XAS: Automatic yet eXplainable Age and Sex determination by combining imprecise per-tooth predictions☆

Nicolás Vila-Blanco [a,d,1], Paulina Varas-Quintana [c,1], Ángela Aneiros-Ardao [c,1], Inmaculada Tomás [a,c,d,*,2], María J. Carreira [a,b,d,*,2]

[a] *Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Spain*
[b] *Departamento de Electrónica e Computación, Escola Técnica Superior de Enxeñaría, Universidade de Santiago de Compostela, Spain*
[c] *Oral Sciences Research Group, Special Needs Unit, Department of Surgery and Medical Surgical Specialities, School of Medicine and Dentistry, Universidade de Santiago de Compostela, Spain*
[d] *Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Spain*

ARTICLE INFO

ABSTRACT

Chronological age and biological sex estimation are two key tasks in a variety of procedures, including human identification and migration control. Issues such as these have led to the development of both semiautomatic and automatic prediction models, but the former are expensive in terms of time and human resources, while the latter lack the interpretability required to be applicable in real-life scenarios. This paper therefore proposes a new, fully automatic methodology for the estimation of age and sex. This first applies a tooth detection by means of a modified CNN with the objective of extracting the oriented bounding boxes of each tooth. Then, it feeds the image features inside the tooth boxes into a second CNN module designed to produce per-tooth age and sex probability distributions. The method then adopts an uncertainty-aware policy to aggregate these estimated distributions. Our approach yielded a lower mean absolute error than any other previously described, at 0.97 years. The accuracy of the sex classification was 91.82%, confirming the suitability of the teeth for this purpose. The proposed model also allows analyses of age and sex estimations on every tooth, enabling experts to identify the most relevant for each task or population cohort or to detect potential developmental problems. In conclusion, the performance of the method in both age and sex predictions is excellent and has a high degree of interpretability, making it suitable for use in a wide range of application scenarios.

1. Introduction

Along with race, chronological age and biological sex are the most important human features examined in anthropological and forensic studies [1]. In archaeology, estimations of these features are fundamental to not only reconstructing the demographics of earlier populations, but also their way of life and environment. In forensic human identification, meanwhile, information of an individual's age and sex is crucial to simplifying further enquiries, such as time-consuming searches of DNA databases, leading to a notable increase in the speed at which this data can be obtained.

Chronological age predictions are also used daily in legal proceedings where the birth date of the subject involved cannot be verified due to either the absence or the limited reliability of their documentation. This issue may arise in relation to migration-control policies or in trials involving individuals without any documentation, as the legislation in most countries makes the identification of legal age a crucial task. It is also important in adoption proceedings concerning undocumented children. Indeed, in all these cases, a somatic maturity examination carried out by an expert is required.

The developmental status of bones has been employed successfully in estimations of both age and sex, and is the only available

* Corresponding authors.
E-mail addresses: nicolas.vila@usc.es (N. Vila-Blanco), paulina.varas@rai.usc.es (P. Varas-Quintana), a.ortodoncia@gmail.com (Á. Aneiros-Ardao), inmaculada.tomas@usc.es (I. Tomás), mariajose.carreira@usc.es (M.J. Carreira).
[1] Researcher.
[2] Co-ordinator.

methodology in examinations of archaeological remains or corpses that have sustained significant damage. Many skeletal elements are used to these ends, including the pubic symphysis, auricular surface and sternal ribs for age estimation [2], and the hip bone [3], the bones of the hands and feet [4] and the scapula [5] to determine sex. Nonetheless, it should be noted that there is no current methodology based on bone development that outperforms any others systematically, as the outcomes of each of them depend on a wide variety of factors. For instance, there are specific age prediction methods for subadults and others that work better in adults [6]. Similarly, sex estimation works better in adults, because dimorphic skeletal indicators do not develop until after puberty [6].

The body part most in use in the field of age prediction field is the dentition. The composition of the teeth (and, specifically, the enamel) makes them the most resistant part of the skeleton. This enables them to resist extreme conditions like very high temperatures, meaning they can be preserved for thousands of years. Although tooth development is guided by both genetic and environmental factors, dental mineralisation is reportedly less affected than bone mineralisation, causing it to be regarded as a better indicator of chronological age [7].

Although the potential of the dentition to estimate biological sex is more limited than that of other skeletal parts, the teeth are still used in cases where they are the only available remains or as a complement to other methods [8]. Specifically, the canines are reported to be the most dimorphic [9], although other teeth also show significant differences [10]. In that respect, the mandible bone has been proven to have a greater dimorphism than any individual tooth [11,12].

X-ray-based imaging techniques took the field a step forward, since they allowed clinicians to assess bone development in situ, using less invasive and faster procedures, and thus enabled them to perform chronological age estimation. In this regard, age prediction from dental radiographic images involves the evaluation of a variety of indicators, as reported in [13]. This includes the formation of the jaw bones and some characteristics of dental evolution, for example, the degree of crown and root completion, the measurement of open apices, or the third molar development and topography of the third molar.

