

Article

# Adult Skeletal Age-at-Death Estimation through Deep Random Neural Networks: A New Method and Its Computational Analysis

David Navega <sup>1,2,\*</sup> , Ernesto Costa <sup>3</sup> and Eugénia Cunha <sup>1,2</sup> 

- <sup>1</sup> Centre for Functional Ecology (CEF), Laboratory of Forensic Anthropology, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal; eugenia.m.cunha@inmlcf.mj.pt
- <sup>2</sup> National Institute of Legal Medicine and Forensic Sciences, 3000-548 Coimbra, Portugal
- <sup>3</sup> Centre for Informatics and Systems of the University of Coimbra (CISUC), Evolutionary and Complex Systems Group (ECOS), Department of Informatics Engineering, University of Coimbra, 3030-290 Coimbra, Portugal; ernesto@dei.uc.pt
- \* Correspondence: dsnavega@gmail.com

**Simple Summary:** Age-at-death is of paramount importance in forensic analysis of skeletal remains. In addition to sex, stature, and population affinity, it constitutes baseline information in the identification process of deceased individuals. Despite its long tradition, in anthropological research age-at-death estimation poses many challenges and unanswered questions. It is undisputedly among the most difficult tasks of the forensic anthropologist and its results are often subject to a lack of performance. In this study, we assessed computationally the efficiency of a holistic approach to skeletal age estimation based on a new proposal for macroscopic examination and the use of machine learning-based models for data analysis. Our results suggest that this approach is key for accurate and efficient age-at-death estimation based on skeletal remains analysis.



**Citation:** Navega, D.; Costa, E.; Cunha, E. Adult Skeletal Age-at-Death Estimation through Deep Random Neural Networks: A New Method and Its Computational Analysis. *Biology* **2022**, *11*, 532. <https://doi.org/10.3390/biology11040532>

Academic Editors: Stefan Bonn and Andrés Moya

Received: 11 February 2022

Accepted: 28 March 2022

Published: 30 March 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Abstract:** Age-at-death assessment is a crucial step in the identification process of skeletal human remains. Nonetheless, in adult individuals this task is particularly difficult to achieve with reasonable accuracy due to high variability in the senescence processes. To improve the accuracy of age-at-estimation, in this work we propose a new method based on a multifactorial macroscopic analysis and deep random neural network models. A sample of 500 identified skeletons was used to establish a reference dataset (age-at-death: 19–101 years old, 250 males and 250 females). A total of 64 skeletal traits are covered in the proposed macroscopic technique. Age-at-death estimation is tackled from a function approximation perspective and a regression approach is used to infer both point and prediction interval estimates. Based on cross-validation and computational experiments, our results demonstrate that age estimation from skeletal remains can be accurately (~6 years mean absolute error) inferred across the entire adult age span and informative estimates and prediction intervals can be obtained for the elderly population. A novel software tool, DRNNAGE, was made available to the community.

**Keywords:** forensic anthropology; age-at-death estimation; machine learning; neural networks



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Forensic anthropology (FA) has become a major component of forensic sciences. During recent decades, a profound change, a true paradigm change, has taken place and forensic anthropology has transformed itself into a discipline with its own theoretical and conceptual corpus and research agenda. It can be stated that the discipline and its attributes have evolved significantly. In fact, this evolution has been so marked and drastic that it can be argued that even some of the most experienced and long-term practicing anthropologists may have trouble conceptualizing and being fully proficient in the many areas now

covered by the discipline [1,2], or in even being able to foresee all possible interdisciplinary and technological developments. Nonetheless, biological profile estimation from human skeletal remains constitutes a pivotal task and inferring age-at-death, sex, stature, and population affinities is a fundamental step of the anthropological analysis in the context of the medico-legal identification process.

In the identification process of human remains, age-at-death is a major screening factor that helps reduce the universe of possible matches. Therefore, an estimate of this biological parameter is a normal request from police forces and judicial entities [3]. This process relies on a meticulous analysis of skeletal and dental structures with an association with chronological age-at-death. Although this is a topic in which significant research has been performed in recent decades, skeletal age estimation of adult remains continues to present many unanswered questions and challenges, especially for the elderly. Determining how to handle age estimation using multiple skeletal age-related traits remains among the problems most commonly identified for which a satisfactory solution has not yet been presented and research further is required [3–10]. Moreover, computational and statistical methods employed in the creation of age estimation techniques have been a topic of debate and contention [11–24].

The present work aims to lay a foundation to tackle some of the challenges of morphoscopic adult skeletal age estimation, especially in terms of its holistic or multifactorial aspect. Several authors argue in favor of multifactorial age estimation to obtain precise and accurate age estimates [9,16,25]. Nonetheless, multifactorial age estimation poses its own challenges and limitations, and is a topic with a clear lack of consensus [5,10]. Conceptually multifactorial age estimation can be argued as being the most effective approach for age estimation because morphological indicators display different age-related trajectories and have different underlying biological processes.

The symphyseal face of the pubic bone, for instance, has been systematically studied, ranging from the pioneering studies that established the morphological analysis of this skeletal marker as an age estimation technique, to modern fully computational frameworks for age estimation [26–34]. However, other skeletal markers and regions that can convey important age-related information, such as the degeneration of vertebral bodies, joint margins, or the roughening of muscle and tendon attachment sites, have received scarce attention as aging markers. The unimpressive accuracy and precision associated with the multiple iterations of pubic symphysis aging techniques, one of the most used and favored techniques for age estimation [5], underlines the idea that further developments and over-analysis of specific skeletal markers in isolation is not likely to result in substantial improvements over the state-of-art of adult age estimation, but rather a more comprehensive array of skeletal markers and features provide a more fertile ground for further developments [35,36].

A multifactorial morphoscopic approach to skeletal analysis does not solve, in itself, the many difficulties faced in the age-at-death assessment. In fact, if not correctly designed, this approach can become methodologically cumbersome from a data collection and analysis perspective. From an analytical and statistical perspective, collecting more data from the skeleton increases the chance of encountering issues of redundancy, multicollinearity, and a dimensionality that hinders the straightforward interpretability and pragmatic value of morphoscopic analysis. From a practical point of view, a more comprehensive analysis of the age-related skeletal features requires a higher level of expertise on how to collect the skeletal features. This issue is of great relevance for approaches that rely on morphoscopic analysis of the skeleton. Moreover, in forensic contexts it is common that the skeletal remains are somehow fragmentary or incomplete due to a multitude of taphonomic factors, which means that not all age-related traits will be available for every unidentified deceased. From a practitioner's perspective, this translates into the need for computational and software tools that can fit or train age-at-death estimation models on a case-by-case basis.

To cope with the difficulties and needs of multifactorial age estimation, novel methods and techniques can be developed by resorting to statistical and machine learning, data

science, and artificial intelligence tools and approaches. More than constantly evolving, machine learning, artificial intelligence and data science are ubiquitous, and have various successful applications within forensic anthropology in domains such as biological profiling or craniofacial identification [13,15,37–41].

This work aims to provide a new method, and its computational analysis, for multifactorial skeletal age-at-death estimation of adult humans supported by a machine learning approach based on a deep randomized neural network. This manuscript is in its essence methodological, presenting both a new macroscopic technique for skeletal analysis and a detailed explanation of a computational framework to obtain age-at-death estimates and model their uncertainty. New age-at-death estimation software, DRNNAGE, that translates the in silico key points of the work presented here into an actionable tool, was developed and is a major research product.

## 2. Materials and Methods

### 2.1. Dataset

#### 2.1.1. Sampled Identified Skeletal Collections

To implement and pursue a computational analysis of the novel age-at-death estimation method proposed in this work, a reference dataset of 500 individuals was constructed. A total of 99 features were collected covering all key traditional age-related and other under-explored skeletal traits. Accounting for laterality, 64 unique traits can be analyzed from the axial and appendicular skeleton using the new macroscopic scoring method, whose rationale and details are described and explored in Section 2.2.

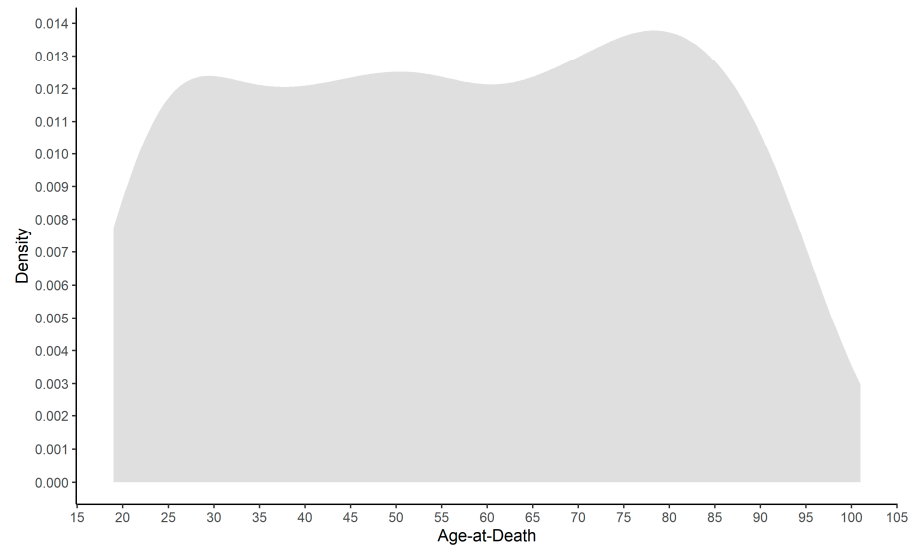
The 500 individuals were sampled from two identified skeletal collections hosted at the Department of Life Sciences at the University of Coimbra, Portugal—the Coimbra Identified Skeletal Collection (CISC) and the 21st Century Identified Skeletal Collection (XXI-ISC). The CISC consists of 505 individuals with age-at-death ranging from 7 to 96 years representing skeletons from the Cemitério da Conchada, that were born between 1817 and 1924 and died from 1904 to 1938 [42]. The XXI-ISC collection is currently composed of 302 skeletons of both sexes, mostly represented by elderly individuals. This collection represents Portuguese nationals who died between 1982 and 2012 and were exhumed between 1999 and 2016 from a main cemetery in Santarém. More details are found in [43,44]. Demographic parameters of the sampled individuals in our study are detailed in Table 1. All sampled individuals presented fully developed long bones. No individual was excluded due to pathology or taphonomy.

**Table 1.** Demographic characterization of reference data sampled from the CISC and XXI-ISC collections.

		CISC		XXI-ISC		Pooled Collections		Pooled Sex
		Female	Male	Female	Male	Female	Male	
Age-at-Death (AGE)	<i>n</i>	168	166	82	84	250	250	500
	Mean	48.482	45.331	81.841	74.881	59.424	55.260	57.34
	Std. Dev.	19.483	18.171	12.889	15.082	23.556	22.141	22.93
	Min.	19	19	38	25	19	19	19
Year of Birth (YOB)	Max.	95	96	101	96	101	96	101
	Mean	1877.286	1879.994	1923.866	1930.560	1892.564	1896.984	1894.774
	Std. Dev.	21.252	19.948	13.137	14.424	28.969	30.096	29.591
	Min.	1830	1836	1904	1908	1830	1836	1830
Year of Death (YOD)	Max.	1911	1917	1970	1982	1970	1982	1982
	Mean	1925.768	1925.325	2005.707	2005.440	1951.988	1952.244	1952.116
	Std. Dev.	6.597	7.343	3.707	3.919	38.051	38.452	38.214
	Min.	1910	1910	2000	1995	1910	1910	1910
	Max.	1936	1936	2012	2011	2012	2011	2012

The sampled reference dataset is composed of 250 male and 250 female individuals who died at the age of 19 to 101 years old (mean = 57.34, SD = 22.93). Age-at-death distribution is homogenous across the age span represented, with the exception of individuals over 95 years old (Figure 1). A homogenous and uniform age-at-death distribution is a simple

yet vital strategy to cope with the problem of age-mimicry [45] and to guarantee that the targeted age span is fully represented.



**Figure 1.** Pooled age-at-death distribution (KDE).

Sampled individuals were born between 1830 and 1982 and died between 1910 and 2012. Despite the large temporal frame represented, there is a continuum and a wide range over the age-at-death distribution that makes this sample particularly suited for age-related research.

### 2.1.2. Data Management and Processing

As previously mentioned, multifactorial age estimation poses many challenges that are mostly related to data management and processing. Two common problems that arise are redundancy and missing data. Redundancy is always involved when bilateral or paired data is collected. The human body is not fully symmetric; yet it is not expected that the left and right diverge drastically under normal conditions. Missing data in FA results mostly from taphonomic factors. To cope with redundancy and missing values, a strategy based on domain heuristics and imputation techniques was pursued. For bilateral traits, the left side was selected as the main source of data. If the left score for a given bilateral trait was missing, the right side was used as a surrogate value. Once this first heuristic was applied, the remaining missing values were imputed using a simple nearest neighbor ( $k = 1$ ) procedure by substituting all missing value of given individual by the values of the nearest neighbor. Jaccard similarity on one-hot encoded data was used to compute the nearest matches. The followed procedure minimized redundancy and dimensionality by reducing the number of skeletal features from 99 to 64. A simple nearest neighbor with  $k = 1$  according to Beretta and Santianello [46] is the preferred strategy to preserve the structure of a dataset. The authors demonstrated that more advanced algorithms reduced imputation error but introduced significant data distortion. To increase the volume and age-related variability of the data available, sexes were pooled. Although this choice seems arbitrary, it is important to note that, in FA, sex is usually estimated during casework. Pooled data models balance out the potential and pitfalls of sex-specific models and their mis-specifications.

Missing values represented 9.52% of the total entries of the data table when bilateral data were considered, and 6.89% when the domain heuristic described was first applied as a naïve imputation mechanism and strategy to handle bilateral data redundancy.

## 2.2. A Novel Technique for Macroscopic Age-At-Death Estimation

A key contribution of the present work to the topic of macroscopic skeletal age estimation in adults is the proposal of new scoring schemes for well-established and under-explored skeletal traits that can be used as biomarkers in age-at-death assessment. The development of a new scoring system emerged from the necessity for standardization of a data collection, and a generation mechanism that was more aligned with a multifactorial approach to age estimation and more suitable multivariate data analysis, while keeping in mind practical aspects such as observation error and ease of application.

The proposed morphoscopic method strives to be comprehensive and to incorporate features from as many skeletal elements as possible. Envisioning the whole skeleton as a biomarker for age estimation, it is more likely that the overall skeletal patterns exhibit a stronger and monotonic relationship with age-at-death, which is pivotal for accurate predictions. The rate and nature of overall skeletal changes also have a greater chance to be consistent across individuals since a holistic approach can encapsulate intra and interpersonal variation with greater finesse [35]. Analyzing multiple traits also offsets the intrinsic limitation to specific traits when analyzed on their own [47].

