Received: 23 October 2023 | Revised: 16 January 2024 | Accepted: 30 January 2024

DOI: 10.1002/ajpa.24912

RESEARCH ARTICLE

AMERICAN JOURNAL OF BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

WILEY

Information fusion for infant age estimation from deciduous teeth using machine learning

Práxedes Martínez-Moreno[1,2] | Andrea Valsecchi[3] | Sergio Damas[4,2] | Javier Irurita[5] | Pablo Mesejo[1,2]

[1]Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain

[2]Andalusian Research Institute in Data Science and Computational Intelligence, University of Granada, Granada, Spain

[3]Panacea Cooperative Research S. Coop., Ponferrada, Spain

[4]Department of Software Engineering, University of Granada, Granada, Spain

[5]Department of Legal Medicine, Toxicology and Physical Anthropology, University of Granada, Granada, Spain

Correspondence
Práxedes Martínez-Moreno, Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain; Andalusian, Research Institute in Data Science and Computational Intelligence, University of Granada, Granada, Spain.
Email: praxedesmm@ugr.es

Funding information
Red.es Skeleton-ID2.0, Grant/Award Number: 2021/C005/00141299; FORAGE, Grant/Award Number: B-TIC-456-UGR20; Consejería de Universidad, Investigación e Innovación "ERDF A way of making Europe"; MCIU/AEI, Grant/Award Number: 10.13039/501100011033; CONFIA, Grant/Award Number: PID2021-122916NB-I00; Universidad de Granada/CBUA

Abstract

Objectives: Over the past few years, several methods have been proposed to improve the accuracy of age estimation in infants with a focus on dental development as a reliable marker. However, traditional approaches have limitations in efficiently combining information from different teeth and features. In order to address these challenges, this article presents a study on age estimation in infants with Machine Learning (ML) techniques, using deciduous teeth.

Materials and Methods: The involved dataset comprises 114 infant skeletons from the Granada osteological collection of identified infants, aged between 5 months of gestation and 3 years of age. The samples consist of features such as the maximum length and mineralization and alveolar stages of teeth. For the purpose of designing a method capable of combining all the information available from each individual, a Multilayer Perceptron model is proposed, one of the most popular artificial neural networks. This model has been validated using the leave-one-out experimental validation protocol. Through different groups of experiments, the study examines the informativeness of the aforementioned features, individually and in combination.

Results: The results indicate that the fusion of different variables allows for more accurate age estimates (RMSE = 66 days) than when variables are analyzed separately (RMSE = 101 days). Additionally, the study demonstrates the benefits of involving multiple teeth, which significantly reduces the RMSE compared to a single tooth.

Discussion: This article underlines the clear advantages of ML-based methods, emphasizing their potential to improve the accuracy and robustness when estimating the age of infants.

KEYWORDS

artificial intelligence, infant age estimation, information fusion, machine learning, physical anthropology

1 | INTRODUCTION

The development of precise, fast, robust, and automatic methods for age estimation is currently an area with a strong research interest in Physical Anthropology in which a majority of recent proposals focus on refining traditional methods with the aim of improving accuracy (Ubelaker et al., 2020). In recent years, various methods of age estimation in infants and young children have been proposed in order to provide more accurate estimates by the analysis of the development of deciduous teeth. There is a consensus that methods based on the development of deciduous teeth have demonstrated substantial reliability (Figueiro et al., 2022; Petrone et al., 2019). In sub-adults, dental development is one of the most stable markers of maturity and age estimation. The chronology of this process is a sequence of events that begins during the first trimester in the uterus, continues through infancy up to late adolescence, and ends with the maturation of the root tips of the third permanent molars (Liversidge, 2008).

Dental age estimation in subadults presents several advantages: it can be applied from prenatal age to adolescence, being the only source of information with this characteristic; in some casework, the enamel is the only element available due to its resistance and hardness; and precision in estimates is increased because dental maturation is influenced by genetics (Irurita et al., 2014), which appears to be independent of both skeletal and secondary sexual maturation, and less influenced by nutritional and other environmental factors. In addition, tooth formation proceeds at a chronologically regular rate, which is not the case for other maturing body systems (Liversidge, 2008). Hence, teeth are the organs that show the greatest correlation between biological and chronological age (Petrone et al., 2019).

Both qualitative and quantitative methods have been traditionally used to estimate the dental age of infants. Some of those qualitative methods analyze the degree of mineralization of dental tissues for each tooth separately (Demirjian et al., 1973; Irurita et al., 2014; Moorrees et al., 1963), whereas others combine both eruption and stages of tooth development offering illustrations in the form of atlases (AlQahtani et al., 2010; Schour & Massler, 1941; Ubelaker, 1989). Metric-based methods usually rely on odontometric variables (Cardoso et al., 2019; Irurita et al., 2013; Liversidge et al., 1993; Minier et al., 2013; Viciano et al., 2018). These methods employ traditional techniques that are easy to use and provide a reasonable level of accuracy. Other more complex approaches can be found in literature that use Artificial Intelligence (AI) (Russell & Norvig, 2020) to estimate age (Buk et al., 2012; Velemínská et al., 2013). These approaches may be more accurate than traditional techniques. However, some of them can be less explainable and interpretable alternatives (Štepanovský et al., 2017).

