1	MAE Ej. 2	MAE Ej. 3	MAE Ej. 4	MAE Ej. 5
Resnet-CNN	RGB	Total	110	7,9559	8,6959	7,9509	8,1714	9,9908
		15 - 18	10	7,6220	11,4919	4,6922	7,6177	18,2280
		19 - 64	98	7,2221	7,7067	7,5093	7,4170	8,4567
		65 - 90	2	45,5809	43,1863	45,8832	47,9058	43,9765
Panorama-CNN	RGB	Total	110	8,6824	8,4229	8,1783	9,0293	9,2644
		15 - 18	10	11,2255	10,1265	5,6543	10,4255	8,8535
		19 - 64	98	8,0021	7,7761	7,9468	8,5652	8,8766
		65 - 90	2	29,3023	31,5948	32,1395	24,7855	30,3232
Panorama-CNN	FM	Total	110	13,8048	10,8471	14,4774	10,5711	15,4329
		15 - 18	10	18,4673	11,9696	23,3531	14,5762	11,2600
		19 - 64	98	13,0031	10,2864	13,3558	9,5977	15,5646
		65 - 90	2	29,7750	32,7101	25,0613	38,2409	29,8486

Tabla A.3: Resultados MAE para niños/adultos/ancianos por ejecución para un conjunto de test sin aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.

Red	Datos	Rango	N muestras	RMSE Ej. 1	RMSE Ej. 2	RMSE Ej. 3	RMSE Ej. 4	RMSE Ej. 5
Resnet-CNN	RGB	Total	110	11,4563	11,8186	11,0969	11,5873	13,4738
		15 - 18	10	11,4793	10,9968	5,3717	10,2310	18,7416
		19 - 64	98	9,3496	9,8636	9,2616	9,3401	10,8738
		65 - 90	2	29,1553	30,9213	31,6210	37,1808	32,8850
Panorama-CNN	RGB	Total	110	10,7899	10,8913	10,7013	11,3014	11,7629
		15 - 18	10	12,8012	12,2729	8,0277	13,6585	10,6424
		19 - 64	98	9,7763	9,8386	10,0335	10,5486	11,1532
		65 - 90	2	30,0111	32,0611	32,3004	25,2945	30,8006
Panorama-CNN	FM	Total	110	15,9380	12,9359	16,7342	13,3244	17,6643
		15 - 18	10	19,0945	12,8721	23,6777	16,0349	13,5631
		19 - 64	98	15,1601	12,2074	15,6264	11,9686	17,6946
		65 - 90	2	29,7795	32,7573	25,1812	38,2404	30,0121

Tabla A.4: Resultados RMSE para niños/adultos/ancianos por ejecución para un conjunto de test sin aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.