Following a component-based approach, up to 64 unique skeletal traits can be scored using the scheme outlined in the next subsections. The covered skeletal traits encode both developmental and degenerative aspects from different anatomical regions. Despite the large number of features analyzed in this proposal, all skeletal features are limited to morphological variables with no more than three classes or stages. Such specifications were established during the several iterations of the development and refinement of the system proposed, and by following guidelines from the literature. Shirley and Montes [48] empirically addressed the old methodological debate of phase versus component-based approach. Their study quantified the observation error of a phase and a component-based method, and the results suggest that a component-based approach offers a more objective scoring if the number of coding possibilities in each component does not exceed three levels of expression.

The following subsections provide a brief overview of the existing scoring methods for specific skeletal region or traits, the novel scoring schemes proposed in this work, and the rationale and difficulties faced during method development. Due to the constraints of space and manuscript presentation, full descriptions of the trait scoring systems developed in this study are provided in Tables S1–S15 of the Supplementary Material. The skeletal scoring systems are also embedded in the developed software (see Section 2.6.4).

### 2.2.1. Cranial and Palatine Suture Scoring

The scoring system used for the cranial and palatine sutures consists of a modification and binarization of the proposal by Boldsen et al. [19]. This system was selected because it incorporates much of the rationale of older methods for scoring ectocranial sutures (neurocranium) and the palatine sutures [49–56]. The simplification to a binary scoring system resulted from the difficulty during preliminary and training sessions to differentiate and consistently score the adjacent stage (i.e., open to juxtaposed or partially obliterated to punctuated). The scoring scheme described in Table S2 should be applied to nine sutural segments from the palatine, the sagittal, coronal, and lambdoid sutures (Table S1).

### 2.2.2. Vertebrae Development and Degeneration Scoring

The fusion of the bodies of the first and second sacral vertebrae is also part of the skeletal markers analyzed in the proposed protocol. This skeletal feature is one of the few developmental traits that persist through early adulthood. Its usefulness as an indicator to distinguish young adults was demonstrated by several researchers [57–59]. This trait was assessed with a binary scale described in Table S3. To incorporate both metamorphic and degenerative traits of the vertebral column, a three-stage scoring scheme was devised, building upon previous work from Snodgrass [60], Watanabe and Terazawa [61], and Albert et al. [62]. The first two methods focus on the degeneration and osteophyte formation on

the margins of the vertebral bodies, whereas the last work focuses on the development of the vertebral epiphyseal rings and body morphology. The proposed system, Table S4, applies to superior and inferior surfaces of the third to seven cervical vertebrae, the first to fifth lumbar vertebrae, and the superior surface of the first sacral vertebra. Table S5 lists all features analyzed in the axial skeleton (excluding sacral auricular surfaces).

### 2.2.3. Joint and Musculoskeletal Degeneration Scoring

Osteoarthritis and enthesal changes have been traditionally analyzed in physical anthropology and bioarcheology as markers of health and biomechanical stress, and tentative indicators of physical activity patterns. According to Milner and Boldsen [35], who advocate a more detailed analysis of this type of skeletal marker, these features collectively contribute to an increase in accuracy and precision of age estimation. The authors base such an assertion on empirical evidence from an experience-based procedure where these types of skeletal traits were extensively used. Several reasons can be noted for why osteoarthritis and enthesal changes have been overlooked or not systematically analyzed in the past as age markers. Broadly speaking, due to their degenerative nature and late onset, it is believed that they provide limited information, distinguishing only in a broad sense young from older individuals. More specifically, osteoarthritis increases with age but has a complex and multifactorial etiology that hinders or masks its relationship with age-at-death. Enthesal changes have traditionally been assessed as musculoskeletal stress markers and as tentative clues to infer physical and occupational activity. This possible relation to activity can interfere in the expression and variation of enthesal morphology and affects its relationship with the aging process. However, recent and systematic studies conducted on identified skeletal collections show that age-at-death is one of the most relevant factors, or even the only one with statistical significance, in the expression of such skeletal traits [63–70].

Developing a scoring procedure for these features proved to be one of the most challenging aspects of method development. The difficulties faced were mostly related to the fact that analyzing joint and musculoskeletal degeneration involves many skeletal elements, which translate into high dimensionality of the collected data. This high dimensionality poses two major problems: increased chance of collinearity, which poses computational issues, and loss of pragmatic value. To tackle the high dimensionality and subsequent issues found when scoring joint and musculoskeletal degeneration, a new binary procedure was developed. The system retains the analysis of the type of traits evaluated in Buikstra and Ubelaker [71] and Henderson et al. [72] but simplifies the scoring to a simple absence or presence of degenerative traits as a whole for any particular anatomical structure. The generic binary scoring system both for joint and musculoskeletal degenerative changes are presented in Tables S7 and S8. The scoring system applies to five major anatomical complexes from the upper and lower limb: shoulder, elbow, hip, knee, and ankle (Table S6). To enhance the analysis of these traits we provide specific scoring descriptions for Stage 1 of some traits (Table S9).

### 2.2.4. Clavicle Sternal and Acromial Ends Scoring

The macroscopic analysis of the clavicle has a long standing in skeletal age estimation. Nonetheless, its focus has been mostly in the epiphyseal fusion of the sternal end [73–76]. Sternal epiphyseal fusion of the clavicle is a key trait to obtain precise age estimate in young adult individuals due to the late total development of this structure around the 30 s. Falys and Prangle [73] were the first to propose a method to score post-epiphyseal changes in the clavicle for age estimation purposes. The authors suggest a scoring system focused on surface topography, porosity, and marginal osteophyte formation, providing a regression model for age estimation. A new scoring scheme that integrates both developmental and degenerative changes in the sternal and acromial ends of the clavicle is proposed. A full description of the traits analyzed is available in Table S10.

### 2.2.5. First Rib Costal Face and Tubercle Scoring

The metamorphosis of the sternal end of the ribs emerged in the mid-1980s as a new age estimation technique. İşcan, Loth and colleagues described multiple morphologic features that characterize the metamorphosis of the sternal end of the ribs, with particular emphasis on the fourth rib costal face [77–80]. This approach proved to be an effective alternative to existing methods. Nonetheless, several disadvantages have been pointed out, such as the difficulty in identifying the fourth rib in disarticulated skeletal remains and the fact the morphology of the costal face is not the only component of the age-related changes in rib morphology. To address these problems, Kunos et al. [81] described a new age estimation method based on the metamorphosis of the costal face, head, and tubercle of the first rib. The first rib has the key advantage of having a morphology that is straightforward to individualize. DiGangi et al. [82] improved upon the work of Kunos et al. [81] and proposed a revised method for age estimation based on the costal face and tubercle morphology. A new scoring method is proposed in this study that build upon previous work by Kunos et al. and DiGangi et al. [81,82]. This new system simplifies the scoring of the costal face morphology to a three-stage coding and the morphology of the tubercle is evaluated in a binary fashion (Table S11).

### 2.2.6. Pubic Symphysis Scoring

The metamorphosis of pubic symphysis is the most popular osteological marker used in adult skeletal age estimation. The previous attention paid to this anatomical structure is not misplaced; however, the over-reliance on this indicator can be explained by the progressive metamorphic features that have enough expression variation to allow an exhaustive morphological description using different scoring schemes and different types of supporting materials such as casts. A simple component-based system was developed focused on the metamorphic and degenerative changes in three features of this structure: rim development, topography, and texture of the symphyseal face. These three components are assessed with a three-stage coding system emphasizing early metamorphic or development traits, such as the presence of billowing (a pattern of transverse ridges and furrows) and late degenerative traits, such as the flattening and erosion of the symphyseal face. A full description of the scoring system is given in Table S12. The proposed system is based on previous work by Todd [30,31] and Brooks and Suchey [26].

### 2.2.7. Sacral and Iliac Auricular Surfaces (Sacroiliac Joint) Scoring

The description of age-related changes in the sacro-iliac joint can be traced back to Sashin [83] and Schunke [84], but its usage as an age indicator its mostly due to the work of Lovejoy and colleagues [85] and Buckberry and Chamberlain [86] on the chronological metamorphosis of the iliac auricular surface, and the age estimation method by Passalacqua [59] based on metamorphic and degenerative changes in the sacrum.

To incorporate age-related features of the sacro-iliac joint, a two-component-based system was developed to assess textural and marginal changes in the sacral and iliac auricular surface. The iliac and sacral auricular surfaces undergo textural changes that are characterized by the transition from a smooth, finely grained surface to a granular, irregular and porotic surface. The margins that delimit the surface tend to manifest osteophytic activity as age progresses. Both the texture and margin features refer to the entire structure but very often the degenerative changes, in particular the margin, are more pronounced in specific areas such as the inferior and anterior apices. Full features descriptions are given in Tables S13 and S14.

### 2.2.8. Acetabulum Scoring

Several age-related changes can be documented in the acetabulum and used for age estimation [87–94]. One key aspect of the acetabulum is the late onset of the age-related changes and its durability and resistance to taphonomic factors. To incorporate this skeletal element in our protocol, a three-stage scoring system for the changes occurring on the

rim, posterior horn, and acetabular fossa was developed. In the spirit of Calce [90], who simplified the method developed by Rissech et al. [91,92], the foundation of the scoring system presented in Table S15 is based on a simplification and adaptation of the method proposed by San-Millán et al. [87,95].

### 2.2.9. Scoring Reliability: Intra-Observer Error

To assess the reproducibility of this new proposed scoring system, 50 individuals were randomly selected and rescored on all possible traits ( $m = 99$ ) by the first author. For bilateral traits, only the left side was used for further intra-observer reliability analysis (first author) to avoid issues that arise from non-independent ratings. Kendall's  $W$  [96] was computed as a concordance coefficient to assess consistency between scoring sessions. This metric ranges from 0 (no agreement) to 1 (perfect agreement).

### 2.3. Feature Analysis Via Sphering and Marginal Correlation Analysis

To assess the relationship of the analyzed traits with age-at-death, we inspected marginal correlation coefficients using Spearman's correlation coefficient ( $\rho$ ) and Pearson's eta coefficient ( $\eta^2$ ). In addition to these two coefficients, we also computed marginal correlations adjusted for inter-trait correlation following Zuber and Strimmer [97]. This technique aims to cope with the myopy of univariate feature selection methods by computing marginal correlations of decorrelated predictors with the target class. First, the data centered and scaled, and then transformed by applying a linear basis that enforces orthogonality among predictors while maintaining the maximum relationship with the original standardized predictors. After this transformation, also known as the Mahalanobis transform or sphering, the predictors covariance matrix is the identity matrix (no correlation). The authors called the adjusted marginal correlations CAR scores and proved that ranking based on these quantities provides a fast and optimal procedure for feature ranking and selection. We suggest [97,98] as primers on feature selection and data sphering based on this approach.

### 2.4. Randomized Neural Networks: Theory and Implementation

From a computational perspective, age-at-death estimation can be viewed as a function approximation problem,  $y = f^*(x)$ , and constitutes one of the core reasons why artificial neural networks were chosen as the predictive technique in this work. In age-at-death estimation,  $y = f^*(x)$  maps the input skeletal traits ( $x$ ) to an age-at-death ( $y$ ). ANNs are function approximation machines that define the mapping  $y = f(x; \theta)$ , where  $\theta$  are the parameters or network weights that result in the best approximation [99].

Artificial neural networks are a class of connectionist, biologically inspired computational models that enable learning from data for a multitude of tasks, such as classification, regression, representation learning, and data compression and generation. ANNs are, in a broad sense the result of two components: architectural design—that is how many layers and neurons comprise the network; and an optimization algorithm—how the parameters of the network are learnt.

In its basic implementation, an ANN is composed of three layers: an input layer, a hidden layer, and an output layer. Two sets of weights are embedded in the network structure: one connecting the inputs to the hidden layer and the other connecting the hidden layer to the output layer. In a neural network, the input is transferred to the hidden layer by means of a non-linear activation function. An activation function and the set of weights define a node of the hidden layer. Such nodes are also known as artificial neurons. An artificial neuron, the key component of an ANN, is a mathematical operator in the form of:

$$h(x) = g\left(\sum_{i=1}^p x_i \omega_i + b\right) \quad (1)$$

where  $g()$  is an activation or transfer function,  $x_i$  and  $\omega_i$  are the  $i$ -th components of the input, and the weight vector  $b$  is the neuron bias. Artificial neurons are, in essence, non-



linear functions with learnable parameters, which ultimately expand the ANN model representational capacity to be able to approximate any output function.

A key aspect of ANN is their flexibility and modularity, which due to their capability can be applied to a vast array of heterogeneous data types and domains. The explosion in the availability and capacity to store and analyze data in the form of images, video, audio, and unstructured text has led to the development of novel ANN training algorithms and architectures, and a transition from shallow (single hidden layer) to deep (multi-layer) networks. It is important to note that not all ANNs are formulated and trained in the same manner. There are specialized architectures to tackle; for instance, data in the form of images that make use of computational operations, such as convolutions and pooling. However, a transversal aspect of modern ANNs is their use of gradient-based learning algorithms, where the parameters of a network are iteratively fine-tuned. Gradient-based learning enables end-to-end training and state-of-the-art performance in many complex tasks, but it is costly and requires considerable amounts of technical knowledge to leverage an ANN to its full potential.

A counterintuitive, yet highly efficient, approach to the training of ANN models is to randomly assign and fix a subset of parameters (i.e., hidden weights) of the network and recast the optimization component to a simpler least squares estimation problem [100,101]. In the context of ANNs, randomization as an intrinsic mechanism of model learning can be traced back to late 1980s and early 1990s, with the proposal of randomized radial basis functions network (RBF) and the random vector functional link network (RVFL) models [102–106]. However, the recent interest in randomized algorithms for training feed-forward neural networks can be attributed to the re-emergence of this approach in the guise of the controversial extreme learning machine (ELM) algorithm [107–110]. According to [111], there is no need to rename this strategy for training neural networks, since all key elements have been previously proposed [102–106], and some of the minor changes introduced by the ELM algorithm, such as the omission of direct links between the input and output layer—present in the RVFL network—can have a deleterious effect in performance. Nonetheless, the ELM algorithm acted as a foundation for many innovations in the field of randomized artificial neural networks (RANNs), such as the development of highly efficient algorithms to compute and cross-validate the output layer analytically [112,113], and its evolution from a framework restricted to shallow networks to a set of techniques and algorithms capable of deep, multi-layered network architectures [114–118].

#### 2.4.1. Efficient Training and Regularization in Randomized Neural Networks

In randomized neural networks, the elements of  $\omega_i$ , the hidden layer weights, are randomly generated from a suitable probability distribution and are not optimized. Only the output weights are learned from data by solving a least squares estimation (LSE) problem expressed as:

$$\beta = H^\dagger Y \quad (2)$$

where  $\beta$  are the output layer weights,  $H^\dagger$  is the Moore–Penrose pseudo-inverse of the matrix  $H$ , which defines the hidden layer, and  $Y$  is a column vector storing the network target output, in our case, age-at-death.  $H^\dagger$  can be computed using several methods; a common approach is through orthogonal projection using Equation (3):

$$H^\dagger = (H^T H)^{-1} H^T \quad (3)$$

From Equations (2) and (3), it is trivial to show that the use of this algorithm yields an age estimate as  $\hat{Y} = H\beta$ , and that the output layer is in fact an ordinary least squares linear regression built on the non-linear feature mapping induced by the hidden layer of the neural network.