One of the main issues posed by the traditional methodology is the difficulty in combining information from different teeth and variables. If there is more than one tooth from which to estimate age, Liversidge (2015) recommends selecting the most appropriate one. Therefore, the selected tooth is the one with a stage of development closest to the definition proposed by the method, but never combining the information from several teeth unless the method specifically offers this possibility. In the case of development atlases, we could consider that they use information from several teeth in a combined way. However, development atlases should only be treated as unreliable guidance methods according to what Roberts et al. have recently stated (Roberts et al., 2023). Authors argue that such atlases do not have the precision and accuracy of other methods (even when analyzing a single tooth) and they lack clarity in expressing the assumed error. Traditional statistical analysis does not allow the design of methods that efficiently combine information from different teeth or the combination of qualitative and quantitative variables. In addition, the error associated with the estimation when traditional analysis methods are used is highly conditioned by the age distribution in both the reference and target samples (Sgheiza & Liversidge, 2023). Advances in Machine Learning (ML) (Murphy, 2022), which is the AI branch devoted to designing machines able to learn directly from data, could allow us to solve many of these problems. They have been successfully applied to different age-estimation methods (Gámez-Granados et al., 2022; Mohamed et al., 2023) to offer more objective, accurate and faster approaches.

In the years 2013 and 2014, two methods were proposed for estimating age in pediatric individuals through the analysis of the maximum length of deciduous teeth (Irurita et al., 2013) and their eruption and mineralization stages (Irurita et al., 2014). Both proposals utilized traditional analysis techniques such as least squares regression or simple frequency analysis for each tooth independently. The sample used was based on the collection of pediatric individuals from Granada (Alemán et al., 2012), one of the best collections available internationally for this age group. These methods can be suitable for age estimation, especially the one using maximum length (Petrone et al., 2019). However, they both have the limitations previously mentioned that are inherent to traditional analysis methods. To verify whether these limitations can be overcome, the aim of this study is to implement and validate a new method based on AI techniques to estimate age more accurately in infants and young children using deciduous teeth, utilizing the same data as in the previous proposals (Irurita et al., 2013, 2014). The method combines information on eruption, mineralization, and maximum tooth length. During this study, several questions have been investigated:

- Is it more informative to use only teeth lengths, emergence stages, maturation stages, or a combination of them?
- Is it better to use a single tooth to estimate age or a combination of teeth?
- Does the proposed method based on ML techniques outperform a traditional one in order to estimate the age of an individual?

The reasons for these questions are directly related to the motivations and contributions of this study:

- The need to work with as much information as possible, if available, when making an estimate for an individual in order to increase the accuracy. This means both considering several teeth at the same time and using multiple sources of

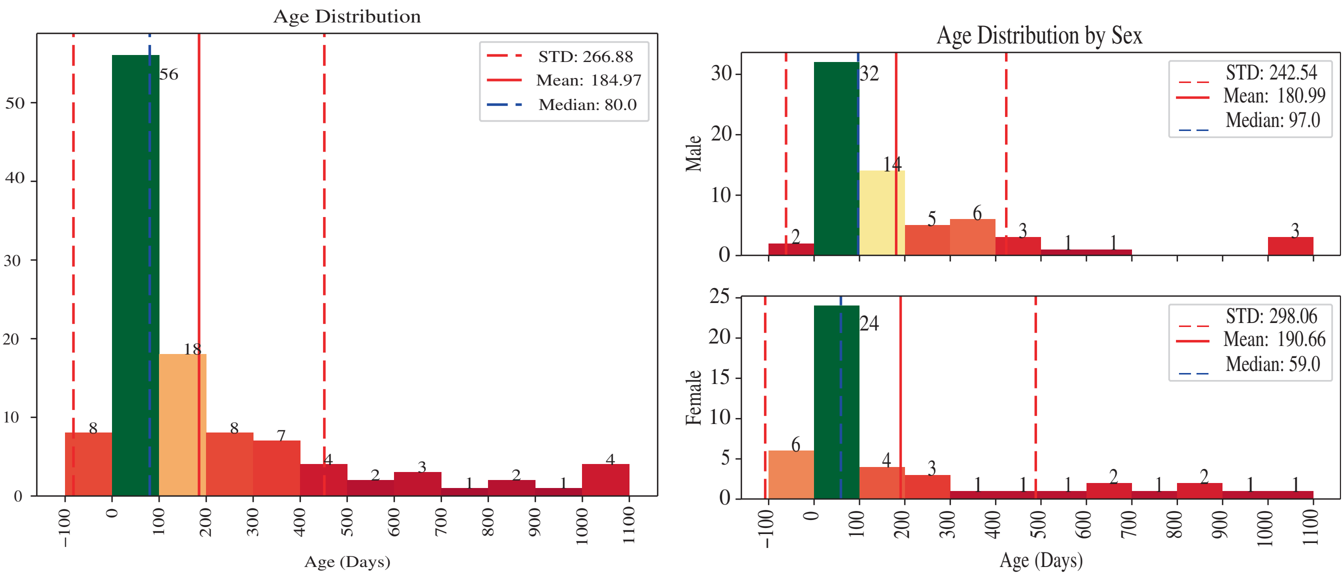

FIGURE 1 Age distribution of individuals within the dataset (left). Age distribution of individuals within the dataset grouped by sex (right): male (top) and female (bottom). The dotted line represents the median of the ages, the solid line represents the mean age, and the dashed lines represent the mean minus and plus the standard deviation. Note that the eight individuals with negative ages are fetuses whose months of gestation at the time of abortion are known from the burial records.

information (maturation stages, emergence stages and maximum tooth lengths).

- The possibility of working with missing data. While a method that employs multiple features is theoretically more powerful, its practical applicability may be limited when some teeth or information about them are missing. However, a method that is able to work with missing data allows the estimation to be carried out despite the lack of information on certain teeth. Moreover, a method that can work with missing information does not depend on the available teeth of an individual, unlike a traditional analysis method. The latter would need information about a specific tooth to perform the estimation.

2 | MATERIALS AND METHODS

The dataset used in this study comes from the Granada osteological collection of identified infants, exhumed from the cemetery of San Jose, Granada, Spain (Alemán et al., 2012). The collection comprises 230 identified individuals from the 20th century in a perfect state of preservation, whose antemortem data are documented, including sex, dates of birth and death, and immediate and underlying causes of death. Ages range from 5 months of gestation to 6 years of age.