It is worth noting that dental panoramic radiographs (orthopantomograms or OPGs) are the least invasive radiological technique for estimating age and sex, as only a single image is required to capture the complete dentition. Other bone structures are also valuable for additional examinations that may be required, including the mandible, the nasal fossa, and the vertebrae.

This paper proposes a new method for predicting age and sex. This relies on deep learning techniques to detect the bounding box of each tooth, with its orientation taken into account. Then, the identified tooth features are reused to estimate the age and sex probability distributions for each tooth. Finally, an uncertainty-aware policy is employed to aggregate these distributions.

2. Related work

Automatic methods for tooth identification using dental images have been approached in various ways. The first ones were aimed at human identification using bitewing images, and conducted tooth detection via the recognition of gaps between the lower and upper jaws and between consecutive teeth, normally by calculating integral projections [14]. However, the most successful methodologies in recent years have used OPGs in methods based on deep learning, more specifically, Convolutional Neural Networks (CNNs). In this regard, the available research has approached the problem from three different angles, as seen in Table 1: (1) localisation — estimation of tooth bounding boxes; (2) segmentation — extraction of tooth mask; and (3) classification — prediction of tooth type or numbering.

Oktay et al. [15] proposed the use of a simple AlexNet to perform both tooth localisation and classification, while successive studies relied on more complex architectures such as Faster R-CNN [18,21,26] or
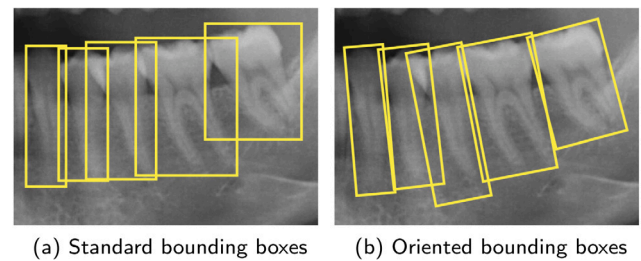

(a) Standard bounding boxes   (b) Oriented bounding boxes

Fig. 1. Comparison of standard (a) and oriented (b) bounding boxes. In the former, there is an obviously greater overlap between neighbouring teeth, which is particularly noticeable in the posterior teeth.

centre-based tooth detectors [27]. In relation to tooth segmentation, Silva et al. [16] and Koch et al. [19] initially tackled this by targeting the dental mask as a whole via the Mask R-CNN and U-Net architectures, respectively. This approach was refined by Jader et al. [17] and Lee et al. [24] to perform instance segmentation, i.e., obtaining the individual mask for each tooth.

The datasets used to validate these methodologies vary in size, with the largest containing 1574 OPGs [18]. However, the one used the most is that compiled by Silva et al. [16], which comprises 1500 images [16,17,19]. The research also vary regarding the ages of the population samples. Indeed, although all the works used OPGs of adult subjects with a fully developed dentition, only a small subgroup of studies included images from children with a deciduous or mixed dentition [16,17,19], and none of these conducted tooth numbering. Certainly, it seems remarkable that in the investigation by Silva et al. they even discarded images belonging to children, though these were available [21].

The studies available conducted tooth localisation using standard bounding boxes, i.e., with rectangles parallel to the x–y axes. This makes box representation simpler – only four coordinates are required – and also means it is easier to implement detection steps like the Region of Interest (ROI) Pooling operation in Faster R-CNN architectures. However, for tooth localisation purposes, this can cause a high degree of overlap between adjacent boxes if the contained teeth are significantly rotated, as shown in Fig. 1. Moreover, standard bounding boxes do not provide data on the tooth aspect ratio (width–height ratio) or its orientation, which is nonetheless valuable information in fields like paediatric dentistry.

More recently, a number of new, fully automatic methodologies to estimate age and sex from OPGs have been proposed (Table 2), most of which are based on deep learning techniques. The available studies on age estimation have different objectives: staging specific teeth [28–30]; numeric age regression [31–35]; age group classification [36]; and legal age classification [37]. The methodologies developed are based on two different principles. In some research, the authors proposed restricting age estimations to image crops or segmentations belonging to a specific tooth or group of teeth [22,28–31]. In these scenarios, the step for detecting these teeth was carried out manually [22,28,29], semiautomatically [31], or automatically [30]. Other studies proposed one-step pipelines to estimate chronological age directly from a raw image, with no intermediate stages [32–35,37]. Regardless of the method adopted, all these authors agreed that age estimation from OPGs is only feasible during the tooth development period.

Automatic sex estimation methodologies using OPGs are rarer, and all of them adopt the same approach: an end-to-end CNN to make predictions directly from images [38–41]. Although the obtained accuracy varied in these studies, it remained above 90% for subjects older than 16. The performance was reportedly worse when younger subjects were included in the samples, especially for those younger than 8 [41]. The almost complete absence of dimorphic skeletal indicators before puberty, as well as the significant tooth development activity at that

Table 1
Main deep learning methods for tooth detection/segmentation in OPG. Loc.: localisation; Seg.: segmentation; Class.: classification; T: tooth type; I: individual tooth; Dent.: dentition; C: children; A: adult.