It has been noted [119] that one can keep the algorithmic simplicity of the least squares solution, while improving its performance and generalization capability by adding a penalty to the output weights. Such a penalty,  $C$ , stabilizes the inversion of matrix  $H$  and

shrinks the coefficients of the output layer towards zero; smaller coefficients lead to smaller error rates on unseen data. Imposing such a constraint on the output weights is a process known as shrinkage or regularization, which in the neural network literature is also named weight decay. This type of regularization is also referred as L2-norm regularization or Tikhonov regularization.

The solution of a regularized RANN is obtained by fitting a ridge regression model [120] as the output layer. The ridge solution,  $\beta_{\text{ridge}}$ , is obtained by substituting Equation (3) as follows:

$$H^\dagger = (H^T H + \frac{I}{C})^{-1} H^T \tag{4}$$

$I$  refers to the identity matrix with dimensions matching  $H^T H$ . Regularization is of paramount importance when training a randomized neural network for age estimation. The solution of the network is obtained by minimizing the squared error as the objective function. LSE-based neural networks lead to unbiased solutions but with high variance if not properly regularized due to the randomness of the initialization [112]. Regularization shrinks the size of the output coefficients towards zero, which is consistent with the theory that smaller weights result in better generalization of neural networks [121,122].

Since the output layer in a RANN is solved as a least squares estimation problem, fortunately, there exist highly efficient, analytical, and closed formulations to assess the leave-one-out (LOO) error, as shown by Shao and Er [112] using Allen’s [123] Prediction Sum of Squares (PRESS) statistic:

$$E_{\text{LOO}} = \frac{1}{n} \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{1 - h_{ii}} \right)^2 \tag{5}$$

where  $h_{ii}$  is the  $i$ -th diagonal element of the hat or projection matrix, which is the matrix that maps the hidden layer parameters to the predicted values of the network, in our case age-at-death. Shao and Er [112] have demonstrated that computing the projection matrix of the network and finding the optimal regularization parameter,  $C$ , under leave-one-out cross-validation (LOO-CV), can be achieved with computational efficiency by performing a singular value decomposition (SVD) of the hidden layer, which, given such an operation, is written as  $H = U \Sigma V^T$ . Using SVD, the network estimate can be written as:

$$\begin{aligned} \hat{Y} &= H \beta \\ \hat{Y} &= H (H^T H + \frac{I}{C})^{-1} H^T Y \\ \hat{Y} &= U (\Sigma^T \Sigma + \frac{I}{C})^{-1} \Sigma^T U^T Y \end{aligned} \tag{6}$$

where  $U (\Sigma^T \Sigma + \frac{I}{C})^{-1} \Sigma^T U^T$  is the projection matrix and it can be noted that only  $(\Sigma^T \Sigma + \frac{I}{C})^{-1} \Sigma^T$  affects the projection matrix for different values of  $C$ .  $\Sigma$  is a diagonal matrix whose element are expressed as  $\phi_i = \frac{\sigma_{ii}^2}{\sigma_{ii}^2 + \frac{1}{C}}$ , where  $\sigma_{ii}$  is the  $i$ -th singular value from the decomposition of  $H$ . SVD makes the regularization of the neural network highly efficient because the diagonal of the projection matrix, which is needed to calculate the LOO error using Equation (6), can be obtained from the following Hadamard products (matrix element-wise multiplication):

$$\gamma = U \circ \Gamma^T = U \circ (\Theta \circ U^T) \tag{7}$$

where  $\Theta = \left(\Sigma^T \Sigma + \frac{I}{c}\right)^{-1} \Sigma^T$ . The diagonal elements of the projection matrix,  $h_{ii}$ , can be obtained by performing a column-wise sum of the elements of  $\gamma$ . The LOO predictions of the network can be obtained analytically as follows:

$$\hat{y}_i = \frac{y_i - f(x_i)}{1 - hat_{ii}} \quad (8)$$

In addition to this highly efficient computational strategy to train a randomized neural network, data standardization and the addition of Gaussian noise to several of the components of the network can also improve performance and accuracy.

#### 2.4.2. From Shallow to Deep Randomized Neural Networks

The mathematical and network formulation presented above pertain to a randomized weights single layer network architecture. Navega and Cunha [124] introduced this model in skeletal age estimation in the formulation of the ELM network (no direct links in the network) and applied it to several traits of the sacroiliac joint. However, several authors proposed different techniques to extend the RANN to deeper architectures [114–118]. To increase the deepness of the network, one can resort to fully randomized approaches or use autoencoding strategies and stack multiple autoencoding RANNs to build a multi-layer network. In this work, due to its simplicity, we follow the proposal of Shi et al. [118] to train deep randomized network models (DRNNs). Following the authors, the first layer of the network is defined as:

$$H^{(1)} = g(XW^{(1)}) \quad (9)$$

where  $X$  is the input matrix, in our case skeletal traits. Every subsequent layer ( $j > 1$ ) is defined as:

$$H^{(j)} = g(H^{(j-1)}W^{(j)}) \quad (10)$$

where  $H^{(j-1)}$  is the previous layer. One can also allow connections from the input to all hidden layers and define the hidden layer as:

$$H^{(j)} = g([H^{(j-1)} X]W^{(j)}) \quad (11)$$

where  $W^1$  and  $W^j$  are the weight matrices between the input-first hidden layer and the inter-hidden layers, respectively. These matrices are randomly assigned and held fixed during the training. The input to output layer is then defined as:

$$D = [H^{(1)} H^{(2)} \dots H^{(j-1)} H^{(j)} X] \quad (12)$$

The design of the deep network is very similar to that of a shallow RANN, and it can be easily seen that the input to output layer consists of non-linear features induced by the hidden layers concatenated to the original input of the network. When the input is reused directly in the output layer, the network is classified as a network with direct link or skip layers. As mentioned above, this is the key difference between ELM and RFVL networks.

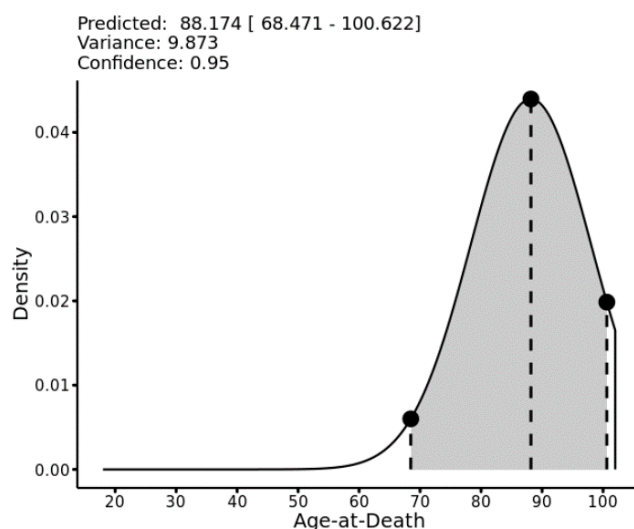
#### 2.4.3. Deep Random Neural Networks as Implicit Ensemble Models

One key advantage of the randomized approach used in this study is that it can enable implicit neural ensemble models [118]. Rather than applying Equation (2) once to solve the output layer weights (solution), Equation (2) can be re-used along the depth of the network for each  $H^{(j)}$  computed from Equations (9) or (10), and obtain an intermediate age-at-death estimate. The final age-at-death estimate can be then obtained by averaging all estimates along the network depth. This feature stabilizes the predictions and offers a different mechanism to train an ensemble model other than training each model independently.

### 2.5. Regression Uncertainty Modeling and Prediction Intervals

The approach followed in this work relies heavily on regression. In Sections 2.4.1 and 2.4.2, we presented the foundation for mathematical age-at-death prediction using RANN models as a regression task. However, we focused only on how point estimates can be obtained, that is, the conditional expectation of age-at-death given a specific skeletal pattern of an individual. Mapping the uncertainty of the point estimate is essential in forensic anthropology, which means that a predictive interval for a preset confidence level should also be part of the analysis and the subsequent report.

In the current work, we follow a simple and generic approach based on modeling the conditional variance associated with each point estimate (network prediction). We recast the prediction interval construction as a regression problem and, using LOO network predictions, we build a regression uncertainty model (RUM) by regressing absolute residuals on predicted age-at-death. We then scale the predicted residual by 1.2533 to obtain a standard deviation associated with each age estimate. The scaling factor is the ratio of the standard deviation to the absolute deviation [125,126]. Assuming normality of the variance around each point estimate, the prediction interval associated with an ANN model is given by the quantiles of a Gaussian or truncated Gaussian parameterized with the conditional mean and standard deviation inferred from the ANN and its associated RUM. The key advantage of this approach is its simplicity compared to likelihood methods [15–17,20,23,127–129] or conformal prediction theory, as in [113,124,130]. In addition to the numerical interval, this approach also allows visualization, as illustrated by Figure 2.



**Figure 2.** Prediction interval visualization using a (truncated) Gaussian uncertainty model.

### 2.6. Computational Analysis: Design, Parameterization, Metrics, and Software

#### 2.6.1. Experimental Design

To assess the performance of DRNN and Gaussian RUM models in multifactorial age estimation from macroscopic skeletal traits we followed a simple template for robust metric assessment based on a resampling Monte Carlo cross-validation (MCCV) scheme. This works as follows: for a given iteration of the scheme, split the dataset into disjoint train and test partitions. Using the training partition, fit a DRNN and RUM models by making use of Equations (5)–(7) to optimize the regularization parameter  $C$  and obtain leave-one-out predictions.  $C$  is optimized as  $2^x$  with  $x \in \{-6, -4, \dots, 12\}$ . With the trained DRNN and RUM models, we predict the age-at-death of the testing sample/partition and compute the MCCV performance metrics. For a given set of skeletal traits, this procedure is repeated 1000 times ( $B = 1000$ ). The training partition is set as 80% of the total data (400 of 500) and the test partition as the remaining (100 of 500). This sampling procedure

was performed without replacement. The core of our computational analysis is organized in two experiments, from now on referred to as experiments A and B:

- (A) The first experiment we conducted was designed to provide a baseline of the accuracy obtained by fitting DRNN models to blocks of traits that have standard or traditional analytical framing. For instance, we fitted models to different anatomical complexes or sets of traits that mimic existing aging standards, i.e., a model for the sutures or the pubis symphysis.
- (B) Our second computational experiment consisted of simulated different proportions of available traits from 90% to 10%. The objective of this experiment was to assess model performance in a more realistic scenario where the forensic anthropologist has skeletal traits available on a case-by-case basis.

In both experiments we computed 95% predictive intervals (95% PI) by setting the uncertainty of parameter  $\sigma = 0.05$ .

### 2.6.2. Network Parameterization

A key aspect of any ANN model is its architecture, that is, how many neurons (or nodes) and layers comprise the network. To leverage the full potential of the DRNN, and to maximize its training speed and efficiency, rather than search for the optimal architecture, we developed a simple heuristic based on the work of Lappas [131]. The author demonstrated that the size of a single layer perceptron can be estimated from the number of samples available. Using his work as a foundation, we propose the following heuristics for setting the architecture of a DRNN. The width, size, or number of neurons of each layer was set as:

$$S = 2^{\lceil \log_2(8(\sqrt{2^k}/k)) \rceil}, k = \log_2(n) \quad (13)$$

where  $n$  is the number of samples. The depth or number of layers was set as:

$$L = 2^{\lceil \log_2(k) \rceil}, k = \log_2(n) \quad (14)$$

Following Equations (13) and (14) as a simple heuristic allows us to have predictable, parsimonious network architectures. In this way, the network allows many computing units for randomized feature extraction distributed over several layers without incurring overparameterization. This heuristic also leverages the simplicity of training a deep neural network using the same mechanisms of a shallow one, while exploiting an implicit ensemble framework (Section 2.4.3). For our experiments, applying the described heuristic defines the network architecture with a rectangular topology comprising eight layers of 32 neurons each, for a total of 256 randomized units.

DRNNs are computationally cheap nonlinear models built by combining regularized linear regression with nonlinear features obtained by using an activation function,  $g(\cdot)$ , with random weights. In this work, we used the rectified linear unit (ReLU) as the nonlinearity of the networks. The ReLU is defined as  $g(z, w) = \max(0, zw)$ , where  $z$  and  $w$  are the layer input and random weight matrices. Since the regularization process involved in the training process described in this work is not scale invariant, during network training normalization by mean centering and variance scaling, Equation (6) was performed on the matrices  $X$ ,  $XW$ ,  $H$ , and  $Y$ . The output of the network was later rescaled before computation of the performance metrics.

ANN architecture selection and design is a non-trivial task often performed through very expensive and complex computational strategies and procedures. The heuristic used and architecture selected in this work emerged from trial-and-error experimentation during the development of the *rwnnet* software package (see Section 2.6.4). This parameterization leverages the benefits and key features of randomized neural networks—fast training and prediction with minimum technical knowledge, given that the model is fully described through linear algebra and matrix operations.

### 2.6.3. Performance Metrics

In our analysis, we evaluate four parameters that any model used in regression task should have, especially one used for age estimation. An age-at-death prediction model—regardless of its underlying mathematical algorithm—should be accurate, unbiased, valid, and efficient. Accuracy refers to the ability of the model of the model to predict age with minimal error. The most straightforward metric to assess this parameter is the mean absolute error (MAE) computed as:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (15)$$

where  $y_i$  and  $\hat{y}_i$  are the known and predicted values, respectively, and  $n$  is the number of evaluated samples.

A model should be unbiased, that is, free of systematic error. A typical pattern of bias or systematic error in age estimation models is the over-estimation of young individuals and under-estimation of the elderly. A robust and comprehensive way to assess bias ( $\hat{\beta}_e$ ) is by computing the slope of the regression line of the residuals,  $e_i = y_i - \hat{y}_i$ , on known values. When minimal to no bias is presented, this value should be close to zero. A positive slope suggests a systematic bias, such as the one describe previously. Bias is computed as:

$$\hat{\beta}_e = \frac{\sum (y_i - \bar{y})(e_i - \bar{e})}{\sum (y_i - \bar{y})^2} \quad (16)$$

where  $\bar{y}$  and  $\bar{e}$  are the means of the known and residual values.

The validity of model, in the context of our study, refers to the ability of a model to contain the known age within the predictive interval and within a reasonable margin close to the nominal uncertainty level allowed. For instance, for an uncertainty level (alpha) of 0.05 (or 5%) we expect that the coverage of the correct proportion of individuals within the predictive interval is close to 0.95 (or 95%). As a validity measure, we compute:

$$P(\alpha) = \frac{\sum_{i=1}^n \delta(y_i, l_i, u_i)}{n} \quad (17)$$

where  $\delta(y_i, l_i, u_i)$  is an indicator function with  $\delta(y_i, l_i, u_i) = 1$ , if  $y_i \geq l_i \wedge y_i \leq u_i$  and  $\delta(y_i, l_i, u_i) = 0$ , and  $l_i$  and  $u_i$  are the values of the lower and upper ends of the predictive interval, respectively.