In order to work with this dataset, several a priori exclusion criteria were applied: unknown age at death, premature birth (which disrupts the correspondence between age and level of development), and the presence of diseases that could affect dental development (Irurita et al., 2013). Note that premature infants are considered to be those whose cause of death, according to the death certificate, was prematurity. These are individuals who died between approximately 1 week and 1 month of life, but biologically they had a development of 7 or 8 months of gestation, so their inclusion in the study could alter the results. Teeth with signs of erosion, resorption or advanced caries were also excluded. In (AlQahtani et al., 2010) it is stated that, at an average age of 2.5 years (912.5 days), all the deciduous teeth are fully developed. From this age, the differences are thus minimal and not so useful for making estimates. For this reason and due to the lack of a representative number of individuals over this age in the dataset, the sample in this study was reduced by removing the cases over the age of 3 years (1095 days). The data thus consists of a set of 114 samples. In total, 854 teeth were analyzed. The distribution of the ages is depicted in Figure 1.

A digital gauge (0.01 mm margin of error) was used to measure the maximum lengths of the teeth (Irurita et al., 2013). The degree of mineralization was assessed macroscopically using the 11 mineralization stages proposed by Irurita et al. (Irurita et al., 2014). The alveolar emergence stage was assessed on the lateral aspect of the mandible or maxilla by distinguishing three readily identifiable stages: (1) alveolar emergence has not started; (2) alveolar emergence has started but without full crown exposure; and (3) complete alveolar emergence, in which the crown is fully exposed. Note that there may be unknown features if a tooth is missing. Table 1 summarizes the information about the features involved.

An initial pre-processing stage was carried out to prepare the data to build and validate the estimation model, in which missing data were imputed. Data imputation (Van Buuren, 2018) is a useful tool to replace the missing values that may be present in the dataset. The basic way of dealing with missing values is to drop whole instances or features, but due to the scarcity of data, this option was discarded. Different imputation methods were thus used in a preliminary study,

TABLE 1 Information about the features of the dataset.

	Type	Description	#Features
ID	Character string	String of characters used to identify the individual	1
Chronological age	Discrete variable (days)	Chronological age of the individual (from −100 to 1095 days)	1
Sex	Categorical variable (male or female)	Sex of the individual	1
Maximum length	Continuous variable (mm)	Maximum length of the tooth, reported in Irurita et al. (2013)	20 (1/tooth)
Development stage	Categorical variable (11 stages)	Development stage of the tooth, reported in Irurita et al. (2014)	20 (1/tooth)
Eruption phase	Categorical variable (3 stages)	Eruption phase of the tooth, reported in Irurita et al. (2014)	20 (1/tooth)
Total number of features in the dataset			63
Total number of instances in the dataset (one per individual)			114

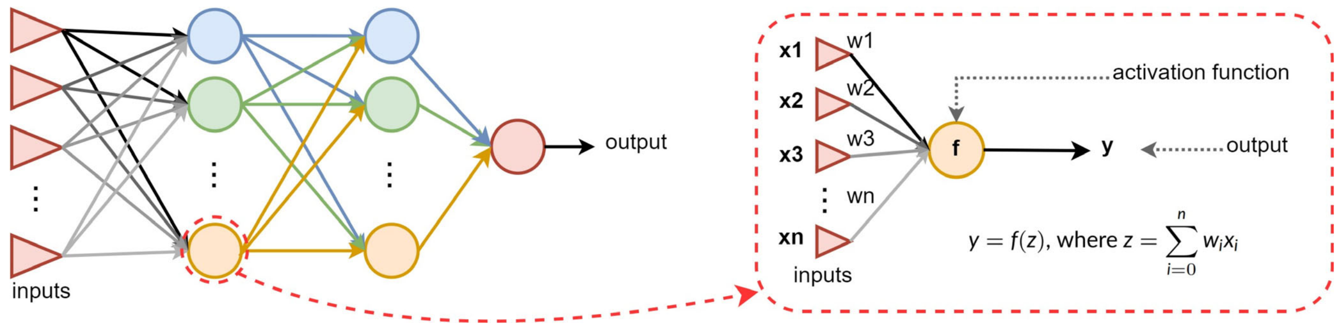

FIGURE 2 Structure of an MLP with N inputs, two hidden layers of M neurons each, and an output (left). Operations that are conducted in each neuron (right). The inputs (x_i) refer to the features of the sample (N in total); the output (y) corresponds to the prediction if it is the output layer, or just the output of the neuron if it is not; the weights of the neuron are represented by w_i. Note that $x_0 = 1$, and w_0 is the bias. The activation function (f) may vary from case to case.

such as a simple imputer which replaces missing values using the mean along each feature, an iterative imputer which employs the k-Nearest Neighbor algorithm, and an iterative imputer using the Random Forest Regressor (Breiman, 2001). The results of this preliminary study demonstrated that the latter alternative performed better. The Random Forest Regressor estimator fits a series of decision trees to multiple subsamples of the dataset and then uses the average to improve accuracy and control over-fitting.[1]

In order to design a method for estimating age that considers both tooth lengths and developmental stages, we used a widely known artificial neural network (Bishop, 1995; Goodfellow et al., 2016) called a Multilayer Perceptron (MLP) (Taud & Mas, 2018). The MLP consists of multiple layers of nodes, with each node being a simple computational unit or neuron. Each neuron performs a weighted sum of its inputs and applies a non-linear activation function to produce an output. The basic structure of an MLP includes an input layer, one or more hidden layers, and an output layer. Each layer is composed of a set of neurons that are connected to other neurons in adjacent layers (see Figure 2). These connections are characterized by weights, which are learned during training to produce the desired output for a given input. Weights are adjusted using backpropagation, which involves computing the gradients of a loss function and using them to update the weights via gradient descent.