A.2. Resultados divididos en los rangos de Todd

Red	Datos	Rango	N muestras	MAE Ej. 1	MAE Ej. 2	MAE Ej. 3	MAE Ej. 4	MAE Ej. 5
Resnet-CNN	RGB	Total	2200	7,9314	8,3572	7,9955	7,9212	9,4296
		0 - 19	240	10,2444	11,3616	7,7281	7,5962	17,7765
		20 - 21	140	7,4668	9,4589	8,2974	7,2980	15,0192
		22 - 24	160	11,7185	11,4324	10,7649	7,4689	11,8624
		25 - 26	40	4,0259	6,5470	2,5099	4,9895	4,6455
		27 - 30	160	6,3039	8,4167	6,8396	5,8505	7,2413
		31 - 34	120	4,3255	3,5232	6,4975	3,5549	4,3556
		35 - 39	300	7,7179	7,6636	7,6226	6,5717	6,5658
		40 - 44	300	6,1822	6,0077	6,6314	6,1659	6,8123
		45 - 49	300	7,0748	5,3253	6,2718	8,0270	5,7829
Panorama-CNN	RGB	50 - 90	440	9,2928	10,8532	10,7253	12,7164	11,0515
		Total	2200	8,9679	8,7470	9,0275	9,5631	9,5110
		0 - 19	240	13,5796	12,5246	10,5341	15,8030	14,0122
		20 - 21	140	13,2367	10,0039	12,9574	17,2151	16,8576
		22 - 24	160	15,1049	10,9515	8,5311	11,8365	9,9747
		25 - 26	40	5,3276	2,5300	3,7151	4,5044	4,9918
		27 - 30	160	9,5048	6,9422	10,0136	12,0430	10,3855
		31 - 34	120	6,3074	5,5699	6,1655	7,9737	6,9406
		35 - 39	300	8,2350	8,8816	7,8664	7,8682	7,1197
		40 - 44	300	5,2846	5,4528	5,0829	4,7709	5,0216
Panorama-CNN	FM	45 - 49	300	5,0436	6,4063	6,5666	5,9190	6,6994
		50 - 90	440	9,4103	11,3232	13,1995	9,7974	11,9518
		Total	2200	8,9952	8,7591	9,6103	11,1693	9,5134
		0 - 19	240	10,3126	9,3661	14,7643	20,0096	9,8630
		20 - 21	140	14,2598	15,0176	18,5323	18,4791	16,1167
		22 - 24	160	9,7669	9,5983	12,8986	12,9747	10,5002
		25 - 26	40	3,5748	3,9037	7,2250	10,7660	4,2014
		27 - 30	160	8,2268	9,7728	10,6481	16,1611	10,5695
		31 - 34	120	10,3953	9,6225	11,4595	11,9409	12,6589
		35 - 39	300	7,7019	8,1539	7,9105	6,5121	9,1526
Panorama-CNN	FM	40 - 44	300	4,4520	4,6542	3,9104	6,4351	5,9597
		45 - 49	300	5,9122	5,8110	4,9961	5,6142	5,7860
		50 - 90	440	12,7926	11,1905	10,2909	11,5669	11,3143

Tabla A.5: Resultados MAE para cada fase de Todd por ejecución para un conjunto de test con aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.

Red	Datos	Rango	N muestras	RMSE Ej. 1	RMSE Ej. 2	RMSE Ej. 3	RMSE Ej. 4	RMSE Ej. 5
Resnet-CNN	RGB	Total	2200	10,8660	11,0961	10,8158	10,7275	12,3852
		0 - 19	240	15,0133	14,6998	12,8191	10,4116	20,8596
		20 - 21	140	11,3527	12,8173	11,9162	10,2541	16,6082
		22 - 24	160	13,4050	12,7079	13,5200	9,4566	13,6571
		25 - 26	40	4,2940	7,2539	2,5838	5,2092	4,7390
		27 - 30	160	8,6572	10,5823	8,7061	8,0164	9,1400
		31 - 34	120	6,0080	4,8651	8,5832	4,6721	5,5325
		35 - 39	300	8,8513	8,8576	9,2440	8,0309	7,6495
		40 - 44	300	7,5672	7,0756	7,9845	7,5324	8,0983
		45 - 49	300	9,5881	6,7326	7,9249	9,9783	7,0167
		50 - 90	440	12,6750	14,4950	13,5769	16,2012	14,4548
Panorama-CNN	RGB	Total	2200	11,2525	11,1452	11,7146	12,3443	12,0306
		0 - 19	240	15,8114	14,9466	13,7812	19,0871	16,2452
		20 - 21	140	14,6032	12,8674	14,7197	19,3135	18,2566
		22 - 24	160	16,5623	12,6613	10,9079	14,1679	12,4297
		25 - 26	40	6,1893	3,2337	4,9737	5,9028	6,0474
		27 - 30	160	11,3991	9,1171	11,7578	13,7228	12,0648
		31 - 34	120	7,5587	6,7862	7,4289	9,8033	8,7993
		35 - 39	300	9,1965	10,0879	9,4980	9,2878	8,6803
		40 - 44	300	6,4467	6,6628	6,6845	6,0819	6,2436
		45 - 49	300	6,2096	7,8311	8,0351	7,1644	8,0253
		50 - 90	440	12,2684	14,2493	16,3339	12,4355	14,7734
Panorama-CNN	FM	Total	2200	11,8915	11,3773	12,1722	14,0285	12,4515
		0 - 19	240	11,7323	11,0583	15,6487	21,0457	12,0253
		20 - 21	140	15,6321	16,5203	19,1908	19,9931	17,8560
		22 - 24	160	12,2216	12,0509	14,5419	14,7193	14,0684
		25 - 26	40	4,8360	4,9492	8,4769	12,6437	5,7137
		27 - 30	160	10,6841	12,6082	12,8109	19,0542	13,3259
		31 - 34	120	13,2264	11,5039	13,5203	13,5722	15,3260
		35 - 39	300	9,4607	9,4807	9,3709	7,8520	11,0931
		40 - 44	300	5,6476	5,8816	5,2766	8,0889	7,6988
		45 - 49	300	7,4271	7,0652	6,1234	6,7704	7,3958
		50 - 90	440	16,9755	15,0280	13,8818	14,9025	15,1558