Finally, a model should thrive to be efficient. Efficiency in this context refers to the width or range of the prediction intervals associated with the regression uncertainty model. A method or model is efficient when it outputs the narrowest predictive interval possible while also maintaining its validity. We compute our measure of efficiency as follows:

$$PIW = Q(u - l, \tau), \text{ with } \tau \in \{0.5; 0.025; 0.975\} \quad (18)$$

where  $Q(\cdot)$  is a quantile function and  $\tau$  a given quantile. One can see that we compute the median of the predictive interval width and its associated 95% confidence interval (quantile-base).

### 2.6.4. Software

All computational work was performed using the R and C++ programming languages with all key software components written by the first author. To perform this work, the `rwnnet`, `rumr`, `rmar`, and `lsmr` packages were used. These packages are available from the respective repositories of the GitHub profile of the first author, <https://github.com/dsnavega> (accessed on 18 March 2022).

Novel software, DRNNAGE, that operationalizes age-at-death estimation following the macroscopic and computational techniques described in this work, was also devel-

oped and is live as a web application at <https://osteomics.com/DRNNAGE> (accessed on 18 March 2022.); its source is available at <https://github.com/dsnavega/DRNNAGE> (accessed on 18 March 2022.). In its current state, we strongly recommend that end users approach their analysis using only default parameters. All problems detected and suggestions should be directed to the corresponding author.

### 3. Results

#### 3.1. Intra-Observer Scoring Error

Overall, the new proposed macroscopic scoring technique presented high intra-observer consistency based on the results on Kendall's  $W$  concordance coefficient [96]. With the exception of RD01 and FM01, 0.751 and 0.716, respectively, all skeletal traits presented a concordance coefficient higher than 0.800. The global average of this coefficient was 0.907. All traits presented a statistically significant concordance between scoring obtained by the first author in two different sessions. The high concordance observed can be explained by the simplicity of the scoring systems used with the large number of traits that were binary coded. Further inter- and intra-observer error analysis is required by an independent third party, due to the nature of the methods employed.

#### 3.2. Marginal Correlation Analysis

Marginal correlation analysis showed that all traits have a statistically significant relationship with age-at-death. The cranial sutures showed the lowest marginal correlation ( $\rho$ : 0.297–0.519,  $\eta^2$ : 0.088–0.249), with palatine sutures explaining less than 10% of the variation in observed age-at-death. The axial traits—cervical and lumbar vertebrae—exhibited a moderate to strong monotonic relationship and explained variation with age-at-death ( $\rho$ : 0.794–0.845,  $\eta^2$ : 0.639–0.725). A similar correlation and explained variation pattern were observed for the clavicle traits ( $\rho$ : 0.710–0.851,  $\eta^2$ : 0.507–0.729), first rib traits ( $\rho$ : 0.763–0.776,  $\eta^2$ : 0.590–0.607), iliac auricular surface traits ( $\rho$ : 0.731–0.789,  $\eta^2$ : 0.539–0.631), and the acetabular traits ( $\rho$ : 0.782–0.818,  $\eta^2$ : 0.625–0.674). A slightly lower marginal correlation was observed for the pubic symphysis traits ( $\rho$ : 0.711–0.731,  $\eta^2$ : 0.523–0.549) and sacral auricular surface traits ( $\rho$ : 0.632–0.704,  $\eta^2$ : 0.398–0.499). Traits from the upper and lower limbs presented a wider range of correlation ( $\rho$ : 0.380–0.789,  $\eta^2$ : 0.145–0.628). When analyzed in the context of feature ranking based on marginal correlations adjusted for inter-trait correlation (CAR scores), the suture traits score was among the worst predictors and its decorrelated components showed no statistically significant relationship with age-at-death. The several appendicular degenerative traits—HM04, UL01, RD01, FM01, FM02, and TB01—also showed no statistically significant correlation when assessed on a Mahalanobis transformed space. Ranking based on CAR scores showed that the top-ranking traits came from all anatomical regions rather than a specific indicator.

#### 3.3. Computational Model Assessment

Results from the two *in silico* experiments performed to assess DRNN models in age-at-death estimation are reported in Tables 2–5. Models based solely on the cranial sutures exhibited the worst performance among all models produced, having a median MAE of 15.300 (Table 2) and a median predictive interval width (PIW) of 68.144 years, which renders the cranial sutures an inaccurate and inefficient set of traits.

**Table 2.** Monte Carlo cross-validation metrics for DRNN models built on pre-specified skeletal traits sets.

		Accuracy	Bias	Validity	Efficiency		
Traits		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95% CI	
Sutures (m = 9)	Median	15.300	0.656	0.950	68.144	51.699	69.759
	95% CI	13.586	0.590	0.900	66.054	46.361	68.312
Axial (m = 16)	Median	17.206	0.732	0.990	69.741	55.776	70.963
	95% CI	8.185	0.198	0.960	38.754	33.732	40.842
Appendicular (m = 23)	Median	7.365	0.137	0.920	37.102	32.272	39.215
	95% CI	9.139	0.260	0.990	40.091	35.029	42.191
Clavicle (m = 2)	Median	7.583	0.167	0.960	37.378	29.109	39.541
	95% CI	6.678	0.103	0.910	35.412	27.613	38.014
First Rib (m = 2)	Median	8.523	0.231	0.990	39.079	30.399	41.061
	95% CI	8.949	0.244	0.960	49.234	17.354	51.610
Pubic symphysis (m = 3)	Median	7.798	0.169	0.920	39.064	15.981	49.962
	95% CI	10.192	0.307	0.990	52.688	18.617	53.098
Sacroiliac complex (m = 6)	Median	9.500	0.277	0.950	48.936	24.334	49.637
	95% CI	8.138	0.204	0.900	46.879	22.499	47.687
Acetabulum (m = 3)	Median	10.831	0.351	0.990	50.903	26.078	51.533
	95% CI	10.897	0.370	0.940	51.210	26.905	56.954
Degenerative traits (m = 39)	Median	9.371	0.280	0.870	48.688	24.520	54.799
	95% CI	12.542	0.459	0.980	55.558	29.058	58.802
Standard traits (m = 16)	Median	8.523	0.223	0.950	44.668	20.378	47.969
	95% CI	7.380	0.145	0.890	39.350	18.596	46.017
All (m = 64)	Median	9.742	0.288	0.990	47.547	21.915	49.720
	95% CI	8.886	0.229	0.970	42.978	31.727	45.742
All (m = 64)	Median	7.758	0.162	0.920	41.201	29.897	43.891
	95% CI	10.006	0.287	1.000	44.509	33.240	47.304
All (m = 64)	Median	6.962	0.147	0.970	33.732	28.882	35.122
	95% CI	6.084	0.085	0.920	32.460	27.570	33.488
All (m = 64)	Median	7.814	0.200	1.000	34.935	30.019	36.656
	95% CI	6.609	0.147	0.950	34.245	12.927	41.087
All (m = 64)	Median	5.561	0.087	0.890	29.701	11.833	39.097
	95% CI	7.598	0.202	0.990	37.857	14.169	42.833
All (m = 64)	Median	5.925	0.117	0.950	30.010	15.631	36.081
	95% CI	5.101	0.060	0.900	26.817	14.464	34.612
All (m = 64)	Median	6.728	0.170	0.990	33.191	16.811	37.515
	95% CI						

**Table 3.** Leave-one-out cross-validation metrics for DRNN models built on pre-specified skeletal traits sets.

		Accuracy	Bias	Validity	Efficiency		
Traits		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95% CI	
Sutures (m = 9)	Median	15.245	0.655	0.953	68.120	51.782	69.796
	95% CI	14.683	0.616	0.940	66.377	46.429	68.371
Axial (m = 16)	Median	15.751	0.692	0.963	69.708	55.878	70.996
	95% CI	8.156	0.200	0.960	38.825	33.594	40.881
Appendicular (m = 23)	Median	7.896	0.184	0.953	37.468	32.131	39.279
	95% CI	8.394	0.213	0.968	39.872	34.902	42.234
Clavicle (m = 2)	Median	7.557	0.169	0.960	37.534	29.035	39.599
	95% CI	7.278	0.155	0.948	35.996	27.542	38.082
First Rib (m = 2)	Median	7.823	0.184	0.970	38.920	30.319	41.109
	95% CI	8.943	0.245	0.963	49.216	17.336	51.768
Pubic symphysis (m = 3)	Median	8.606	0.228	0.953	47.184	15.969	50.112
	95% CI	9.248	0.263	0.970	51.238	18.597	53.252



**Table 3.** Cont.

		Accuracy	Bias	Validity	Efficiency		
Traits		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95% CI	
First Rib (m = 2)	Median	9.409	0.275	0.950	48.897	24.356	49.811
	95% CI	9.067	0.255	0.938	47.036	22.502	47.862
Pubic symphysis (m = 3)	Median	9.751	0.296	0.960	50.829	26.102	51.724
	95% CI	10.898	0.370	0.932	51.113	27.029	57.040
Sacroiliac complex (m = 6)	Median	10.436	0.343	0.922	48.668	24.616	54.949
	95% CI	11.315	0.398	0.945	53.003	29.217	58.909
Acetabulum (m = 3)	Median	8.438	0.220	0.950	44.765	20.350	48.037
	95% CI	8.075	0.200	0.940	42.461	18.607	46.091
Degenerative traits (m = 39)	Median	8.741	0.239	0.960	46.755	21.893	49.800
	95% CI	8.833	0.229	0.965	43.051	31.541	45.832
Standard traits (m = 16)	Median	8.490	0.210	0.955	41.302	29.726	43.995
	95% CI	9.116	0.247	0.975	44.535	33.054	47.395
All (m = 64)	Median	6.929	0.147	0.963	33.744	28.816	35.194
	95% CI	6.694	0.133	0.953	32.530	27.499	33.566
	Median	7.154	0.157	0.973	34.829	29.946	36.715
	95% CI	6.561	0.145	0.948	34.283	12.952	41.170
	Median	6.277	0.132	0.935	32.464	11.853	39.222
	95% CI	6.855	0.157	0.960	36.027	14.122	42.921
	Median	5.899	0.118	0.950	30.057	15.558	36.141
	95% CI	5.677	0.110	0.940	28.758	14.403	34.644
	Median	6.121	0.127	0.963	31.485	16.668	37.620
	95% CI						

**Table 4.** Monte Carlo cross-validation metrics for DRNN models built on different fractions of available skeletal traits.

		Accuracy	Bias	Validity	Efficiency		
Available Traits %		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95% CI	
90% (m ≈ 57)	Median	5.964	0.120	0.950	30.354	15.851	36.215
	95% CI	5.136	0.062	0.900	27.067	14.466	34.554
80% (m ≈ 51)	Median	6.773	0.169	0.990	33.422	18.081	37.705
	95% CI	6.026	0.121	0.950	30.498	16.004	36.261
70% (m ≈ 44)	Median	5.211	0.061	0.900	27.183	14.213	34.498
	95% CI	6.851	0.172	0.990	33.584	18.492	37.902
60% (m ≈ 38)	Median	6.072	0.125	0.950	30.805	16.206	36.454
	95% CI	5.152	0.062	0.900	27.528	14.001	34.600
50% (m ≈ 32)	Median	6.924	0.180	0.990	34.004	19.666	38.405
	95% CI	6.131	0.125	0.950	30.964	16.352	36.649
40% (m ≈ 25)	Median	5.316	0.065	0.900	27.513	13.893	34.672
	95% CI	7.049	0.179	0.990	34.320	20.532	38.692
30% (m ≈ 19)	Median	6.237	0.129	0.950	31.479	16.717	36.969
	95% CI	5.293	0.064	0.900	27.820	13.757	34.930
20% (m ≈ 12)	Median	7.180	0.179	0.990	34.854	22.119	39.250
	95% CI	6.360	0.134	0.950	32.125	17.165	37.429
10% (m ≈ 6)	Median	5.441	0.074	0.900	28.500	13.910	35.075
	95% CI	7.380	0.193	0.990	35.636	23.292	40.166
	Median	6.570	0.140	0.950	33.163	17.933	38.137
	95% CI	5.565	0.075	0.900	29.036	13.905	35.393
	Median	7.651	0.201	0.990	36.916	25.407	40.861
	95% CI	6.951	0.153	0.950	35.263	19.946	39.694
	Median	5.857	0.086	0.900	31.082	14.074	36.427
	95% CI	8.139	0.218	0.990	39.625	28.892	43.619
	Median	8.026	0.196	0.950	39.618	26.914	43.025
	95% CI	6.592	0.119	0.900	34.681	15.495	38.368
	Median	9.683	0.276	0.990	46.043	34.276	49.479
	95% CI						

**Table 5.** Leave-one-out cross-validation metrics for DRNN models built on different fractions of available skeletal traits.

Available Traits %		Accuracy	Bias	Validity	Efficiency		
		MAE	$\hat{\beta}_e$	$P(\alpha)$	PIW	PIW 95% CI	
90% (m ≈ 57)	Median	5.942	0.121	0.953	30.276	15.745	36.278
	95% CI	5.699	0.110	0.940	28.748	14.339	34.599
80% (m ≈ 51)	Median	6.198	0.131	0.965	31.797	18.048	37.772
	95% CI	5.970	0.122	0.953	30.476	15.941	36.332
70% (m ≈ 44)	Median	5.702	0.108	0.940	28.860	14.162	34.574
	95% CI	6.235	0.132	0.965	31.963	18.470	37.938
60% (m ≈ 38)	Median	6.028	0.124	0.953	30.711	16.182	36.518
	95% CI	5.737	0.108	0.938	28.960	14.013	34.697
50% (m ≈ 32)	Median	6.376	0.137	0.965	32.583	19.643	38.435
	95% CI	6.078	0.125	0.953	30.975	16.342	36.716
40% (m ≈ 25)	Median	5.768	0.108	0.938	29.070	13.872	34.756
	95% CI	6.441	0.140	0.965	33.017	20.569	38.732
30% (m ≈ 19)	Median	6.173	0.128	0.953	31.502	16.684	37.040
	95% CI	5.819	0.111	0.938	29.410	13.724	34.989
20% (m ≈ 12)	Median	6.648	0.146	0.968	33.900	22.110	39.305
	95% CI	6.305	0.132	0.953	32.146	17.153	37.511
10% (m ≈ 6)	Median	5.903	0.114	0.935	29.839	13.905	35.130
	95% CI	6.797	0.153	0.968	34.565	23.287	40.214
	Median	6.501	0.138	0.953	33.097	17.923	38.203
	95% CI	6.046	0.118	0.935	30.583	13.899	35.468
	Median	7.096	0.163	0.965	35.986	25.377	40.943
	95% CI	6.957	0.154	0.953	35.321	19.986	39.742
	Median	6.316	0.127	0.935	32.096	14.117	36.479
	95% CI	7.674	0.184	0.968	38.931	28.768	43.707
	Median	7.952	0.192	0.955	39.733	26.846	43.076
	95% CI	6.968	0.154	0.940	35.229	15.515	38.419
		9.214	0.256	0.973	46.437	34.087	49.551

Modeling based on specific anatomical regions resulted in a DRNN with a median MAE ranging from 7.583 to 10.897 years (Table 2); focusing solely on this metric, it is reasonable to state that, on its own, different anatomical regions perform similarly in age estimation. The same can be said for the metrics of bias, validity, and efficiency. Predictive interval width is perhaps the most distinctive metric for practical applications. Anatomical regions with strong developmental signs, such as the clavicle or the pubis, tend to provide narrower predictive intervals for younger individuals.