Work	Loc.	Seg.	Class.		Dent.		Method	#Images
			T	I	C	A		
Oktay [15]	✓	✗	✓	✗	✗	✓	Modified AlexNet	100
Silva [16]	✗	✓	✗	✗	✓	✓	Mask R-CNN	1,500
Jader [17]	✓	✓	✗	✗	✓	✓	Mask R-CNN	1,500
Tuzoff [18]	✓	✗	✓	✓	✗	✓	Faster R-CNN + VGG-16	1,574
Koch [19]	✗	✓	✗	✗	✓	✓	U-Net	1,500
Leite [20]	✓	✓	✗	✗	✗	✓	Deeplab-v3 + FCN-ResNet101	153
Silva [21]	✓	✓	✓	✓	✗	✓	PANet	543
Kim [22]	✓	✗	✓	✗	✗	✓	Faster R-CNN + Heuristics	303
Muramatsu [23]	✓	✗	✓	✗	✗	✓	DetectNet + ResNet50	100
Lee [24]	✓	✓	✗	✗	✗	✓	Mask R-CNN	50
Muresan [25]	✓	✓	✗	✗	✗	✓	ERFNet + Histogram classification	1,000
Mahdi [26]	✓	✗	✓	✓	✗	✓	Faster R-CNN + Heuristics	1,000
Chung [27]	✓	✗	✓	✓	✗	✓	Center detector + Bbox regressor	818
Ours	✓[a]	✗	✓	✓	✓	✓	Rotated R-CNN	**1,746**

[a] The result is an oriented bounding box.

Table 2
Main automatic methods for age/sex estimation from OPG. DSC: development stage classification; NA: numeric age; AGC: age group classification; LAD: legal age determination; BSC: binary sex classification.

	Work	Method	Required teeth[a]	Age	Target
Age	De Tobel [28]	AlexNet[b]	38	–	DSC
	Čular [31]	Active Appearance Model+ Radial Basis Network	48	10–25	NA
	De Back [32]	Bayesian CNN	–	5–25	NA
	Merdietio [29]	DenseNet201[b]	38	–	DSC
	Banar [30]	YOLO-like CNN + U-Net+ DenseNet201	38	7–24	DSC
	Vila-Blanco [33]	Two-path CNN	–	5–89	NA
	Kim [36]	ResNet152[b]	18, 28, 38, 48	–	AGC
	Wallraff [34]	ResNet18	–	11–20	NA
	Hou [35]	PC-DARTS-based CNN	–	0–93	NA
	Guo [37]	SE-ResNet101	–	5–24	LAD
Sex	Ilić [38]	VGG-16	–	–	BSC
	Milošević [39]	VGG-16	–	19–85	BSC
	Ke [40]	Multi-Feature Fusion Model	–	16–70	BSC
	Vila-Blanco [41]	Two-path CNN	–	5–90	BSC

[a] FDI notation.
[b] This method requires an initial manual crop of the regions of interest.

age, hinder the capacity to identify sexually dimorphic features in images belonging to children and, as a consequence, the sex classification process.

There are noticeable differences in the proposed automatic age and sex prediction methodologies regarding their interpretability and potential value. Age prediction approaches to third molar staging are easy to understand, since the steps proposed are based on well-known scoring systems that rely on particular tooth characteristics [28]. However, these do require the presence of this tooth and are therefore only useful for estimating age in the period when the third molar develops. On the other hand, the methods for predicting the numeric age or binary sex directly from raw images are particularly valuable, because of the very accurate estimations they produce. Moreover, they do not require the presence of specific teeth and, as a consequence, are applicable in a wider range of age cohorts. Although the heatmaps provided in some of the studies reflect the attention behaviour of these methodologies [33,39], the models' interpretability nonetheless lags well behind that of their traditional manual and semiautomatic counterparts, ultimately hindering their potential applications.

3. Materials and methods

A dataset of 1746 OPGs collected with a direct digital panoramic unit (Orthophos Plus DS; Sirona USA, Charlotte, NC) was provided by the School of Medicine and Dentistry, Universidade de Santiago de Compostela (Spain) for use in this study. The image collection protocol was approved by the institution's ethics committee and followed the requirements of the 1964 Declaration of Helsinki and its later amendments [42]. This included the complete anonymisation of the images, with data on the chronological age (calculated as the days elapsed between the birth date and the image acquisition date) and sex of each subject being the only information retained. No exclusion criteria were set for the image collection process. In terms of the database demographics, the number of female and male images was almost the same for every age group, as seen in Table 3. To ensure a good balance between developing and fully developed dentitions, a similar quantity of subjects younger and older than 20 was included.