Tabla A.6: Resultados RMSE para cada fase de Todd por ejecución para un conjunto de test con aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.

Red	Datos	Rango	N muestras	MAE Ej. 1	MAE Ej. 2	MAE Ej. 3	MAE Ej. 4	MAE Ej. 5
Resnet-CNN	RGB	Total	110	7,9559	8,6959	7,9509	8,1714	9,9908
		0 - 19	12	8,3126	12,4361	5,5283	8,1234	19,4906
		20 - 21	7	7,2323	8,9566	6,8303	9,7780	16,5051
		22 - 24	8	10,9577	12,8129	9,1918	4,6329	9,3991
		25 - 26	2	9,3743	11,4269	9,2597	9,5899	9,3269
		27 - 30	8	4,9220	8,0102	5,9934	5,2420	7,4782
		31 - 34	6	4,0221	3,4917	6,7192	3,7258	2,9868
		35 - 39	15	7,4971	6,7433	7,3873	5,8176	6,4737
		40 - 44	15	5,9513	6,3646	6,0109	6,1192	7,4830
		45 - 49	15	8,4702	5,1802	7,9864	8,6672	5,8453
		50 - 90	22	10,2760	11,8141	11,7890	13,7881	12,7703
Panorama-CNN	RGB	Total	110	8,6824	8,4229	8,1783	9,0293	9,2644
		0 - 19	12	12,1044	12,5993	5,2835	9,5022	8,3543
		20 - 21	7	13,8104	8,8566	13,9004	18,3150	18,9407
		22 - 24	8	14,7487	8,7233	5,3222	7,0063	5,9133
		25 - 26	2	3,7319	2,1470	1,2017	1,6476	2,5518
		27 - 30	8	10,6987	6,3041	8,7968	11,9133	10,1549
		31 - 34	6	4,7231	5,3529	6,1767	9,5100	8,1790
		35 - 39	15	8,3287	7,8968	8,2253	8,3143	7,1605
		40 - 44	15	4,8656	4,9016	5,5550	5,5227	4,2848
		45 - 49	15	5,3418	7,3815	6,2902	6,7917	8,9908
		50 - 90	22	8,8964	11,5452	12,9743	10,4475	13,4993
Panorama-CNN	FM	Total	110	13,8048	10,8471	14,4774	10,5711	15,4329
		0 - 19	12	18,7808	12,3242	23,7055	14,8577	11,4966
		20 - 21	7	24,3938	19,0262	26,4310	18,3901	25,7490
		22 - 24	8	17,7900	14,2276	20,6874	11,6351	17,2422
		25 - 26	2	14,5316	10,4538	17,5029	4,3803	17,4294
		27 - 30	8	16,5627	16,3594	21,2071	17,7360	22,9483
		31 - 34	6	16,5327	11,4134	17,1938	9,7987	19,6925
		35 - 39	15	13,9417	10,1542	14,9934	5,5860	17,7979
		40 - 44	15	11,1354	7,6500	10,5236	5,3824	15,5240
		45 - 49	15	11,2540	5,9644	8,2380	6,2441	13,4228
		50 - 90	22	7,9252	10,0680	6,5174	13,4131	9,2597