MLPs can learn complex non-linear relationships between inputs and outputs and can be used to solve a wide range of problems in fields such as computer vision, natural language processing and speech recognition. As mentioned in the Section 1, the idea is to use as much information as possible about the problem under discussion, that is, the maturation stages, emergence stages and maximum lengths of the available teeth of an individual. An MLP can thus address this problem by using the selected features as input to produce an estimate of age as output, learning the relationship between the teeth' characteristics and the individuals' age.

In this study, several experiments were carried out to build an accurate model. Some of these experiments were aimed at determining the best number of hidden layers and neurons that make up our MLP. However, due to the scarcity of the data and to avoid overfitting, the initial number of hidden layers was decided to be two, each with 100 neurons. On the other hand, the purpose of our model is to estimate the age of an individual using the information about their deciduous teeth and sex. Therefore, the input layer of the model would consist of as many features as were selected in the ongoing experiment, and the output would be a single value corresponding to the estimated age in days. For instance, if the experiment is performed using all the information available about the subject, there will be 61 features as input.

The validation of the model in each experiment is done employing the leave-one-out technique. The model is trained from scratch as many times as there are instances in the dataset. At each step, the training set consists of all the instances except one, and the test set comprises this remaining sample. This process is repeated until the model is tested on all the instances, so an estimate is obtained for each one. The reason for applying this methodology is the desire to use as many samples as possible to train the model, taking into account the paucity of data. Note that at each step, the model is trained from scratch, so the test sample is not observed before. Once the estimates are obtained for each instance of the sample, the results are assessed using the Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y} - y)^2}{N}}. \tag{1}$$

The estimated ages (collected from each leave-one-out step) are expressed as $\hat{y}$ and the actual ages are expressed as y, while N is the total number of estimates. The RMSE provides a measure of how much the predicted values differ from the actual values on average. The smaller the value of RMSE, the better the predictive model's performance. In this case, the RMSE will express the difference in days between the estimated and actual ages on average. Additionally, we employed the Friedman and Nemenyi-Friedman tests for model validation, along with various metrics, including percentiles (5%, 95%, and 99%), mean absolute error (MAE), and standard deviation (SD) of the test residuals. The residuals are the difference in days between the estimated age and the actual age of an individual.

Finally, a preliminary study was performed, in which the pre-processing stage was extended by testing whether it was more appropriate to merge the subjects' hemi-arches or not. Hemi-arches are half of the dental arch in either the upper or lower jaw. The right hemi-arch includes the teeth on the right side, while the left hemi-arch includes the teeth on the left side. Merging hemi-arches refers to the process of combining the separate halves of the dental arch, specifically the right and left hemi-arches. Some contributions such as (Irurita et al., 2014) and (Irurita et al., 2013) encountered no significant differences between both hemi-arches. According to the results obtained, there was insufficient evidence to conclude that the differences between merging and not merging hemi-arches were significant. Therefore, the decision to merge the hemi-arches was made since the number of features involved is lower (half of them). This simplifies the model, as well as the training time and the work of the practitioner. In addition, the number of empty features in the data is reduced. On the other hand, as part of the pre-processing, dimensionality reduction techniques were applied to assess whether they would lead to an improvement in the accuracy of the model. The techniques used were Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA). The former showed better results.

Moreover, hyperparameter tuning was performed in the experiments on both the imputer and the model in order to obtain accurate results. Hyperparameters are the settings and configurations that are set by the programmer or user, rather than learned from the data during training. For instance, in neural networks, these settings include items such as the number of layers in the network, the size of each layer, the activation functions used, the learning rate for the optimization algorithm, and more. In this study, the hyperparameters are tuned by setting different values for some of them and evaluating the performance of the model on the validation set. The interested reader is referred to Table A1, Appendix A, for more details on the configuration of the models.

3 | EXPERIMENTS AND RESULTS

This section summarizes the main results and insights obtained to answer the three questions raised in Section 1.

3.1 | Information fusion: Lengths and stages

Firstly, the initial question raised is addressed: *Is it more informative to use only teeth lengths, emergence stages, maturation stages, or a combination of them?* In *MLP*, the model takes into account both the length and the stages of the teeth to estimate the age. In this experiment, the PCA dimensionality reduction technique was used over the lengths. Only the maximum lengths and only the stages of the teeth were used for both training and testing the corresponding model in experiments *MLP_lengths* and *MLP_stages*, respectively. Note that before the *MLP* experiment, another study was carried out that also involved both the length and the stages of the teeth, but did not use the PCA technique. The RMSE values obtained for the training (58.37) and test (77.74) phases were lower than those of *MLP_lengths* and *MLP_stages*, but higher than those of *MLP*. Therefore, for the sake of simplicity, this previous experiment has been excluded from the comparison below.

Table 2 contains the train and test RMSE values and the 5%, 95% and 99% percentiles, MAE, and SD of the residuals from the test phase in each experiment. It is observed that the best results were obtained in *MLP*, while the worst ones were found in *MLP_lengths*, where only teeth lengths were involved. This means that working with lengths only does not provide much information. This was also the case for the experiment *MLP_stages*, which does not stand out regarding the different metrics employing only the maturation and emergence stages. Hence, the combination of lengths and stages proved to be more useful and informative than either lengths only or emergence and maturation stages only.

Although the RMSE values and the rest of the metrics are useful for assessing and comparing the implemented models, statistical tests were used to make a more detailed comparison. The results from both the Friedman and the Nemenyi-Friedman tests are summarized in Table 3. The p-value calculated by the Friedman test indicates that

TABLE 2 Results obtained from experiments *MLP*, *MLP_lengths*, and *MLP_stages*.