All the radiographs used in the study were greyscale JPEG images with a depth of 8 bits. The image width varied from 2,700 to 3,200 pixels, but all of them were 1552 pixels high. The image annotation procedure was carried out by two trained and previously calibrated experts using the Labelbox platform [43]. This involved tracing the oriented bounding box of each tooth, preserving its natural orientation, as shown in Fig. 1(b).

Due to the properties of an OPG's geometry, the maxillary teeth are shown as having a greater overlap than their mandibular counterparts, especially in the premolar region, while the maxillary sinuses can interfere with the visibility of the apices [44]. Moreover, the anterior teeth in both the maxilla and the mandible are usually affected by ghost shadows produced by the cervical spine [45]. Consequently, the experiments were conducted using the mandibular canines, premolars, and molars.

Fig. 2 shows the methodology used in this study to estimate age and sex completely automatically. This has two main elements: tooth
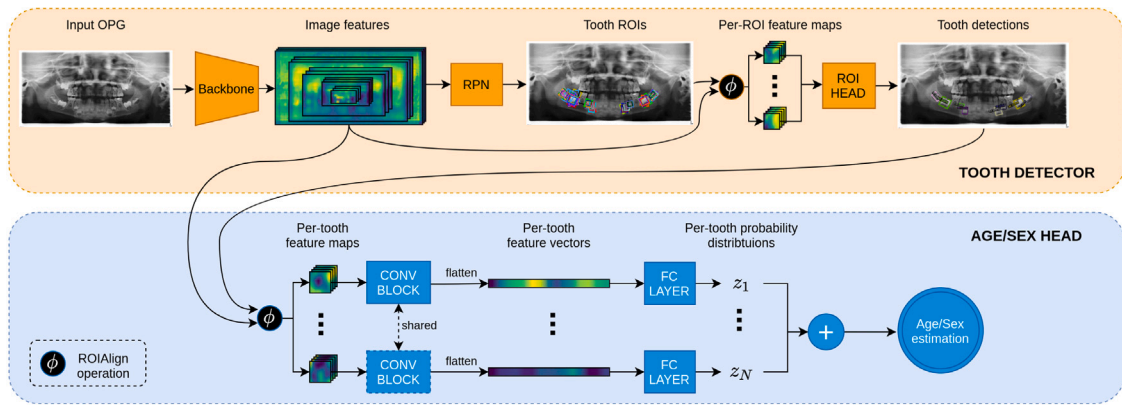
Fig. 2. Proposed approach for estimating age and sex: (1) the teeth are detected in the OPG with an R-CNN-based architecture. This involves the feature extraction process using an FPN backbone, the proposal of teeth locations (ROIs) using an RPN, and the refinement and classification of each tooth ROI using an ROI head; and (2) the per-tooth probability distributions of the age and sex are estimated. This process consists of evaluating the features extracted by the backbone at the estimated tooth locations, and applying a convolutional block and a fully connected layer to obtain the per-tooth distribution parameters. Finally, the estimated distributions are aggregated to calculate the age and sex.

Table 3
Dataset demographics.

Age group	Female	Male
(5, 10]	140	150
(10, 15]	150	150
(15, 20]	146	149
(20, 25]	108	104
(25, 30]	45	45
(30, 35]	45	45
(35, 40]	49	47
(40, 45]	49	48
(45, 50]	49	49
(50, 55]	47	47
(55, 60]	44	40
	872	874

detection and age and sex prediction. In the first step, the selected subset of teeth is detected in the image, with the position of each tooth obtained via oriented bounding boxes; its number is based on the FDI notation [46]. In the second step, the image information retrieved by the CNN at the tooth locations is fed into a second CNN to estimate the per-tooth chronological age and sex probability distributions, which are then aggregated to obtain the final prediction.

The detection network shown in the upper part of Fig. 2 is based on the Faster R-CNN architecture [47] and so is composed of three main modules: the feature extraction network, also called the backbone; the Region Proposal Network (RPN); and the ROI head. First, the CNN-based backbone is applied to extract a set of multiscale features from the input image. Then, those features are fed into an RPN, which classifies a grid of boxes (also known as anchors) according to the probability of it containing a tooth, with the most likely locations referred to as Regions of Interest (ROIs). Next, the learned features are fed into an ROI pooler, which produces equal-size feature maps for each ROI. Finally, the ROI head is applied to the per-ROI feature maps to refine the localisation of each ROI and determine whether it belongs to one of the tooth classes or the background class. This work includes some Faster R-CNN improvements proposed in a later study. Specifically, the Feature Pyramid Network (FPN) is used as the feature extractor [48] (on top of a ResNet50 network) and ROIAlign as the ROI pooler operation [49].

Once the teeth are localised and classified, the ROIAlign operation is performed again to evaluate the features extracted by the backbone in the final tooth positions, resulting in a per-tooth stack of 256 feature maps of size 7×7. These features are then processed to obtain the chronological age and sex, as shown at the bottom of Fig. 2. The age and sex estimation modules are referred to herein as the age head and sex head, respectively.