Combining traits from different regions provided an improvement over models built on specific anatomic regions. Using 16 traits from standard age-related traits—clavicle, first rib, pubic symphysis, sacroiliac complex (auricular surfaces, S1 body surface, and S1-S2 fusion), resulted in a MAE of 6.609 (5.561–7.598, 95% CI) and reduced the prediction bias considerably when compared to any model built on the same anatomical regions independently (Table 2), and a PIW of 34.245 (12.927–41.087, PIW 95% CI). A model based only on degenerative traits (m = 39) resulted in a MAE of 6.962 (6.084–7.814, 95% CI) and median PIW of 33.732 (28.882–33.122, PIW 95% CI). From our results, multifactorial age estimation models provide improved efficiency, as reflected in narrower predictive intervals (Figures 3–5).

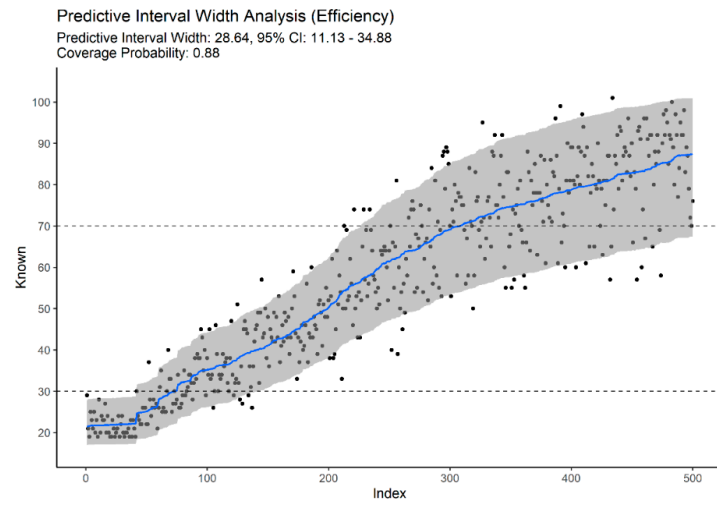


Figure 3. Predictive efficiency of standard age-related traits,  $\alpha = 0.1$ .

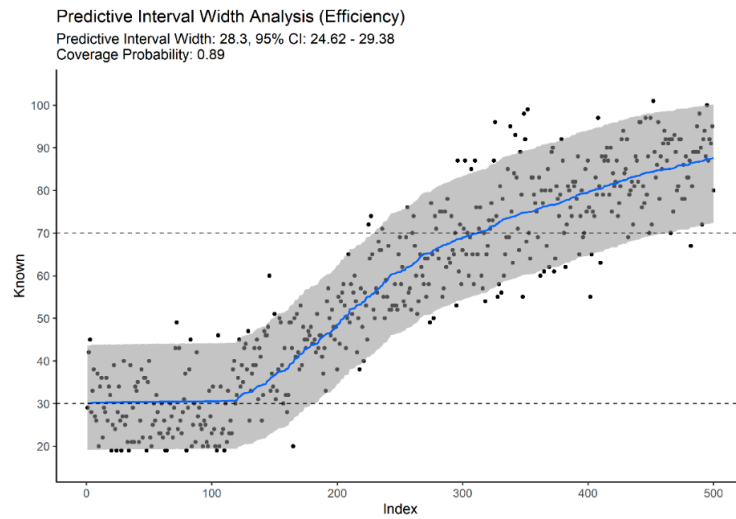


Figure 4. Predictive efficiency of degenerative traits of the axial and appendicular skeleton,  $\alpha = 0.1$ .

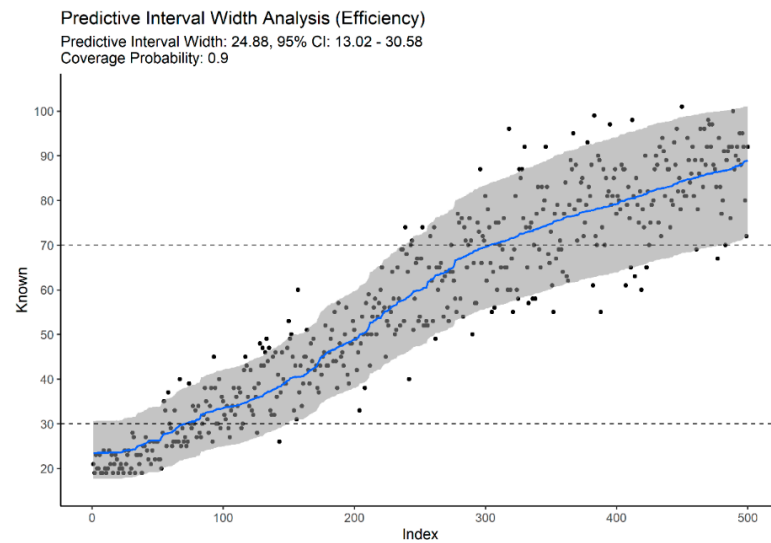


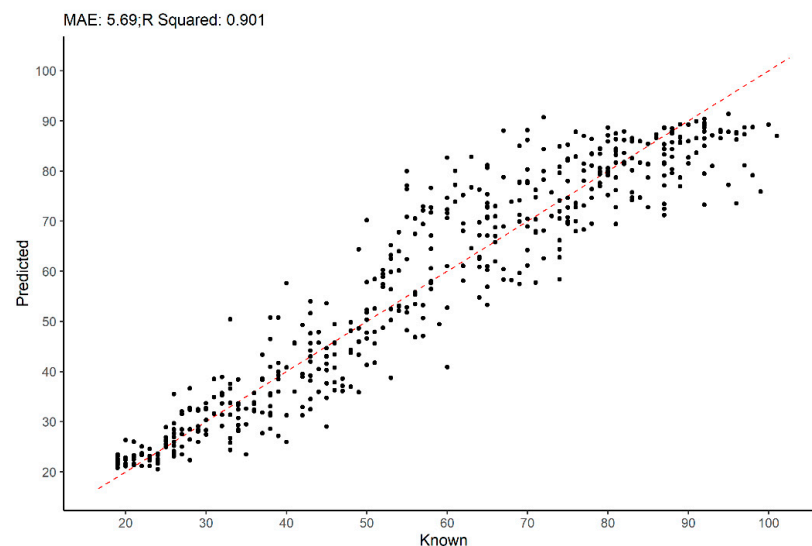
Figure 5. Predictive efficiency of full traits, DRNN-RUM model,  $\alpha = 0.1$ .

From Figures 3–5, we can also observe that multifactorial models provide accurate and efficient estimates across the entire adult lifespan, solving the problem of open-ended and unspecific age-at-death estimates for the elderly. Figure 4 illustrates the importance of non-standard traits to accurately predict advanced age-at-death. Based solely on degenerative traits of the vertebrae, limb joint, and musculoskeletal attachment sites, we can obtain estimates for the elderly that are comparable to more classical traits (Figure 3) or full-set models (Figure 5). The downside of relying solely on this type of indicator for age-at-death estimation is the wider intervals for young adults with no degenerative traits (95% PI ~18 to 46 years vs. ~18 to 32 if traits with sharp developmental stages are present).

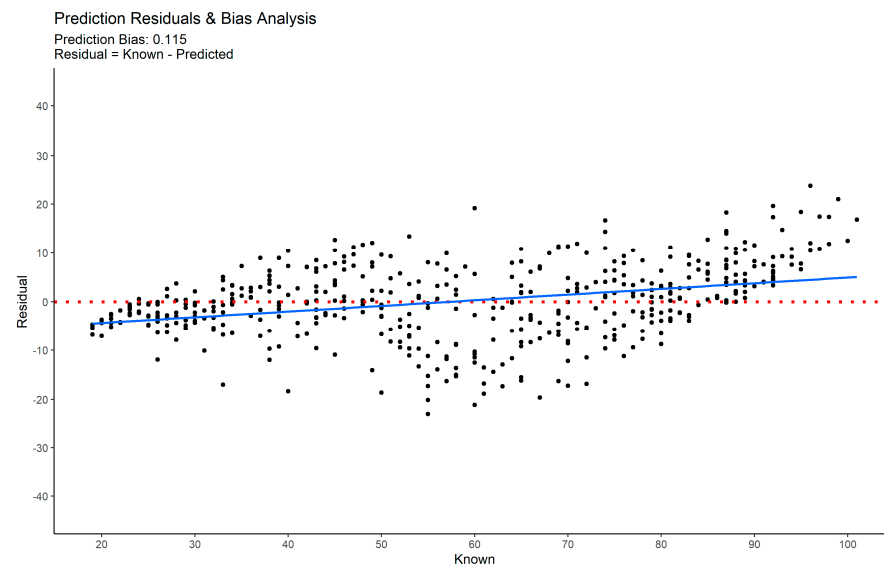
The best performing models in experiment A were those built on the full feature set ( $m = 64$ ), with a mean absolute error of 5.925 (5.110–6.728, 95% CI), and PIW of 30.010 (15.63–36.081, PIW 95% CI) years. The prediction bias for this model was 0.117 (0.060–0.170, 95% CI), which represents a two-to-six-fold reduction in the prediction bias compared to other models built on specific anatomical regions individually (Table 2). Results from experiment B (Tables 4 and 5) showed that similar results can be obtained using different proportions of traits selected at random.

An important remark to make regarding our results based on the two computational experiments is that analytical LOOCV, implicitly performed during model optimization, showed little to no disparity with the results obtained during the repeats of the Monte Carlo cross-validation procedure ( $B = 1000$  repeats) where 20% of the data was used as a proper test set.

The accuracy of our approach can be visualized in Figure 6, where a scatter plot of known vs. predicted age-at-death is depicted. From this figure, one can infer that the predictions obtained using our approach maintain a similar level of error—dispersion around the identity line (dashed red line)—across the entire adult age span, and slightly more accurate for individuals under 40 years. For individuals over 90 years old at death, there is an observable under-estimation. It is also possible to visualize, Figure 7, that a deep RANN model using multiple traits produces minimally biased estimates.



**Figure 6.** Known vs. predicted age-at-death using a full set of traits (LOOCV,  $n = 500$ ).



**Figure 7.** Prediction bias plot for the multifactorial ( $m = 64$ ) RANN model.

Regarding the validity of the models trained in our computational experiments, results show that the predictive intervals contained the known age-at-death without significant deviation from the nominal level of uncertainty (median of  $P(\alpha) \sim 0.95$ , with variation between 0.87 and 0.99). Multifactorial models also show a systematical reduction in prediction bias when compared to models based only on a specific anatomical structure.

#### 4. Discussion

The main objective of this work was to investigate the fundamental issue of age-at-death estimation in the forensic analysis of human remains, and propose a new method and its computational analysis from a perspective of multifactorial analysis of the adult skeleton. Several age estimation methods have been previously developed, focusing on specific anatomical structures or regions such as the cranium, the ribs, or the pelvic joints. Nonetheless, it is well known that no single skeletal indicator is capable of producing accurate and efficient age estimates across the entire human age span. Determining how to report age estimates using multiple indicators or traits remains an open issue, with experts resorting to different heuristics that often are not standardized and lack a valid computational or statistical grounding [5]. In the literature, there are techniques that use multiple skeletal indicators for age estimation but are often limited to the cranial sutures and the pelvic joints [20,23,132]. More generic procedures for multifactorial analysis have also been proposed [133,134], but with poor adoption in forensic casework because they require seriation or advanced mathematical knowledge to be put into action.

The current study provides strong support for multifactorial or multi-trait analysis of the skeleton as a way of obtaining accurate and efficient age estimates across the entire span of adulthood. Results from computational experiment A suggest that using each skeletal indicator or anatomical region separately provides limited improvement over existing methods. One striking remark from this experiment was the performance of the models solely based on the axial (vertebrae) and appendicular (limbs) skeleton. In previous studies, these traits have been considered to be only useful for providing a general estimate or limited in value for age prediction [135,136]; nonetheless, our results are consistent with those of more recent publications that assess their predictive utility and urge reconsideration of these traits as valid age-related traits [64,66]. For instance, if these traits all present a Stage 0, one can infer without any computation that the age-at-death of the deceased is between approximately 18 and 46 years (Figure 4, considering  $\sigma = 0.1$ ). Our results also indicate that the inclusion of these traits is pivotal to solve the problem of open-ended age intervals and poor age estimation for the elderly. On their

own, degenerative axial and appendicular traits allow estimation of the age-at-death of the elderly with an improved accuracy and efficiency compared to more standard traits such as the pelvic joints (i.e., pubic symphysis, acetabulum, iliac auricular surface). The neural model based on the full set of traits described in the novel macroscopic age estimation proposed here provided the best performance results in respect to all metrics analyzed. This can be attributed to the fact that having more features allows the deep neural models to operate at their maximum potential regarding what they do best—extracting novel features from existing ones using, in our case, random weights and a non-linearity (ReLU function) as a mechanism to combine multiple traits, which ultimately allows the output layer to operate in a non-linear regime, despite it being, in practice, a regularized linear model. Moreover, the multitude of traits scored also permits the models to encapsulate the intra- and inter-variability of skeletal morphology with greater finesse, which is manifested as more efficient (narrower) predictive intervals that reflect the heteroskedastic nature associated with the senescence process.

Although the main goal of the computational experiment A was to establish a baseline of performance of multifactorial age-at-death estimation compared to more traditional modeling approaches based on specific anatomical blocks or regions, experiment B aimed to assess the performance of neural models for age-at-death estimation in a more realistic setting, where the expert may not be able to use the pre-specified models or the full set of traits due to the availability of skeletal elements or the multitude of factors that make it impossible to score all traits defined in this macroscopic technique. This computational experiment also provides, both directly and indirectly, answers to several questions that may arise regarding the approach and technique used, and proposed in this work from a more pragmatical and casework view: Does the skeleton need to be complete to reap the maximum benefits of this protocol? Which combination of traits works best or is necessary? How practical is the method?

The results demonstrated that the accuracy of the full-set model ( $m = 64$ ) can be maintained to large degree using smaller random combinations of traits, which ultimately are dictated on a case-by-case basis in a forensic setting. Once again, this can be explained by the capacity of the neural models to extract and combine information from the skeletal traits in an optimal way in terms of prediction. It is important to note here that models based on randomized proportions of traits presented performance metrics superior to most models based on specific anatomical regions, which reinforces our thesis that the multifactorial or multi-trait models are crucial for improving the state-of-art in forensic skeletal age estimation.

Finding an optimal or minimum number of traits is, from a combinatorial and practical point of view, an intractable problem, for which a solution can only be approximated with such a large number of traits ( $m = 64$ ). However, such a solution would be computational wasteful and of little pragmatic value because, as in the situation of the full trait set, the optimal or minimum trait set can result in a non-applicable model due to the availability of skeletal elements during casework. This is the main reason why, in our study, we opted for a randomized evaluation of smaller traits set. Ultimately, we developed the DRNNAGE software to operationalize the age estimation procedure described in this manuscript, in a manner that is flexible and practical for the expert applying it, bearing in mind that each case will be limited by its own available skeletal traits. DRNNAGE allows the expert to compute the optimal network and associated uncertainty model based only on the traits that the forensic expert can score. Thus, in that regard, the usefulness of the estimates obtained is limited by biology and taphonomy itself, rather than the technical implementation.