	RMSE		Percentiles			Statistical test measures	
Experiment	Train	Test	5%	95%	99%	MAE	SD
MLP	**56.83**	**65.63**	**1.8**	**128.03**	**207.96**	**41.32**	**65.63**
MLP_lengths	74.66	95.68	3.17	164.25	477.72	53.1	95.58
MLP_stages	70.6	84.8	2.68	172.62	235.17	59.31	84.43

Note: All values are reported in days as the unit. The lowest value of each column is in bold.

TABLE 3 Results from the Friedman and the Nemenyi-Friedman tests for experiments *MLP*, *MLP_lengths*, and *MLP_stages*.

Friedman test		Statistic = 34.111	p-Value = 3.916e-08
Nemenyi Friedman test			
	MLP	MLP_lengths	MLP_stages
MLP	1.000	**0.001**	**0.001**
MLP_lengths	**0.001**	1.000	0.597
MLP_stages	**0.001**	0.597	1.000

Note: The *p*-values lower than α (0.05) are in bold.

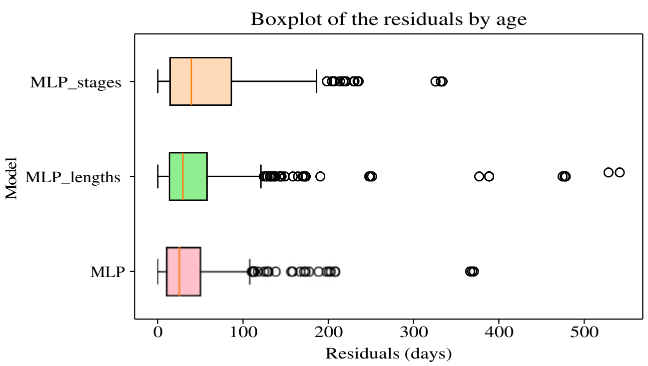

FIGURE 3 Comparison of the test residuals of the experiments *MLP*, *MLP_lengths*, and *MLP_stages*.

there are significant differences between some of the results obtained by the models. According to the p-value of the Nemenyi- Friedman test, the differences between *MLP* and *MLP_lengths* are significant, as well as those between *MLP* and *MLP_stages*. Nonetheless, it can be observed that there is insufficient evidence to infer that there are significant differences between *MLP_lengths* and *MLP_stages*. It can thus be concluded that the model used in *MLP* is the most promising. Figure 3 shows the test residuals of the three experiments.

3.2 | Information fusion: Single tooth versus combination of teeth

In this section, the experiments aim at answering the second question posed: *Is it better to use a single tooth to estimate age or a combination of teeth?* Hence, the best-performing model of the previous experimental study (*MLP*), which takes into account the combination of all the available teeth of an individual, was compared to a method that uses only a tooth in order to perform the age estimation task: a traditional analysis method. The latter consists in an exponential regression method, which models the relationship between a dependent variable and an independent variable that exhibits exponential growth or decay. This model type is really common in this field since the relationship between the chronological age and the maximum length of teeth is exponential (Irurita et al., 2014). As previously mentioned, these methods typically focus on examining a particular tooth, considering only one of its features.

The traditional method was trained and tested using only real data, that is, no imputation techniques were applied. The left and right hemi-arches were also merged, which is the procedure that actual practitioners would follow (Irurita et al., 2013, 2014). Note that the names of the teeth of the right hemi-arch were used to refer to them despite both sides having been merged. In order to work with a single tooth and one of its features, the maximum length of tooth 83 was chosen. The reason for this was that in (Irurita et al., 2013), authors reported that the canines were the teeth with a higher value of R^2. This variable represents the proportion of variation in the dependent variable (age) that can be explained by the independent variable (tooth length) included in the model. A higher R^2 value implies a better fit of the model to the data. In addition, there were more instances in the dataset with a value of the length of the lower canine (tooth 83), 63 in total, than that of the upper canine (tooth 53), 56 in total. This leads to training and testing phases with more examples. Figure 4 plots the distribution of the actual ages of the instances where tooth 83 was present (left) and the fit of the implemented traditional method to these instances (right).

Table 4 summarizes the RMSE values and the test residuals' percentiles (5%, 95% and 99%), MAE, and SD of the experiments *MLP* and *traditional_method*. A noticeable difference exists between the values of each case, providing evidence that the method proposed in this study is more accurate. Consequently, the second of the raised questions is thus answered: a method using a combination of multiple teeth turns out to be better at estimating the age of infants than a traditional analysis method that relies on a single tooth. Figure 5 shows the test residuals of these experiments.

3.3 | ML-based method vs. traditional method

Finally, in order to answer the third raised question, the results of the experiments *MLP* and *traditional_method* are thoroughly compared: *Does the proposed ML-based method outperform a traditional one to*