The architecture of the age head is first composed of a convolutional block shared by all tooth classes. This is used to extract per-tooth high-level features, which we expected would be associated with tooth development, and so with chronological age. The convolutional block consists of an initial convolutional layer with 512 output feature maps of size 7×7 pixels plus a max pooling operation, followed by a second convolutional layer with 1024 output feature maps of size 4×4 plus an average pooling operation. The output of this block is then flattened into a one-dimensional per-tooth-feature vector of size 1024. As the dental development stages have different timings for each tooth, performing a tooth-wise estimation without taking into account the specific tooth type can be misleading. Consequently, per-tooth fully connected layers are applied after the convolutional block. This enabled the development features to be summarised differently for each kind of tooth.

It is well known that some teeth are more relevant in chronological age estimations than others, meaning that more weight should be attached to assessments of their age. In this regard, the per-tooth fully connected layers do not produce a single output with a point estimation of age, but a tuple of two values parameterising a Gaussian distribution $\mathcal{N}_t(\mu_t, \sigma_t^2), t \in T$, with T being the set of detected teeth. Gaussian parameters are especially interpretable in age estimations, as they provide both a measure of centrality (mean) and one of uncertainty (variance).

To constrain the variance parameter to the positive domain, Bishop has suggested adopting an exponential activation function [50]. However, the rapid growth of this function can lead to unstable behaviour, especially in datasets with high variance. Accordingly, we followed the approach proposed in [51] and used a modified version of the ELU activation function (1). As can be seen, the ELU' function retains the ELU behaviour (exponential growth for lower values and linear growth for higher ones) and adds $1 + \epsilon$ to achieve a strictly positive output.

$$ELU'(x) = ELU(x) + 1 + \epsilon$$
$$ELU(x) = \begin{cases} \alpha(e^x - 1) & \text{for} \quad x \leq 0 \\ x & \text{for} \quad x > 0 \end{cases} \tag{1}$$

The aggregation of the estimated per-tooth age distributions is carried out by following the principle that the most uncertain estimation should contribute to the final estimation to a lesser extent than the less uncertain estimation. Consequently, the overall estimated age is calculated through a weighted sum of the means of the per-tooth estimated distributions via

$$\hat{y} = \sum_{t \in T} \mu_t \cdot w_t$$
$$w_t = \frac{1/[(\sigma_t^2)^n]}{\sum_{t \in T} 1/[(\sigma_t^2)^n]} \tag{2}$$

with T being the set of detected teeth, μ_t and σ_t^2 the mean and variance of the estimated distribution for tooth t, and w_t the weight associated with the estimation of tooth t. As the variance of the estimated per-tooth distributions can be seen as a measure of estimation uncertainty, this weight gives less importance to those estimated per-tooth distributions with a high variance. The amount of penalisation can be controlled through the parameter n, also called the penalisation term. The higher the n parameter is, the closer the aggregated estimation will be to the most confident per-tooth estimation.

The architecture of the sex head is almost identical to that of the age head. In this case, the convolutional block is aimed at catching the most sexually dimorphic features of the teeth. Again, the dimorphism proved not to be the same for all the teeth, so the fully connected layers applied at the end of the convolutional layer are not shared across the different tooth types. As this is a binary classification problem, the output of the per-tooth fully connected layers corresponds to a single value p_t which parameterises a Bernoulli distribution such that p_t is the probability that tooth t belongs to a female and $1 - p_t$ is the probability that the same tooth belongs to a male. Again, the mean $\mu_t = p_t$ and variance $\sigma_t^2 = p_t(1 - p_t)$ of the per-tooth estimated probability distributions provide the centrality and uncertainty of the predictions.

To constrain the output p_t to the [0,1] range, a sigmoid activation function is used at the end of the per-tooth fully connected layers. The estimation aggregation is carried out in the same way as in the age head using (2).

4. Experiments and performance evaluation

Although the tooth detector and the age and sex heads worked together to perform age and sex estimations directly from an image, the training process was split into two steps. First, the tooth detector was trained without the age and sex heads, and the detection accuracy was reported. Then, the detector layers were frozen and the age and sex heads were added and trained. In this way, the computational cost was lower for each training process and the convergence was faster. Every experiment was performed using the Detectron2 platform [52].

4.1. Tooth detection

For the experiments related to the tooth detector, the initial dataset of 1746 images was split into training, validation, and test subsets, which contained 60%, 20%, and 20% of the cases, respectively. In particular, a stratified split approach was followed, and therefore the age and sex distribution of the dataset, which is set out in Table 3, was preserved across the three subsets. Image augmentation techniques were applied to increase the size and variability of the data, specifically, the horizontal flip, translation in both axes, rotation, and the brightness and contrast disturbance.

In relation to the training hyperparameters, the batchsize was set to 8 images for the FPN and RPN, and 512 ROIs per image for the ROI head. The optimal grid of RPN anchors was calculated using K-means on the training set, resulting in: rotations of −52, −23, 0, 23, and 52 degrees; aspect ratios of 0.3, 0.6, and 1.4; and sizes of 30^2, 42^2, and 53^2 pixels. The model was trained for 10,000 iterations and validated every 600. The learning rate was initially set to 0.01 and multiplied by a factor of 0.1 at iterations 4000 and 7000. At the end of the training step, the model state was reverted to the validation point where the loss function was minimal. The performance metrics were obtained with the test set.