From a practitioner perspective, marginal correlation analysis and the performance of the developed models clearly suggest that there is room for improvement in our approach regarding the issue of the traits to be used. For instance, our results suggest that there is little to be gained from including the cranial sutures, which, from a predictive modeling standpoint, resulted in the worst model on its own using our scoring protocol. Similar conclusions were reached by Jooste et al. [137], who also investigated the cranial sutures in the

context of a multifactorial approach. To maximize the potential of the framework proposed in this work, it is important to bear in mind that domain and expert knowledge is of utmost importance; this can also be said of any other machine learning or computationally heavy approach. The practical aspect of this method can be improved if applied with the rationale of the well-known Two-Step Procedure proposed by Baccino et al. [138]. This procedure and heuristic for age-at-death estimation suggests age indicators should be combined logically or hierarchically rather than by brute force (i.e., averaging). In the context of our proposal, this translates into the following: if several traits with sharp metamorphic or developmental stages exhibit Stage 0—i.e., clavicle sternal end, S1-S2 fusion, pubic symphysis components—a neural model is trained using those traits and the other traits are ignored. The same rationale can be applied if the traits that encode a strong degenerative signal, such as the vertebrae and limb traits, are scored with their maximum stage (Stages 1 or 2). In this case, we have demonstrated that age estimation can be accurate and efficient when relying solely on these traits. As a final remark and suggestion to improve age estimation with our method, but also with any other method that employs a multifactorial or multi-trait approach, rather than focusing on an optimal or minimal number of traits to use, one should focus on the representational power of the traits analyzed and, whenever possible, use traits that represent both metamorphic and degenerative aspects of the skeletal development and senescence, as argued by Winburn [88].

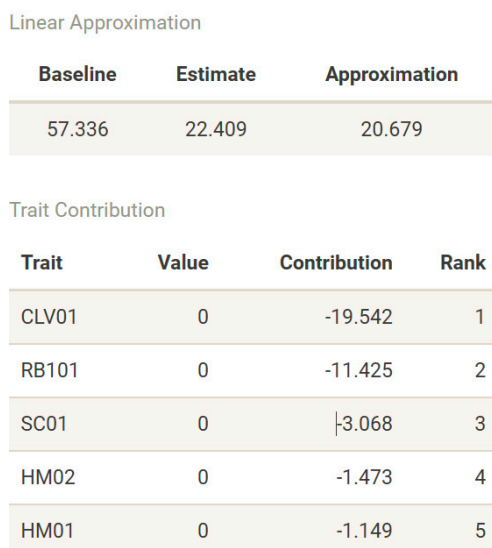
The present work provides a solution to the problem of multifactorial age estimation based on the macroscopic analysis of the skeleton. Multifactorial skeletal age estimation is systematically noted as being the most accurate way to achieve an age estimation in adults, but is obtained through a plethora of procedures and heuristics that are often subjective and lack a clearly well-defined statistical or computational rationale [3,5]. As noted by Ritz-Timme et al. [3], a comparison of different methods with regard to their performance based on published data is an exercise that can only be undertaken with severe limitations and caution. The existing methods have been developed on samples of differing sizes, unbalanced age distributions, and different population backgrounds. There is no standardized array of statistical parameters used to assess an age estimation method, and different statistical procedures have been applied. In many cases, there is a lack of detail regarding the procedures used, and often only an incomplete analysis performance is pursued (i.e., focusing only on MAE and point estimate accuracy). In the context of our research subject, these limitations are exacerbated by the fact that, to the best of our knowledge, no other study in the literature has pursued a systematic analysis of adult skeletal age estimation using such a vast and diverse array of morphoscopic traits based on a single reference dataset. Nonetheless, a brief analysis of the most recent and comprehensive validation studies clearly demonstrates that our multifactorial approach offers improved accuracy (MAE < 8 years) in relation to other skeletal age estimation methods [137,139–141]. Independent validation of the method and software tools proposed here on samples from different temporal and biogeographic origins are of utmost importance to ascertain the broader impact and significance in archaeology, forensic anthropology, and medicine.

Artificial intelligence, statistical, and machine learning approaches are now ubiquitous in forensic and biological sciences. Several cases in the literature illustrate the usefulness of such approaches in adult macroscopic age-at-death estimation [13–15,22,24,124]. Although these approaches usually allow for flexible and non-parametric modeling with improved predictive performance, it also results in more opaque or black-box models from a non-expert perspective. These approaches also require proper validation and model selection techniques to avoid overfitting [142]. In this study, we applied a resampling approach to cross-validation based on Monte Carlo cross-validation for fair model assessment, and we also used a robust, analytical, computationally efficient leave-one-out cross-validation strategy to set the regularization parameter of the networks developed in experiments A and B. Randomization rather than optimization of the hidden layers, combined with an efficient C++ implementation of our models, allowed the construction of software that

enables on-the-fly computation and validation (LOOCV) of deep architecture models for any combination of traits with minimal to no technical knowledge on the part of the user.

The problem of interpretability and explainability is a current issue in computational systems using machine learning techniques and constitutes an active topic of research in artificial intelligence [143]. A detailed methodological and implementation analysis will be the focus of a future work, but we briefly describe here how we handle the issue of explainability and interpretability in age-at-death using the neural networks with our software. As previously stated, we can look at the neural network fitted using the techniques described in this manuscript as a regularized linear model operating on the non-linear features extracted by the hidden layers concatenated with the original input (skip layer). We can exploit this property and use the intuitive and additive nature intrinsic to linear models and build a linear surrogate model to explain or interpret any neural network and its predictions.

In DRNNAGE, we regress the cross-validated predictions of the DRNN model on the original input of the network. We decorrelate the input data using the previously described sphering technique and standardize it to zero mean and unit variance. This results in a surrogate model where the intercept or baseline is the average of network estimates, and a new estimate can be “explained” by the sum of the contributions of individual traits to arrive at an approximation of the network estimate (Figure 8).



**Figure 8.** Explanation of an estimate by a linear surrogate model as performed by DRNNAGE software.

Our results suggest that a regression-based framework produces accurate age estimation in adult individuals. Prediction intervals can be estimated with ease and computational efficiency. Bayesian approaches [16,20,23] could have been used for this purpose but they encapsulate a different philosophy to data analysis and are more restrictive in regard to assumptions, parameterization, and computational efficiency compared to the ANN approach we pursued here. Recent contributions suggest that Bayesian approaches do not radically improve age-at-death estimation or outperform regression-based approaches [144,145].

The predictive modeling or function approximation approach pursued in this work is, at the same time, its strongest point and its key limitation. Although neural networks as function approximation machines allowed us to obtain individual accurate age estimates, a predictive modeling strategy—regardless of the underlying algorithm—can only demonstrate that there is an efficient mapping in the form of  $y = f^*(x)$ . Such a strategy does not explain the underlying biology of the skeletal traits. Fully understanding the biology of the skeletal traits used in age estimation is perhaps the greatest challenge of



this problem, and perhaps the solution for more refined age estimation based solely on the skeletal morphology.

Despite the promising results, the current research did not emerge in a vacuum, nor has it any pretension to be a one-size-fits-all solution to skeletal age estimation, because it was inspired by significant work that was previously developed on this topic, see [16,19,24,35,140].

An important technical and methodological aspect that deserves a detailed analysis in the future is intra- and interobserver error. The results demonstrate the proposed scoring method is highly reproducible. This can be explained by the fact that most traits are encoded in a binary fashion; nonetheless, more data are required from an independent third party that applies the method as described here.

One last aspect that deserves discussion is the dataset employed in this study. The constructed dataset aimed to be uniform and homogeneous in respect to age-at-death and sex. At the moment, it only represents Portuguese nationals over a broad time span; thus, it would be important to expand the dataset to include individuals from other regions, and ascertain possible population and temporal differences in the performance of the proposed method.

## 5. Conclusions

The work presented here is an important and valuable contribution to the field of age-at-death estimation. Our results clearly demonstrated that a multifactorial approach improves accuracy and precision over single anatomic regions, as established in traditional adult skeletal aging methods. Multifactorial neural models introduce a two-to-six-fold reduction in the mean absolute error and prediction bias compared to standard models. This research also demonstrated that it is possible to produce informative age estimates for the elderly and that nonstandard skeletal traits are pivotal in the later stage of the adult age span. As an age estimation technique developed with forensic casework as its applicational domain, proper validation by other researchers and practitioners is most needed as we are aware that our results, as solid as they are, reflect only *in silico* performance and cross-validation. This work clearly demonstrated that neural network models offer excellent predictive accuracy. A current issue to be further investigated in future research work is the problem of interpretability and explainability. We briefly alluded to how this problem can be tackled using a global surrogate modeling approach, but other techniques will be investigated in the future so that age-at-death estimation can be approached with computationally accurate and intelligible techniques.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biology11040532/s1>. Table S1: Scoring system for suture obliteration. Table S2: List of cranial and palatine suture segment analyzed. Table S3: Scoring system for S1-S2 fusion. Table S4: Scoring system for vertebral body development and degeneration. Table S5: List of traits analyzed in the cervical, lumbar, and sacral vertebrae. Table S6: List of traits used to assess joint and musculoskeletal degeneration of the limbs. Table S7: Generic scoring system for joint degeneration traits. Table S8: Generic scoring system for musculoskeletal degeneration traits. Table S9: Stage 1 specific descriptions for selected joint and musculoskeletal degeneration traits. Table S10: Scoring system for clavicle age-related traits. Table S11: Scoring system for the first rib age-related traits. Table S12: Scoring system for the pubic symphysis age-related traits. Table S13: Scoring system for the sacral auricular age-related traits. Table S14: Scoring system for the iliac auricular age-related traits. Table S15: Scoring system for the acetabular age-related traits.

**Author Contributions:** Conceptualization, D.N.; Data curation, D.N.; Methodology, D.N.; Software, D.N.; Supervision, E.C. (Ernesto Costa) and E.C. (Eugénia Cunha); Validation, D.N.; Visualization, D.N.; Writing—original draft, D.N., E.C. (Ernesto Costa) and E.C. (Eugénia Cunha); Writing—review and editing, D.N., E.C. (Ernesto Costa) and E.C. (Eugénia Cunha). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Fundação para a Ciência e Tecnologia, grant number SFRH/BD/99676/2014.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data used for this manuscript preparation is embedded in the software DRNNAGE and it is available from GitHub, see Section 2.6.4. Raw data is available upon request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Dirkmaat, D.C.; Cabo, L.L.; Ousley, S.D.; Symes, S.A. New perspectives in forensic anthropology. *Am. J. Phys. Anthropol.* **2008**, *137*, 33–52. [[CrossRef](#)] [[PubMed](#)]
2. Dirkmaat, D.C.; Cabo, L.L. Embracing the New Paradigm. In *A Companion to Forensic Anthropology*, 1st ed.; Dirkmaat, D.C., Ed.; Blackwell Publishing Ltd.: Hoboken, NJ, USA; pp. 1–40.
3. Ritz-Timme, S.; Cattaneo, C.; Collins, M.; Waite, E.R.; Schütz, H.W.; Kaatsch, H.-J.; Borrman, H.I.M. Age estimation: The state of the art in relation to the specific demands of forensic practise. *Int. J. Legal Med.* **2000**, *113*, 129–136. [[CrossRef](#)] [[PubMed](#)]
4. Ferrante, L.; Cameriere, R. Statistical methods to assess the reliability of measurements in the procedures for forensic age estimation. *Int. J. Legal Med.* **2009**, *123*, 277–283. [[CrossRef](#)] [[PubMed](#)]
5. Garvin, H.M.; Passalacqua, N.V. Current Practices by Forensic Anthropologists in Adult Skeletal Age Estimation. *J. Forensic Sci.* **2011**, *57*, 427–433. [[CrossRef](#)] [[PubMed](#)]
6. Franklin, D. Forensic age estimation in human skeletal remains: Current concepts and future directions. *Leg. Med.* **2010**, *12*, 1–7. [[CrossRef](#)] [[PubMed](#)]
7. Rösing, F.; Graw, M.; Marré, B.; Ritz-Timme, S.; Rothschild, M.; Röttscher, K.; Schmeling, A.; Schröder, I.; Geserick, G. Recommendations for the forensic diagnosis of sex and age from skeletons. *HOMO J. Comp. Hum. Biol.* **2007**, *58*, 75–89. [[CrossRef](#)]
8. Kimmerle, E.H.; Prince, D.A.; Berg, G.E. Inter-Observer Variation in Methodologies Involving the Pubic Symphysis, Sternal Ribs, and Teeth. *J. Forensic Sci.* **2008**, *53*, 594–600. [[CrossRef](#)]
9. Martrille, L.; Ubelaker, D.H.; Cattaneo, C.; Seguret, F.; Tremblay, M.; Baccino, E. Comparison of Four Skeletal Methods for the Estimation of Age at Death on White and Black Adults. *J. Forensic Sci.* **2007**, *52*, 302–307. [[CrossRef](#)]
10. Buckberry, J. The (mis)use of adult age estimates in osteology. *Ann. Hum. Biol.* **2015**, *42*, 323–331. [[CrossRef](#)]
11. Bocquet-Appel, J.-P.; Masset, C. Farewell to paleodemography. *J. Hum. Evol.* **1982**, *11*, 321–333. [[CrossRef](#)]
12. Samworth, R.; Gowland, R. Estimation of adult skeletal age-at-death: Statistical assumptions and applications. *Int. J. Osteoarchaeol.* **2006**, *17*, 174–188. [[CrossRef](#)]
13. Kotěrová, A.; Navega, D.; Štepanovský, M.; Buk, Z.; Brůžek, J.; Cunha, E. Age estimation of adult human remains from hip bones using advanced methods. *Forensic Sci. Int.* **2018**, *287*, 163–175. [[CrossRef](#)] [[PubMed](#)]
14. Navega, D.; Costa, E.; Cunha, E. Lost in the woods: The value of tree ensemble modelling for adult age-at-death estimation from skeletal degeneration. *La Rev. Med. Légale* **2017**, *8*, 181–182. [[CrossRef](#)]
15. Navega, D.; Coelho, J.D.; Cunha, E.; Curate, F. DXAGE: A New Method for Age at Death Estimation Based on Femoral Bone Mineral Density and Artificial Neural Networks. *J. Forensic Sci.* **2017**, *63*, 497–503. [[CrossRef](#)]
16. Konigsberg, L.W. Multivariate cumulative probit for age estimation using ordinal categorical data. *Ann. Hum. Biol.* **2015**, *42*, 368–378. [[CrossRef](#)]
17. Lucy, D.; Aykroyd, R.G.; Pollard, A.M. Nonparametric calibration for age estimation. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2002**, *51*, 183–196. [[CrossRef](#)]
18. Konigsberg, L.; Frankenberg, S.R. Estimation of age structure in anthropological demography. *Am. J. Phys. Anthropol.* **1992**, *89*, 235–256. [[CrossRef](#)]
19. Boldsen, J.L.; Milner, G.R.; Konigsberg, L.W.; Wood, J.W. Transition analysis: A new method for estimating age from the skeletons. In *Paleodemography: Age Distributions from Skeletal Samples*, 1st ed.; Hoppa, R.D., Vaupel, J.W., Eds.; Cambridge University Press: Cambridge, UK; pp. 73–106.
20. Milner, G.R.; Boldsen, J.L. Transition analysis: A validation study with known-age modern American skeletons. *Am. J. Phys. Anthropol.* **2012**, *148*, 98–110. [[CrossRef](#)]
21. Steadman, D.W.; Adams, B.J.; Konigsberg, L.W. Statistical basis for positive identification in forensic anthropology. *Am. J. Phys. Anthropol.* **2006**, *131*, 15–26. [[CrossRef](#)]
22. Corsini, M.-M.; Schmitt, A.; Bruzek, J. Aging process variability on the human skeleton: Artificial network as an appropriate tool for age at death assessment. *Forensic Sci. Int.* **2005**, *148*, 163–167. [[CrossRef](#)]
23. Martins, R.; Oliveira, P.E.; Schmitt, A. Estimation of age at death from the pubic symphysis and the auricular surface of the ilium using a smoothing procedure. *Forensic Sci. Int.* **2012**, *219*, 287.e1–287.e7. [[CrossRef](#)] [[PubMed](#)]
24. Buk, Z.; Kordik, P.; Bruzek, J.; Schmitt, A.; Snorek, M. The age at death assessment in a multi-ethnic sample of pelvic bones using nature-inspired data mining methods. *Forensic Sci. Int.* **2012**, *220*, 294.e1–294.e9. [[CrossRef](#)] [[PubMed](#)]
25. Baccino, E.; Ubelaker, D.H.; Hayek, L.-A.C.; Zerilli, A. Evaluation of Seven Methods of Estimating Age at Death from Mature Human Skeletal Remains. *J. Forensic Sci.* **1999**, *44*, 931–936. [[CrossRef](#)] [[PubMed](#)]