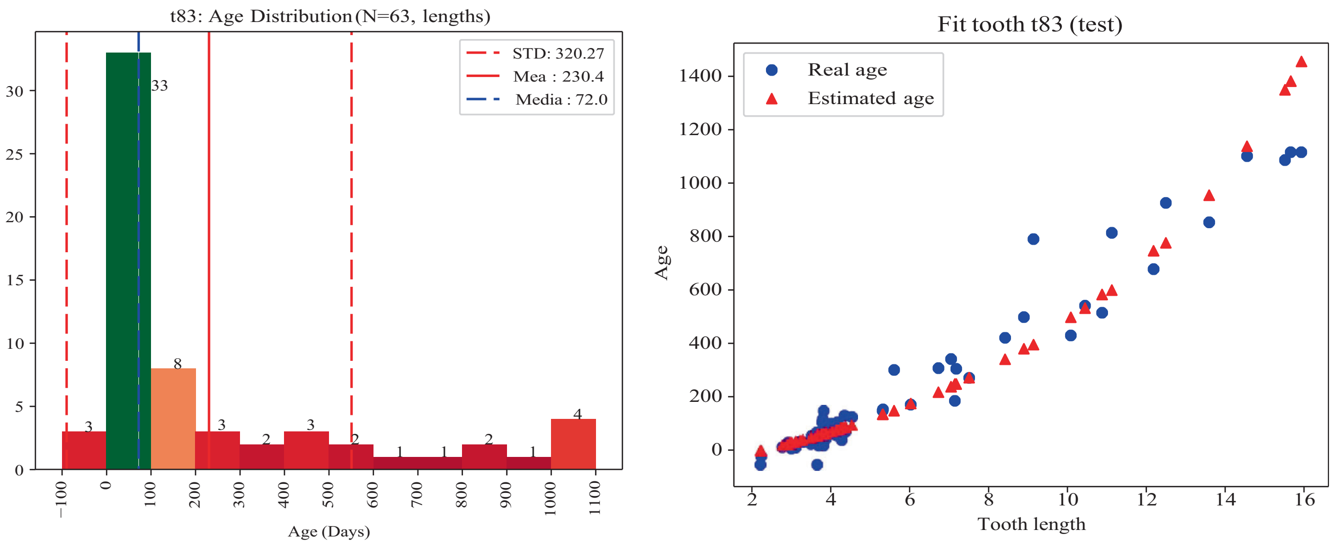

FIGURE 4 Distribution of actual age (in days) for tooth 83 (the hemi-arches are merged) on the left. Fit of the traditional analysis method (named *traditional_method*) using the maximum length of this tooth on the right.

TABLE 4 Results obtained from experiments *MLP* and *traditional_method*.

Model		RMSE		Percentiles			Statistical test measures	
Experiment	Tooth	Train	Test	5%	95%	99%	MAE	SD
MLP	All teeth	**56.83**	**65.63**	**1.8**	**128.03**	**207.96**	**41.32**	**65.63**
Traditional_method	83 (N = 63)	93.22	101.02	2.35	270.9	369.59	58.25	101.0

Note: All values are reported in days as the unit. The minimum values of each column are in bold.

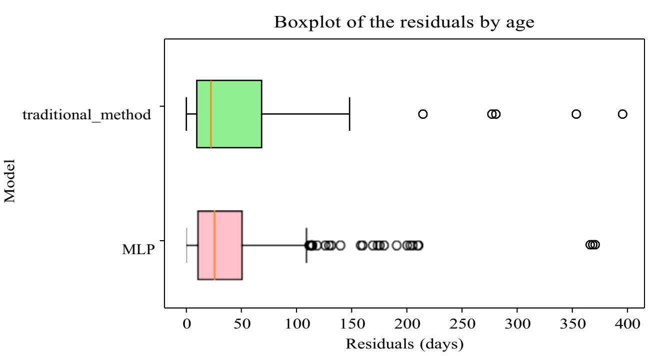

FIGURE 5 Comparison of the test residuals of the experiments *MLP* and *traditional_method*.

estimate the age of an individual? For that purpose, the bootstrap technique (Hesterberg, 2011) was used. Bootstrapping is a statistical technique used to estimate the accuracy of a statistical estimator by resampling the original data. The basic idea is to create many samples of the original data set by drawing samples from it with replacement. The estimator is then calculated for each of these bootstrap samples, and the distribution of these estimates is used to infer the accuracy of the original estimator. In this study, the estimator was the RMSE and the number of samples drawn was 10,000. In addition, the test residuals' MAE, SD, and percentiles (5%, 95% and 99%) were also calculated.

As shown in Table 5, the model of the experiment *MLP* was the one performing better, reaching lower error values. Therefore, the third question is answered: an ML-based method turns out to be more accurate in the age estimation task in infants than a traditional analysis method. Moreover, the answer to the first question is strengthened: employing a combination of lengths and emergence and maturation stages, as the ML-based method does, implies better results than involving only one of these features, as performed by the traditional method.

Furthermore, another advantage of our proposal is its straightforward use. When estimating the age of an individual, the practitioner just needs to provide information related to the length, and the maturation and alveolar stages of available teeth in the right hemi-arch, along with the individual's sex. If data is present for both hemi-arches, the practitioner can choose either one or input the average. In cases where data is exclusively available for the left hemi-arch, the user should input that information. If certain features are not available, "NA" should be entered. Note that our model requires a minimum of three features to function properly; otherwise, it will not work due to scarce data. Once the required information is provided, the software automatically imputes missing data, pre-processes it, and estimates the age in days. To perform this, we saved checkpoints of the imputer,

TABLE 5 Bootstrap results from experiments *MLP* and *traditional_method* (tooth 83).

Bootstrap	Mean results						
Experiment	No.	RMSE	5%	95%	99%	MAE	SD
MLP	10,000	**65.36**	**1.96**	**131.86**	**249.44**	**41.34**	**65.27**
Traditional_method (tooth 83)	10,000	99.74	2.67	241.01	343.55	58.44	98.93

Note: All values except the number of samples are reported in days as the unit.

the PCA model, and the MLP model after a posterior training stage that was performed with the whole dataset. When a new estimation is to be made, the three checkpoints are loaded and used directly, without the need for a training stage relying on the data. It is worth mentioning that there is no final application available for users since the software is currently in an experimental phase.