The main evaluation metric for the tooth detector was the mean average precision (mAP) [53]. This consists of first establishing a minimum overlap – intersection over union (IoU) – between the object ground truth and the prediction from which the detection is considered to be correct. Then, the Average Precision (AP) is used to summarise the precision/recall curve for each tooth class. Finally, the metric is averaged over all the classes. Specifically, three different versions of the mAP metric are reported: mAP@0.5, which uses an IoU of 0.5; mAP@0.75, which takes an IoU of 0.75; and mAP@[0.5:0.95], which averages the mAP over 10 different IoU values, from 0.5 to 0.95.

4.2. Age and sex estimation

The training–validation–test split performed for the tooth detector experiments was preserved in the age and sex estimation step. However, as noted above, there are specific age ranges where it is not feasible to estimate chronological age or sex. Thus, the training, validation, and test subsets were filtered to preserve only the age ranges where the objective is approachable according to the literature. For the case of age estimation experiments, the three subsets were filtered so they contained only subjects younger than 25 (1097 cases). Similarly, sex experiments were carried out in a reduced version of the subsets in which only the subjects older than 16 were preserved (1096 cases).

The loss function used to train the age head combined the negative log-likelihood, which was calculated as the negative logarithm of the Probability Density Function (PDF) for a normally distributed variable, and the L2 regularisation loss on the weights of the neurons that compute the mean and variance components. The overall loss for the age head $\mathcal{L}_a$ was calculated using

$$\mathcal{L}_a = \frac{\sum_{t \in T} -\ln \left[\frac{1}{\sqrt{2\pi}\sigma_t} \quad e^{\frac{-(y-\mu_t)^2}{2\sigma_t^2}} \right]}{|T|} + \lambda \left(\sum W_l^2 \right) \tag{3}$$

where y is the real age, λ is the weight of the L2-loss (set empirically to 10^{-2}), and W_l is the tensor of the weights of the per-tooth fully connected layers.

To increase the training stability, the bias and weights of the per-tooth fully connected layers were initialised in such a way that the predicted per-tooth age distributions at the beginning of the training process corresponded to the age distribution of the training dataset: specifically, $\mu = 15$ years and $\sigma^2 = 13$ years.

The age estimation performance was evaluated using regression metrics on the residuals: real age minus estimated age; absolute errors: absolute value of the residuals; and the explained variance. The median value is employed for the residuals, as it has proved to be more robust in relation to this issue [33]. For the absolute errors, the mean, standard deviation, median, and interquartile range are reported. The 95th percentile of the absolute errors was used as a measure of performance in the worst-case scenarios. Finally, the coefficient of determination R^2 was reported.

The formula for the binary cross-entropy was used to train the sex head:

$$\mathcal{L}_s = \frac{\sum_{t \in T} \left[y \ln \left(p_t(y) \right) + \left(1 - y \right) \ln \left(1 - p_t(y) \right) \right]}{|T|} \tag{4}$$

where y is the real sex of the subject in the image: 1 = female; 0 = male.

Sex estimation capabilities were assessed with traditional classification metrics, i.e., the overall accuracy, precision, recall, and F1 score for each class of female and male.

The impact of the uncertainty-aware distribution aggregation was analysed for both the age and sex heads. Specifically, we tested four policies based on different n values in (2). The first attaches equal weight to each per-tooth estimation, i.e., $n = 0$. In the second, the per-tooth estimations are weighted linearly by their uncertainty, i.e., $n = 1$. In the third version, the highly uncertain estimations are heavily penalised by increasing the parameter n in (2). The impact of the n value on the performance was also analysed. Finally, the fourth policy considers the prediction produced by the less-uncertain tooth ($n \to \infty$).

5. Results

5.1. Tooth detection performance

The detection performance is reported in Table 4, along with the details of other tooth detection networks available in the literature. The

Table 4
Comparison of the available tooth detection systems with the proposed approach.[a]

Model	mAP@[0.5:0.95]	mAP@0.5	mAP@0.75
Silva (2020) [21]	74.0	**99.7**	89.0
Chung (2020) [27]	**81.0**	91.0	90.0
Ours	77.6	96.4	**94.4**

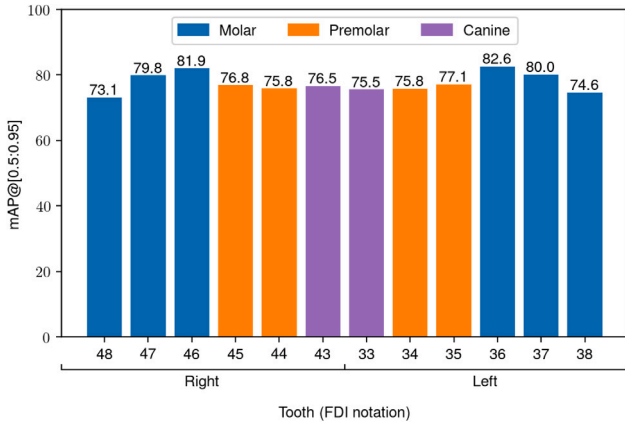[a]Only those methods reporting AP-based measurements were included.