Tabla A.7: Resultados MAE para cada fase de Todd por ejecución para un conjunto de test sin aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.

Red	Datos	Rango	N muestras	RMSE Ej. 1	RMSE Ej. 2	RMSE Ej. 3	RMSE Ej. 4	RMSE Ej. 5
Resnet-CNN	RGB	Total	110	11,4563	11,8186	11,0969	11,5873	13,4738
		0 - 19	12	12,1321	15,5210	7,1995	11,0927	22,0112
		20 - 21	7	9,6258	10,2376	9,6595	11,7392	17,7956
		22 - 24	8	12,1109	12,7998	11,0088	5,4742	11,5523
		25 - 26	2	5,1564	5,9884	2,2329	4,6368	4,1556
		27 - 30	8	8,5503	12,0394	10,6665	8,6927	9,3395
		31 - 34	6	3,3850	1,7601	7,4393	3,7772	4,2510
		35 - 39	15	9,4695	8,5713	9,0381	7,2463	7,4868
		40 - 44	15	7,2905	7,0801	7,2194	7,1995	8,9043
		45 - 49	15	10,8885	6,4376	9,4284	10,9341	6,7697
		50 - 90	22	12,3744	13,4763	13,2409	15,9378	14,4696
Panorama-CNN	RGB	Total	110	10,7899	10,8913	10,7013	11,3014	11,7629
		0 - 19	12	14,3451	15,1863	7,4764	12,6975	10,2848
		20 - 21	7	14,6160	11,7608	14,8901	20,1044	20,6504
		22 - 24	8	15,7087	9,7432	8,3004	8,9551	7,8240
		25 - 26	2	4,8386	2,1472	1,2636	2,2440	2,5627
		27 - 30	8	12,4380	8,0588	9,6464	12,8393	11,0239
		31 - 34	6	5,3778	5,7732	7,6919	11,1772	9,7415
		35 - 39	15	9,2249	9,7319	10,1608	9,3687	9,2165
		40 - 44	15	5,8420	6,0174	7,5041	7,0972	5,4596
		45 - 49	15	6,0629	8,7198	7,4912	8,0487	10,2921
		50 - 90	22	11,8125	14,6475	15,7903	12,6526	15,9709
Panorama-CNN	FM	Total	110	15,9380	12,9359	16,7342	13,3244	17,6643
		0 - 19	12	19,6461	13,2327	23,9987	16,3359	13,6004
		20 - 21	7	24,6615	20,5749	26,8990	18,9603	26,4789
		22 - 24	8	18,3858	15,1681	20,8436	13,5334	18,4735
		25 - 26	2	14,6731	10,7892	17,5567	5,2026	17,4482
		27 - 30	8	17,2077	17,1045	21,5528	20,5175	23,6186
		31 - 34	6	17,4816	12,4360	17,9059	11,0438	20,7564
		35 - 39	15	15,9846	11,7423	15,9157	6,8545	19,2684
		40 - 44	15	12,4647	9,2906	11,3609	6,6478	17,4764
		45 - 49	15	12,9568	7,0859	9,5657	7,4643	14,7808
		50 - 90	22	11,7457	13,2881	9,5066	16,5047	12,4003

Tabla A.8: Resultados RMSE para cada fase de Todd por ejecución para un conjunto de test con aumento de datos. **Datos:** RGB, imágenes RGB; FM, mapas de características.