4 | DISCUSSION AND CONCLUSION

Age estimation in Physical Anthropology is currently an area of considerable research interest, with a strong emphasis on enhancing the accuracy and robustness of traditional methods through the introduction of innovative approaches (like AI-based methods). Within this context, our main goal is to develop precise, fast, reliable, and automated techniques focused on age estimation in infants using deciduous teeth. Our findings have demonstrated the value of fusing information from various sources when estimating the age of an individual. Experiments *MLP_lengths*, where only teeth lengths were involved, and *MLP_stages*, where only maturation and emergence stages were used, yielded a worse performance in terms of RMSE. The obtained RMSE values were 95.68 days and 84.8 days, respectively. These errors were higher than the recorded in the experiment *MLP* (65.63 days), which considered both lengths and stages. On the other hand, it was also investigated whether it is preferable to use a combination of several teeth rather than a single one to perform the age estimation task. This resulted in an RMSE of 101.02 days (*traditional_method*) using a single tooth, which is significantly worse than the experiment *MLP* using multiple teeth. Finally, the proposed ML-based method was compared to a traditional one, showing that our model performed better in the age estimation task. Moreover, it is worth noting that, thanks to the application of missing data imputation techniques, our method can address the age estimation task despite the existence of missing values in the dataset, i.e. when there is no information from all the teeth of an individual. This constitutes an advantage with respect to a traditional analysis method, which needs information from a specific tooth to perform the task.

Furthermore, these results also stand out significantly even when compared to other methods frequently employed for infant age estimation. Notable examples within similar age ranges include: utilizing the diaphysis length of long bones, which yields an RMSE of 84 days (Cardoso et al., 2014); employing the ilium and the scapula, resulting in RMSE values of 76 and 106 days, respectively (Figueiro et al., 2022); and considering the frontal bone, which leads to an RMSE of 73 days (Smith et al., 2021). It is crucial to stress the significance of this comparison, bearing in mind that the last RMSE values quoted correspond specifically to the training stage and not the validation stage as in this study. These results clearly demonstrate the potential of the techniques employed in the present contribution. Moreover, it is worth noting that even better estimates could be achieved by combining the information from the mentioned variables, rather than solely selecting only one of them as the most suitable for age estimation. Furthermore, alternative methods based on the study of radiographs also present the opportunity to combine the information provided by different teeth for the task of age estimation (Cameriere et al., 2016; Demirjian et al., 1973). However, a major limitation of these techniques is that, unlike our method, they cannot be used effectively when a tooth is missing.

To explore future directions and potential advances in the field, several areas warrant further investigation. Our method could be compared to other models based on different ML techniques, such as Random Forest or Support Vector Machine, or even other neural networks or ensemble methods. By evaluating the performance of our method against these alternative approaches, we can gain deeper insights into its strengths and limitations. Nonetheless, it would require a larger sample, so that the complexity of the models could be increased and further validation of the performance of our method could be addressed. Ultimately, we have not implemented a final version of our application software for public use yet. Hence, we aim to address this in future works. We also aim to ensure its accessibility for interested practitioners by integrating it into forensic identification software such as Skeleton-ID.[2]

In summary, the results reported in this article highlight the clear advantages of utilizing ML-based methods over traditional ones: (1) ML-based methods enable the integration of qualitative and quantitative variables from multiple teeth, enhancing the accuracy of age estimation; (2) these methods result in a reduction of the error assumed; (3) ML-based methods offer the capability to perform estimations even in cases where teeth are missing, ensuring robustness and applicability in various scenarios.

AUTHOR CONTRIBUTIONS

Práxedes Martínez-Moreno: Conceptualization (lead); formal analysis (lead); investigation (lead); methodology (lead); software (lead); validation (lead); visualization (lead); writing – original draft (lead). **Andrea Valsecchi:** Conceptualization (equal); formal analysis (equal); writing – review and editing (equal). **Sergio Damas:** Conceptualization (equal); supervision (equal); writing – review and editing (equal). **Javier Irurita Olivares:** Conceptualization (equal); data curation (lead); resources (lead); supervision (equal); writing – review and editing

(equal). **Pablo Mesejo:** Conceptualization (equal); funding acquisition (lead); project administration (lead); supervision (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

The authors thank D. José Antonio Muñóz, Managing Director, Maribel Martín, Service Coordinator, and all EMUCESA staff at the San José cemetery in Granada for their invaluable assistance and are grateful to the Magistrate Judge (Court of First Instance no. 5), responsible for the Registry Office of Granada. This work was supported by grant CONFIA (PID2021-122916NB-I00) funded by MCIU/AEI/10.13039/501100011033 and by grant FORAGE (B-TIC-456-UGR20) funded by Consejería de Universidad, Investigación e Innovación, both funded by "ERDF A way of making Europe"; and by Red.es under grant Skeleton-ID2.0 (2021/C005/00141299). Funding for open access charge: Universidad de Granada/CBUA.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

ORCID

Práxedes Martínez-Moreno https://orcid.org/0000-0002-4091-3547

Javier Irurita https://orcid.org/0000-0003-1676-9773

ENDNOTES

[1] Over-fitting is a problem that typically occurs with complex models that are designed with insufficient data. The model's design is based on a training phase with just a few samples, where the model's performance is often outstanding or even extraordinary. However, the model's estimates on new data are significantly worse or even really poor, that is, the models do not generalize (Bishop, 1995).

[2] https://skeleton-id.com.

REFERENCES

Alemán, I., Irurita, J., Valencia, A. R., Martínez, A., López-Lázaro, S., Viciano, J., & Botella, M. C. (2012). Brief communication: The granada osteological collection of identified infants and young children. *American Journal of Physical Anthropology*, *149*(4), 606–610.

AlQahtani, S., Hector, M., & Liversidge, H. (2010). Brief communication: The london atlas of human tooth development and eruption. *American Journal of Physical Anthropology*, *142*(3), 481–490.

Bishop, C. (1995). *Neural networks for pattern recognition*. Oxford University Press.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Buk, Z., Kordik, P., Bruzek, J., Schmitt, A., & Snorek, M. (2012). The age at death assessment in a multi-ethnic sample of pelvic bones using nature-inspired data mining methods. *Forensic Science International*, *220*(1–3), 294.