Fig. 3. Detection accuracy for each tooth type. The first and second molars (46, 45, 45, and 36) were clearly the best performers for this purpose.

proposed approach outperformed the others in terms of the mAP@0.75 metrics (94.4 vs 89.0 and 90.0); it remained in second place overall for the mAP@[0.5:0.95] and mAP@0.5 metrics.

The mAP@[0.5:0.95] metric is represented in Fig. 3. As can be seen, the highest detection accuracy was achieved in teeth 46 and 36, corresponding to the first molars, with values of 81.9 and 82.6, respectively. The next most accurate teeth were 47 and 37, also known as the second molars, with values of 79.8 and 82.6. The detection performance fell in the rest of the tooth types, reaching a minimum of 73.1 and 74.6 in teeth 48 and 38, i.e., the third molars.

5.2. Age estimation performance

The age estimation performances for the different aggregation policies can be seen in Table 5 and Fig. 4 (blue colour). When the aggregation consisted of averaging the means of the per-tooth distributions ($n = 0$ in (2)), there was a noticeable systematic overestimation of age (the median E was -0.20 years), while the mean AE was 1.11 years. The worst 5% of the estimations yielded an AE greater than or equal to 3.29 years. The inclusion of a parameter that penalises the uncertainty of the per-tooth estimations linearly ($n = 1$) helped to slightly reduce the overestimation, with the median then E -0.17 years. The metrics based on the AE were also improved, it being remarkable that the mean AE was below 1 year and the median AE decreased from 0.82 to 0.63 years. Moreover, the explained variance improved from 92% to 94%. When only the tooth with the least uncertainty was taken into account, the AE-based metrics tended to be slightly worse, being particularly noticeable in the mean (1.11 years) and the IQR (1.54).

As shown in Fig. 4, when the n parameter was increased further, leading to exponential uncertainty penalisations, the age estimation errors fell until they reached a minimum at $n = 3$, and started to increase thereafter. When the n value was set to that optimal point, the overestimation was reduced to almost zero, with the median being E -0.06 years. Every other metric was also improved, and it was particularly noticeable that the AE interquartile range decreased to 1.11 years and the 95th percentile to less than 3 years (2.88).

Fig. 5 contains an example of the estimated per-tooth age distributions for a patient aged 18 years and 3 months. In this case,
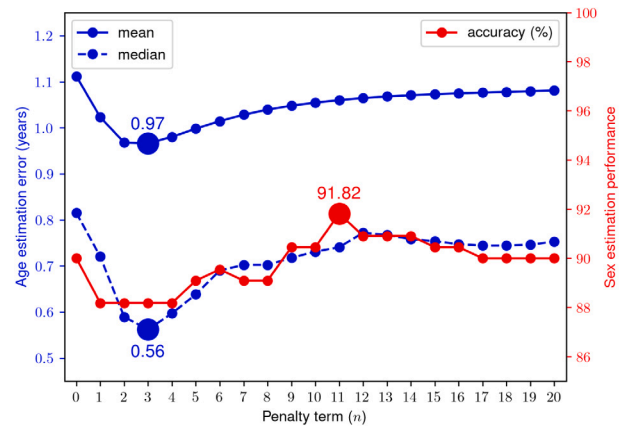

Fig. 4. Impact of the penalty term, n, on the age and sex estimations (blue and red lines, respectively) for values greater than or equal to 1. The best estimates for age were obtained with $n = 3$, and the most accurate sex classifications with $n = 11$.
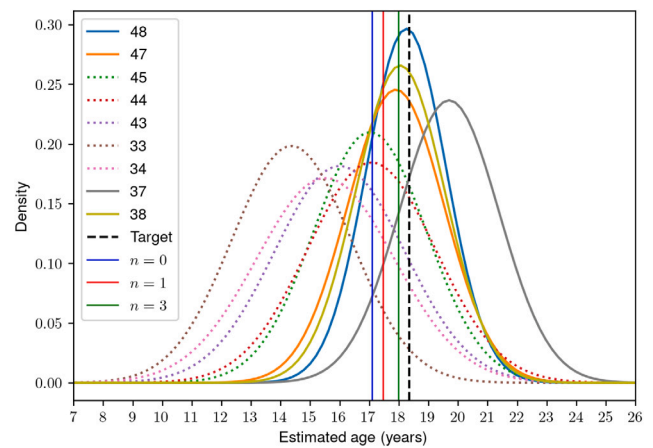