26. Brooks, S.; Suchey, J.M. Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Hum. Evol.* **1990**, *5*, 227–238. [[CrossRef](#)]
27. Hanihara, K.; Suzuki, T. Estimation of age from the pubic symphysis by means of multiple regression analysis. *Am. J. Phys. Anthropol.* **1978**, *48*, 233–239. [[CrossRef](#)] [[PubMed](#)]
28. Gilbert, B.M.; McKern, T.W. A method for aging the female Os pubis. *Am. J. Phys. Anthropol.* **1973**, *38*, 31–38. [[CrossRef](#)] [[PubMed](#)]
29. McKern, T.W.; Stewart, T.D. *Skeletal Age Changes in Young American Males, Analyzed from the Standpoint of Age Identification*; Cambridge University Press: Natick, MA, USA, 1957.
30. Todd, T.W. Age changes in the pubic bone. I. The male white pubis. *Am. J. Phys. Anthropol.* **1920**, *3*, 285–334. [[CrossRef](#)]
31. Todd, T.W. Age changes in the pubic bone. *Am. J. Phys. Anthropol.* **1921**, *4*, 1–70. [[CrossRef](#)]
32. Katz, D.; Suchey, J.M. Age determination of the male Os pubis. *Am. J. Phys. Anthropol.* **1986**, *69*, 427–435. [[CrossRef](#)]
33. Stoyanova, D.; Algee-Hewitt, B.F.; Slice, D.E. An enhanced computational method for age-at-death estimation based on the pubic symphysis using 3D laser scans and thin plate splines. *Am. J. Phys. Anthropol.* **2015**, *158*, 431–440. [[CrossRef](#)]
34. Slice, D.E.; Algee-Hewitt, B.F. Modeling Bone Surface Morphology: A Fully Quantitative Method for Age-at-Death Estimation Using the Pubic Symphysis. *J. Forensic Sci.* **2015**, *60*, 835–843. [[CrossRef](#)] [[PubMed](#)]
35. Milner, G.R.; Boldsen, J.L. Skeletal Age Estimation: Where We Are and Where We Should Go. In *A Companion to Forensic Anthropology*, 1st ed.; Dirkmaat, D.C., Ed.; Blackwell Publishing Ltd.: Hoboken, NJ, USA, 2017; pp. 224–238.
36. Milner, G.R.; Boldsen, J.L. Estimating Age and Sex from the Skeleton, a Paleopathological Perspective. In *A Companion to Paleopathology*, 1st ed.; Grauer, A.L., Ed.; Blackwell Publishing Ltd.: Hoboken, NJ, USA, 2008; pp. 268–284.
37. Vilas-Boas, D.; Wasterlain, S.; d’Oliveira Coelho, J.; Navega, D.; Gonçalves, D. SPINNE: An app for human vertebral height estimation based on artificial neural networks. *Forensic Sci. Int.* **2019**, *298*, 121–130. [[CrossRef](#)] [[PubMed](#)]
38. Scott, G.R.; Pilloud, M.A.; Navega, D.; Coelho, J.D.; Cunha, E.; Irish, J.D. rASUDAS: A New Web-Based Application for Estimating Ancestry from Tooth Morphology. *Forensic Anthropol.* **2018**, *1*, 18–31. [[CrossRef](#)]
39. Navega, D.; Coelho, C.; Vicente, R.; Ferreira, M.T.; Wasterlain, S.; Cunha, E. AncesTrees: Ancestry estimation with randomized decision trees. *Int. J. Leg. Med.* **2014**, *129*, 1145–1153. [[CrossRef](#)]
40. Damas, S.; Wilkinson, C.; Kahana, T.; Veselovskaya, E.; Abramov, A.; Jankauskas, R.; Jayaprakash, P.; Ruiz, E.; Navarro, F.; Huete, M.; et al. Study on the performance of different craniofacial superimposition approaches (II): Best practices proposal. *Forensic Sci. Int.* **2015**, *257*, 504–508. [[CrossRef](#)]
41. Mesejo, P.; Martos, R.; Ibáñez; Novo, J.; Ortega, M. A Survey on Artificial Intelligence Techniques for Biomedical Image Analysis in Skeleton-Based Forensic Human Identification. *Appl. Sci.* **2020**, *10*, 4703. [[CrossRef](#)]
42. Cunha, E.; Wasterlain, S. The Coimbra identified osteological collections. In *Skeletal Series and Their Socio-Economic Context (Documenta Archaeobiologiae; Bd. 5)*, 1st ed.; Grupe, G., Peters, J., Eds.; Verlag Marie Leidorf GmbH, Rahden/Westf.: Rahden, Germany, 2013; pp. 23–34.
43. Ferreira, M.T.; Vicente, R.; Navega, D.; Gonçalves, D.; Curate, F.; Cunha, E. A new forensic collection housed at the University of Coimbra, Portugal: The 21st century identified skeletal collection. *Forensic Sci. Int.* **2014**, *245*, 202.e1–202.e5. [[CrossRef](#)]
44. Ferreira, M.T.; Coelho, C.; Makhoul, C.; Navega, D.; Gonçalves, D.; Cunha, E.; Curate, F. New data about the 21st Century Identified Skeletal Collection (University of Coimbra, Portugal). *Int. J. Leg. Med.* **2020**, *135*, 1087–1094. [[CrossRef](#)]
45. Usher, B.M. Reference samples: The first step in linking biology and age in the human skeleton. In *Paleodemography*; Hoppa, R.D., Vaupel, J.W., Eds.; Cambridge University Press: Cambridge, UK, 2018; pp. 29–47.
46. Beretta, L.; Santaniello, A. Nearest neighbor imputation algorithms: A critical evaluation. *BMC Med. Inform. Decis. Mak.* **2016**, *16*, 197–208. [[CrossRef](#)]
47. Kemkes-Grotenthaler, A. Aging through the ages: Historical perspectives on age indicators methods. In *Paleodemography: Age Distributions from Skeletal Samples*, 1st ed.; Hoppa, R.D., Vaupel, J.W., Eds.; Cambridge University Press: Cambridge, UK, 2016; pp. 48–72.
48. Shirley, N.R.; Montes, P.A.R. Age Estimation in Forensic Anthropology: Quantification of Observer Error in Phase versus Component-Based Methods. *J. Forensic Sci.* **2014**, *60*, 107–111. [[CrossRef](#)]
49. Todd, T.W.; Lyon, D.W. Cranial Suture Closure, Its Progress and Age Relationship. Part I. *Am. J. Phys. Anthropol.* **1925**, *8*, 325–384.
50. Todd, T.W.; Lyon, D.W. Cranial suture closure Part II. *Am. J. Phys. Anthropol.* **1925**, *9*, 23–44. [[CrossRef](#)]
51. Meindl, R.S.; Lovejoy, O. Ectocranial suture closure: A revised method for the determination of skeletal age at death and blind tests of its accuracy. *Am. J. Phys. Anthropol.* **1985**, *68*, 57–66. [[CrossRef](#)] [[PubMed](#)]
52. Perizonius, W.R.K. Closing and non-closing sutures in 256 crania of known age and sex from Amsterdam (a.d. 1883–1909). *J. Hum. Evol.* **1984**, *13*, 201–216. [[CrossRef](#)]
53. Mann, R.W.; Jantz, R.L.; Bass, W.M.; Willey, P.S. Maxillary Suture Obliteration: A Visual Method for Estimating Skeletal Age. *J. Forensic Sci.* **1991**, *36*, 781–791. [[CrossRef](#)]
54. Mann, R.W.; Symes, S.A.; Bass, W.M. Maxillary Suture Obliteration: Aging the Human Skeleton Based on Intact or Fragmentary Maxilla. *J. Forensic Sci.* **1987**, *32*, 148–157. [[CrossRef](#)]
55. Acsadi, J.; Nemeskéri, G. *History of Human Life Span and Mortality*; Akadémiai Kiadó: Budapest, Hungary, 1970.
56. Masset, C. Age estimation on the basis of cranial sutures. In *Age Markers in the Human Skeleton*; İşcan, M.Y., Ed.; Charles C Thomas: Springfield, IL, USA, 1989; pp. 71–103.
57. Ríos, L.; Weisensee, K.; Rissech, C. Sacral fusion as an aid in age estimation. *Forensic Sci. Int.* **2008**, *180*, 111.e1–111.e7. [[CrossRef](#)]

58. Belcastro, M.G.; Rastelli, E.; Mariotti, V. Variation of the degree of sacral vertebral body fusion in adulthood in two European modern skeletal collections. *Am. J. Phys. Anthropol.* **2007**, *135*, 149–160. [[CrossRef](#)]
59. Passalacqua, N.V. Forensic Age-at-Death Estimation from the Human Sacrum. *J. Forensic Sci.* **2009**, *54*, 255–262. [[CrossRef](#)]
60. Snodgrass, J.J. Sex Differences and Aging of the Vertebral Column. *J. Forensic Sci.* **2004**, *49*, 1–6. [[CrossRef](#)]
61. Watanabe, S.; Terazawa, K. Age estimation from the degree of osteophyte formation of vertebral columns in Japanese. *Leg. Med.* **2006**, *8*, 156–160. [[CrossRef](#)] [[PubMed](#)]
62. Albert, M.; Mulhern, D.; Torpey, M.A.; Boone, E. Age Estimation Using Thoracic and First Two Lumbar Vertebral Ring Epiphyseal Union. *J. Forensic Sci.* **2010**, *55*, 287–294. [[CrossRef](#)] [[PubMed](#)]
63. Alves-Cardoso, F.; Assis, S. Can osteophytes be used as age at death estimators? Testing correlations in skeletonized human remains with known age-at-death. *Forensic Sci. Int.* **2018**, *288*, 59–66. [[CrossRef](#)] [[PubMed](#)]
64. Milella, M.; Belcastro, M.G.; Mariotti, V.; Nikita, E. Estimation of adult age-at-death from enthesal robusticity: A test using an identified Italian skeletal collection. *Am. J. Phys. Anthropol.* **2020**, *173*, 190–199. [[CrossRef](#)]
65. Milella, M.; Belcastro, M.G.; Zollikofer, C.P.; Mariotti, V. The effect of age, sex, and physical activity on enthesal morphology in a contemporary Italian skeletal collection. *Am. J. Phys. Anthropol.* **2012**, *148*, 379–388. [[CrossRef](#)]
66. Winburn, A.P.; Stock, M.K. Reconsidering osteoarthritis as a skeletal indicator of age at death. *Am. J. Phys. Anthropol.* **2019**, *170*, 459–473. [[CrossRef](#)]
67. Calce, S.E.; Kurki, H.K.; Weston, D.A.; Gould, L. Effects of osteoarthritis on age-at-death estimates from the human pelvis. *Am. J. Phys. Anthropol.* **2018**, *167*, 3–19. [[CrossRef](#)]
68. Calce, S.E.; Kurki, H.K.; Weston, D.A.; Gould, L. Principal component analysis in the evaluation of osteoarthritis. *Am. J. Phys. Anthropol.* **2016**, *162*, 476–490. [[CrossRef](#)]
69. Brennaman, A.L.; Love, K.R.; Bethard, J.D.; Pokines, J.T. A Bayesian Approach to Age-at-Death Estimation from Osteoarthritis of the Shoulder in Modern North Americans. *J. Forensic Sci.* **2016**, *62*, 573–584. [[CrossRef](#)]
70. Calce, S.E.; Kurki, H.K.; Weston, D.A.; Gould, L. The relationship of age, activity, and body size on osteoarthritis in weight-bearing skeletal regions. *Int. J. Paleopathol.* **2018**, *22*, 45–53. [[CrossRef](#)]
71. Buikstra, J.E.; Ubelaker, D.H. Standards for Data Collection from Human Skeletal Remains. *Fayetteville* **1995**. [[CrossRef](#)]
72. Henderson, C.Y.; Mariotti, V.; Pany-Kucera, D.; Villotte, S.; Wilczak, C. Recording Specific Enthesal Changes of Fibrocartilaginous Entheses: Initial Tests Using the Coimbra Method. *Int. J. Osteoarchaeol.* **2012**, *23*, 152–162. [[CrossRef](#)]
73. Falys, C.G.; Prangle, D. Estimating age of mature adults from the degeneration of the sternal end of the clavicle. *Am. J. Phys. Anthropol.* **2014**, *156*, 203–214. [[CrossRef](#)]
74. Langley-Shirley, N.; Jantz, R.L. A Bayesian Approach to Age Estimation in Modern Americans from the Clavicle. *J. Forensic Sci.* **2010**, *55*, 571–583. [[CrossRef](#)] [[PubMed](#)]
75. Owings, W.P.A.; Myers, S.J.; Webb, P.A.O.; Suchey, J.M. Epiphyseal union of the anterior iliac crest and medial clavicle in a modern multiracial sample of American males and females. *Am. J. Phys. Anthropol.* **1985**, *68*, 457–466. [[CrossRef](#)]
76. Cardoso, H.F.V. V Age estimation of adolescent and young adult male and female skeletons II, epiphyseal union at the upper limb and scapular girdle in a modern Portuguese skeletal sample. *Am. J. Phys. Anthropol.* **2008**, *137*, 97–105. [[CrossRef](#)]
77. Can, M.Y.; Loth, S.R. Determination of age from the sternal rib in white males: A test of the phase method. *J. Forensic Sci.* **1986**, *31*, 122–132.
78. İşcan, M.Y.; Loth, S.R.; Wright, R.K. Age Estimation from the Rib by Phase Analysis: White Males. *J. Forensic Sci.* **1984**, *29*, 1094–1104. [[CrossRef](#)]
79. Can, M.Y.; Loth, S.R. Metamorphosis at the sternal rib end: A new method to estimate age at death in white males. *Am. J. Phys. Anthropol.* **1984**, *65*, 147–156.
80. İşcan, M.Y.; Loth, S.R.; Wright, R.K. Racial Variation in the Sternal Extremity of the Rib and Its Effect on Age Determination. *J. Forensic Sci.* **1987**, *32*, 452–466. [[CrossRef](#)]
81. Kunos, C.A.; Simpson, S.W.; Russell, K.F. Hershkovitz I First rib metamorphosis: Its possible utility for human age-at-death estimation. *Am. J. Phys. Anthropol.* **1999**, *110*, 303–323. [[CrossRef](#)]
82. DiGangi, E.A.; Bethard, J.D.; Kimmerle, E.H.; Konigsberg, L.W. A new method for estimating age-at-death from the first rib. *Am. J. Phys. Anthropol.* **2009**, *138*, 164–176. [[CrossRef](#)] [[PubMed](#)]
83. Sashin, D.A. Critical analysis of the anatomy and the pathologic changes of the sacro-iliac joints. *J. Bone Jt. Surg.* **1992**, *12*, 891–910.
84. Schunke, G.B. The anatomy and development of the sacro-iliac joint in man. *Anat. Rec.* **1938**, *72*, 313–331. [[CrossRef](#)]
85. Lovejoy, C.O.; Meindl, R.S.; Pryzbeck, T.R.; Mensforth, R.P. Chronological metamorphosis of the auricular surface of the ilium: A new method for the determination of adult skeletal age at death. *Am. J. Phys. Anthropol.* **1985**, *68*, 15–28. [[CrossRef](#)]
86. Buckberry, J.; Chamberlain, A. Age estimation from the auricular surface of the ilium: A revised method. *Am. J. Phys. Anthropol.* **2002**, *119*, 231–239. [[CrossRef](#)]
87. San-Millán, M.; Rissech, C.; Turbón, D. New approach to age estimation of male and female adult skeletons based on the morphological characteristics of the acetabulum. *Int. J. Leg. Med.* **2016**, *131*, 501–525. [[CrossRef](#)]
88. Winburn, A.P. Validation of the Acetabulum as a Skeletal Indicator of Age at Death in Modern European-Americans. *J. Forensic Sci.* **2018**, *64*, 989–1003. [[CrossRef](#)]
89. Mays, S. A Test of a Recently Devised Method of Estimating Skeletal Age at Death using Features of the Adult Acetabulum. *J. Forensic Sci.* **2013**, *59*, 184–187. [[CrossRef](#)]