Cameriere, R., Pacifici, A., Pacifici, L., Polimeni, A., Federici, F., Cingolani, M., & Ferrante, L. (2016). Age estimation in children by measurement of open apices in teeth with bayesian calibration approach. *Forensic Science International*, *258*, 50–54.

Cardoso, H. F., Abrantes, J., & Humphrey, L. T. (2014). Age estimation of immature human skeletal remains from the diaphyseal length of the long bones in the postnatal period. *International Journal of Legal Medicine*, *128*, 809–824.

Cardoso, H. F., Meyers, J., & Liversidge, H. M. (2019). A reappraisal of developing deciduous tooth length as an estimate of age in human immature skeletal remains. *Journal of Forensic Sciences*, *64*(2), 385–392.

Demirjian, A., Goldstein, H., & Tanner, J. M. (1973). A new system of dental age assessment. *Human Biology*, *45*(2), 211–227.

Figueiro, G., Irurita Olivares, J., & Alemán Aguilera, I. (2022). Age estimation in infant skeletal remains by measurements of the pars lateralis. *International Journal of Legal Medicine*, *136*(6), 1675–1684.

Gámez-Granados, J. C., Irurita, J., Pérez, R., González, A., Damas, S., Alemán, I., & Cordón, O. (2022). Automating the decision making process of todd's age estimation method from the pubic symphysis with explainable machine learning. *Information Sciences*, *612*, 514–535.

Goodfellow, I. J., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hesterberg, T. (2011). Bootstrap. *WIREs Computational Statistics*, *3*(6), 497–526.

Irurita, J., Alemán, I., López-Lázaro, S., Viciano, J., & Botella, M. C. (2014). Chronology of the development of the deciduous dentition in mediterranean population. *Forensic Science International*, *240*, 95–103.

Irurita, J., Aleman, I., Viciano, J., Luca, S., & Botella, M. (2013). Evaluation of the maximum length of deciduous teeth for estimation of the age of infants and young children: Proposal of new regression formulas. *International Journal of Legal Medicine*, *128*, 345–352.

Liversidge, H. M. (2008). Dental age revisited. In *Technique and application in dental anthropology* (pp. 234–252). Cambridge University Press.

Liversidge, H. M. (2015). Controversies in age estimation from developing teeth. *Annals of Human Biology*, *42*(4), 397–406.

Liversidge, H. M., Dean, M. C., & Molleson, T. I. (1993). Increasing human tooth length between birth and 5.4 years. *American Journal of Physical Anthropology*, *90*(3), 307–313.

Minier, M., Maret, D., Dedouit, F., Vergnault, M., Mokrane, F. Z., Rousseau, H., … Rougé, D. (2013). Fetal age estimation using msct scans of deciduous tooth germs. *International Journal of Legal Medicine*, *128*, 177–182.

Mohamed, E. G., Redondo, R. P. D., Koura, A., EL-Mofty, M. S., & Kayed, M. (2023). Dental age estimation using deep learning: A comparative survey. *Computation*, *11*(2).

Moorrees, C. F. A., Fanning, E. A., & Hunt, E. E., Jr. (1963). Formation and resorption of three deciduous teeth in children. *American Journal of Physical Anthropology*, *21*(2), 205–213.

Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT Press.

Petrone, S., Garizoain, G., García Mancuso, R., & Inda, A. M. (2019). Maximum length of deciduous dentition as an indicator of age during the first year of life: Methodological validation in a contemporary osteological collection. *Forensic Science International*, *303*, 109928.

Roberts, G., Lucas, V. S., Camilleri, S., Jayaraman, J., Kasper, K. A., & Lewis, J. M. (2023). Questions of logic in atlas methods of dental age estimation. *Journal of Forensic and Legal Medicine*, *96*, 102505.

Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Schour, I., & Massler, M. (1941). *Development of human dentition chart*. American Dental Association.

Sgheiza, V., & Liversidge, H. (2023). Reference and target sample age distribution impacts between model types in dental developmental age estimation. *International Journal of Legal Medicine*, *137*, 383–393.

Smith, D. E., Humphrey, L. T., & Cardoso, H. F. (2021). Age estimation of immature human skeletal remains from mandibular and cranial bone dimensions in the postnatal period. *Forensic Science International*, *327*, 110943.

Štepanovský, M., Ibrová, A., Buk, Z., & Velemínská, J. (2017). Novel age estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods. *Forensic Science International*, *279*, 72–82.

Taud, H., & Mas, J. (2018). Multilayer perceptron (mlp). In *Geomatic approaches for modeling land change scenarios* (pp. 451–455). Springer International Publishing.

Ubelaker, D., Cordero, Q., & Linton, N. (2020). Recent research in forensic anthropology. *European Journal of Anatomy*, *24*(3), 221–227.

Ubelaker, D. H. (1989). *Human skeletal remains. Excavation, analysis, interpretation*. Taraxacum.

Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC Press.

Velemínská, J., Pilnỳ, A., Čepek, M., Kot'ová, M., & Kubelková, R. (2013). Dental age estimation and different predictive ability of various tooth types in the czech population: Data mining methods. *Anthropologischer Anzeiger*, *70*(3), 331–345.

Viciano, J., De Luca, S., Irurita, J., & Alemán, I. (2018). Age estimation of infants through metric analysis of developing anterior deciduous teeth. *Journal of Forensic Sciences*, *63*(1), 20–30.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Martínez-Moreno, P., Valsecchi, A., Damas, S., Irurita, J., & Mesejo, P. (2024). Information fusion for infant age estimation from deciduous teeth using machine learning. *American Journal of Biological Anthropology*, e24912. https://doi.org/10.1002/ajpa.24912