90. Calce, S.E. A new method to estimate adult age-at-death using the acetabulum. *Am. J. Phys. Anthropol.* **2012**, *148*, 11–23. [[CrossRef](#)]
91. Rissech, C.; Estabrook, G.F.; Cunha, E.; Malgosa, A. Using the Acetabulum to Estimate Age at Death of Adult Males. *J. Forensic Sci.* **2006**, *51*, 213–229. [[CrossRef](#)] [[PubMed](#)]
92. Rissech, C.; Estabrook, G.F.; Cunha, E.; Malgosa, A.; Badalló, M.D.C.R. Estimation of Age-at-Death for Adult Males Using the Acetabulum, Applied to Four Western European Populations. *J. Forensic Sci.* **2007**, *52*, 774–778. [[CrossRef](#)] [[PubMed](#)]
93. Rougé-Maillart, C.; Vielle, B.; Jousset, N.; Chappard, D.; Telmon, N.; Cunha, E. Development of a method to estimate skeletal age at death in adults using the acetabulum and the auricular surface on a Portuguese population. *Forensic Sci. Int.* **2009**, *188*, 91–95. [[CrossRef](#)] [[PubMed](#)]
94. Rougé-Maillart, C.; Jousset, N.; Vielle, B.; Gaudin, A.; Telmon, N. Contribution of the study of acetabulum for the estimation of adult subjects. *Forensic Sci. Int.* **2007**, *171*, 103–110. [[CrossRef](#)] [[PubMed](#)]
95. San-Millán, M.; Rissech, C.; Turbón, D. Application of the recent SanMillán–Rissech acetabular adult aging method in a North American sample. *Int. J. Leg. Med.* **2019**, *133*, 909–920. [[CrossRef](#)]
96. Kendall, M.G.; Smith, B.B. The Problem of m Rankings. *Ann. Math. Stat.* **1939**, *10*, 275–287. [[CrossRef](#)]
97. Zuber, V.; Strimmer, K. High-Dimensional Regression and Variable Selection Using CAR Scores. *Stat. Appl. Genet. Mol. Biol.* **2011**, *10*. [[CrossRef](#)]
98. Kessy, A.; Lewin, A.; Strimmer, K. Optimal Whitening and Decorrelation. *Am. Stat.* **2018**, *72*, 309–314. [[CrossRef](#)]
99. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2015.
100. Scardapane, S.; Wang, D. Randomness in neural networks: An overview. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2017**, *7*, e1200. [[CrossRef](#)]
101. Gallicchio, C.; Scardapane, S. Deep Randomized Neural Networks. *Stud. Comput. Intell.* **2020**, *896*, 43–68. [[CrossRef](#)]
102. Schmidt, W.F.; Kraaijeveld, M.A.; Duin, R.P.W. Feedforward neural networks with random weights. In Proceedings of the 11th IAPR International Conference on Pattern Recognition. Volume II. Conference B: Pattern Recognition Methodology and Systems, The Hague, The Netherlands, 30 August 1992; IEEE Comput. Soc. Press: Piscataway, NJ, USA, 1992; pp. 1–4.
103. Pao, Y.-H.; Takefuji, Y. Functional-link net computing: Theory, system architecture, and functionalities. *Computer (Long Beach Ca.)* **1992**, *25*, 76–79. [[CrossRef](#)]
104. Broomhead, D.S.; Lowe, D. Multivariable functional interpolation and adaptive networks. *Complex Syst.* **1988**, *2*, 321–355.
105. Pao, Y.-H.; Park, G.-H.; Sobajic, D.J. Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing* **1994**, *6*, 163–180. [[CrossRef](#)]
106. Igel'nik, B.; Pao, Y.-H. Stochastic choice of basis functions in adaptive function approximation and the functional-link net. *IEEE Trans. Neural Netw.* **1995**, *6*, 1320–1329. [[CrossRef](#)] [[PubMed](#)]
107. Huang, G.-B. What are Extreme Learning Machines? Filling the Gap between Frank Rosenblatt's Dream and John von Neumann's Puzzle. *Cogn. Comput.* **2015**, *7*, 263–278. [[CrossRef](#)]
108. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]
109. Huang, G.; Huang, G.; Bin Song, S.; You, K. Trends in extreme learning machines: A review. *Neural Netw.* **2015**, *61*, 32–48. [[CrossRef](#)]
110. Huang, G.-B. An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels. *Cogn. Comput.* **2014**, *6*, 376–390. [[CrossRef](#)]
111. Wang, L.P.; Wan, C.R. Comments on “The Extreme Learning Machine. *IEEE Trans. Neural Netw.* **2008**, *19*, 1494–1495. [[CrossRef](#)]
112. Shao, Z.; Er, M.J. Efficient Leave-One-Out Cross-Validation-based Regularized Extreme Learning Machine. *Neurocomputing* **2016**, *194*, 260–270. [[CrossRef](#)]
113. Wang, D.; Wang, P.; Shi, J. A fast and efficient conformal regressor with regularized extreme learning machine. *Neurocomputing* **2018**, *304*, 1–11. [[CrossRef](#)]
114. Tissera, M.D.; McDonnell, M. Deep extreme learning machines: Supervised autoencoding architecture for classification. *Neurocomputing* **2016**, *174*, 42–49. [[CrossRef](#)]
115. Tang, J.; Deng, C.; Huang, G.-B. Extreme Learning Machine for Multilayer Perceptron. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *27*, 809–821. [[CrossRef](#)] [[PubMed](#)]
116. Zhou, H.; Huang, G.-B.; Lin, Z.; Wang, H.; Soh, Y.C. Stacked Extreme Learning Machines. *IEEE Trans. Cybern.* **2014**, *45*, 2013–2025. [[CrossRef](#)]
117. Yu, W.; Zhuang, F.; He, Q.; Shi, Z. Learning deep representations via extreme learning machines. *Neurocomputing* **2015**, *149*, 308–315. [[CrossRef](#)]
118. Shi, Q.; Katuwal, R.; Suganthan, P.; Tanveer, M. Random vector functional link neural network based ensemble deep learning. *Pattern Recognit.* **2021**, *117*, 107978. [[CrossRef](#)]
119. Huang, G.-B.; Zhou, H.; Ding, X.; Zhang, R. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2012**, *42*, 513–529. [[CrossRef](#)]
120. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **1970**, *12*, 55–67. [[CrossRef](#)]
121. Bartlett, P.L. The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network. *IEEE Trans. Inf. Theory* **1998**, *44*, 525–536. [[CrossRef](#)]

122. Bartlett, P.L. For valid generalization, the size of the weights is more important than the size of the network. *Adv. Neural Inf. Processing Syst.* 9 (NIPS 1996) **1996**, 23, 35–39.
123. Allen, D.M. The Relationship between Variable Selection and Data Augmentation and a Method for Prediction. *Technometrics* **1974**, 16, 125. [[CrossRef](#)]
124. Navega, D.; Cunha, E. Extreme learning machine neural networks for adult skeletal age-at-death estimation. In *Statistics and Probability in Forensic Anthropology*; Obertová, Z., Stewart, A., Cattaneo, C., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 209–225.
125. Milborrow, S. Variance Models in Earth. Available online: <http://www.milbo.org/doc/earth-varmod.pdf> (accessed on 18 March 2022).
126. Geary, R.C. The Ratio of the Mean Deviation to the Standard Deviation as a Test of Normality. *Biometrika* **1935**, 27, 310. [[CrossRef](#)]
127. Konigsberg, L.W.; Herrmann, N.P.; Wescott, D.J.; Kimmerle, E.H. Estimation and Evidence in Forensic Anthropology: Age-at-Death. *J. Forensic Sci.* **2008**, 53, 541–557. [[CrossRef](#)] [[PubMed](#)]
128. Müller, H.-G.; Love, B.; Hoppa, R.D. Semiparametric method for estimating paleodemographic profiles from age indicator data. *Am. J. Phys. Anthropol.* **2001**, 117, 1–14. [[CrossRef](#)] [[PubMed](#)]
129. Fieuws, S.; Willems, G.; Larsen-Tangmose, S.; Lynnerup, N.; Boldsen, J.; Thevissen, P. Obtaining appropriate interval estimates for age when multiple indicators are used: Evaluation of an ad-hoc procedure. *Int. J. Leg. Med.* **2015**, 130, 489–499. [[CrossRef](#)]
130. Papadopoulos, H.; Haralambous, H. Reliable prediction intervals with regression neural networks. *Neural Netw.* **2011**, 24, 842–851. [[CrossRef](#)]
131. Lappas, G. Estimating the Size of Neural Networks from the number of available training data. *Neural Netw.* **2007**, 68–77. [[CrossRef](#)]
132. Rougé-Maillart, C.; Telmon, N.; Rissech, C.; Malgosa, A.; Rougé, D. The determination of male adult age at death by central and posterior coxal analysis—a preliminary study. *J. Forensic Sci.* **2004**, 49, 1–7. [[CrossRef](#)]
133. Lovejoy, C.O.; Meindl, R.S.; Mensforth, R.P.; Barton, T.J. Multifactorial determination of skeletal age at death: A method and blind tests of its accuracy. *Am. J. Phys. Anthropol.* **1985**, 68, 1–14. [[CrossRef](#)]
134. Anderson, M.F.; Anderson, D.T.; Wescott, D.J. Estimation of adult skeletal age-at-death using the Sugeno fuzzy integral. *Am. J. Phys. Anthropol.* **2009**, 142, 30–41. [[CrossRef](#)]
135. Listi, G.A. The Use of Enteseal Changes in the Femur and Os Coxa for Age Assessment. *J. Forensic Sci.* **2015**, 61, 12–18. [[CrossRef](#)] [[PubMed](#)]
136. Listi, G.A.; Manhein, M.H. The Use of Vertebral Osteoarthritis and Osteophytosis in Age Estimation. *J. Forensic Sci.* **2012**, 57, 1537–1540. [[CrossRef](#)] [[PubMed](#)]
137. Jooste, N.; L'Abbé, E.N.; Pretorius, S.; Steyn, M. Validation of transition analysis as a method of adult age estimation in a modern South African sample. *Forensic Sci. Int.* **2016**, 266, 580.e1–580.e7. [[CrossRef](#)] [[PubMed](#)]
138. Baccino, E.; Sinfield, L.; Colomb, S.; Baum, T.P.; Martrille, L. Technical note: The two step procedure (TSP) for the determination of age at death of adult human remains in forensic cases. *Forensic Sci. Int.* **2014**, 244, 247–251. [[CrossRef](#)]
139. Miranker, M. A Comparison of Different Age Estimation Methods of the Adult Pelvis. *J. Forensic Sci.* **2016**, 61, 1173–1179. [[CrossRef](#)]
140. Merritt, C.E. Inaccuracy and bias in adult skeletal age estimation: Assessing the reliability of eight methods on individuals of varying body sizes. *Forensic Sci. Int.* **2017**, 275, 315.e1–315.e11. [[CrossRef](#)]
141. Hagelthorn, C.L.; Alblas, A.; Greyling, L. The accuracy of the Transition Analysis of aging on a heterogenic South African population. *Forensic Sci. Int.* **2019**, 297, 370.e1–370.e5. [[CrossRef](#)]
142. Valsecchi, A.; Olivares, J.I.; Mesejo, P. Age estimation in forensic anthropology: Methodological considerations about the validation studies of prediction models. *Int. J. Leg. Med.* **2019**, 133, 1915–1924. [[CrossRef](#)]
143. Burkart, N.; Huber, M.F. A Survey on the Explainability of Supervised Machine Learning. *J. Artif. Intell. Res.* **2021**, 70, 245–317. [[CrossRef](#)]
144. Jooste, N.; Pretorius, S.; Steyn, M. Performance of three mathematical models for estimating age-at-death from multiple indicators of the adult skeleton. *Int. J. Leg. Med.* **2021**, 1–13. [[CrossRef](#)]
145. Nikita, E.; Nikitas, P. Skeletal age-at-death estimation: Bayesian versus regression methods. *Forensic Sci. Int.* **2019**, 297, 56–64. [[CrossRef](#)] [[PubMed](#)]