Fig. 5. Example of the estimated per-tooth age distributions for a subject aged 18 years and 3 months, and the final prediction according to the different aggregation policies ($n \in \{0, 1, 3\}$). The solid and dotted per-tooth distributions correspond to the teeth producing the most and least confident outcomes, respectively.

the most anterior teeth – 33 and 43 (canines) and 34, 44, and 45 (premolars) – yielded age distributions centred well away from the real age. These distributions also had a high degree of variance. As a counterpart, the most posterior teeth – 37, 47, 38 and 48 (molars) – produced distributions with a mean closer to the target age and a noticeable lower variance. The first aggregation policy ($n = 0$), which does not take into account the variance of the distributions, produced poor estimations because of the influence of the distributions of the most anterior teeth, which significantly underestimated the real age. The second aggregation policy ($n = 1$), which weights each per-tooth estimation by the inverse of the distribution variance, produced an estimation closer to the target age. This is because its reliance on the widest distributions is less than on the tightest. Finally, the third aggregation policy ($n = 3$, according to the optimal value reported in Fig. 4), which penalises the uncertainty exponentially, achieved a final estimation even closer to the real age of the subject, as evidenced by the significant increase in the difference regarding the weights assigned to the most and least reliable estimations.

The estimations produced by the best aggregation policy (exponential uncertainty penalisation with $n = 3$) are showed graphically in Fig. 6, and the performance metrics for each age range are also reported in Table 6. It can be seen that the behaviour is more reliable in the younger age groups, especially in those subjects younger than 10, as the estimations are very close to the ideal performance and the mean

Table 5

Impact of the aggregation policy of age head on the estimation performance.

Aggregation policy	E	AE					R^2
	med.	μ	σ	med.	IQR	p95	
No uncertainty ($n = 0$)	−0.20	1.11	1.04	0.82	1.33	3.29	0.92
Linear ($n = 1$)	−0.17	0.99	0.96	0.63	1.23	3.12	**0.94**
Exp ($n = n^*, n > 1$)	**−0.06**	**0.97**	**0.94**	**0.56**	**1.11**	2.88	**0.94**
Only best tooth ($n \to \infty$)	**0.02**	1.11	1.03	0.74	1.54	**2.82**	0.92

Table 6

Performance metrics of the proposed age estimation method per age range, with the optimal n penalisation parameter ($n = 3$).

Age range	E(med.)	AE ($\mu \pm \sigma$)
(5,7]	0.28	0.44 ± 0.32
(7,9]	0.21	0.43 ± 0.34
(9,11]	0.28	0.71 ± 0.55
(11,13]	0.13	0.96 ± 0.79
(13,15]	0.14	1.22 ± 1.02
(15,17]	0.26	0.82 ± 0.78
(17,19]	0.22	0.76 ± 0.65
(19,21]	−0.30	0.91 ± 0.67
(21,23]	−1.69	1.76 ± 1.28
(23,25]	−2.90	2.87 ± 0.90

Table 7

Comparison of the available automatic age estimation systems and the proposed approach.

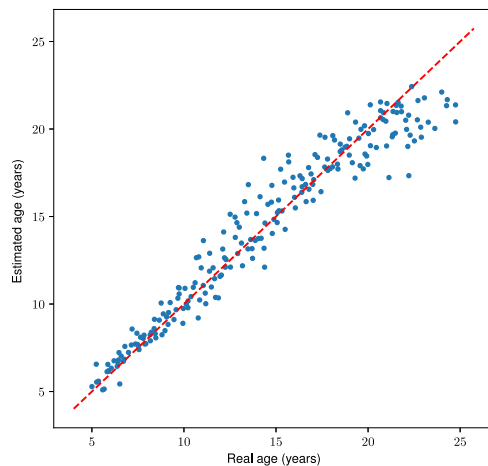| Model	Age range	E (med.)	AE ($\mu \pm \sigma$)
Čular (2017) [31]	10–25	–	2.28 ± 2.17
De Back (2019) [32]	5–25	–	1.75 ± –
Vila-Blanco (2020) [33]	5–25	0.07	1.17 ± 1.11
Wallraff (2021) [34]	11–20	−0.30	1.08 ± –
Hou (2021) [35]	0–93	–	1.64 ± –
Ours	5–25	**−0.06**	**0.97 ± 0.94**

Fig. 6. Real age versus the predictions of the proposed method, with the optimal n penalisation parameter ($n = 3$). The red dashed line shows the ideal performance.

absolute error is lower than 0.45 years. On the contrary, the absolute error in subjects older than 20 exceeds 1.75 years, motivated by a noticeable tendency to underestimate age.

Table 7 contains our comparison between the performances of our proposed method and those of other approaches that also aim to estimate numeric age automatically. As can be seen, the methodology described in this study performed better than any other, with improvements to the mean AE ranging from 0.11 (vs Wallraff et al. [34]) to 1.31 years (vs Čular et al. [31]). Furthermore, the distribution of the AE was tighter than in other studies based on its standard deviation. Finally, the overestimation of our method was close to zero, which is comparable in absolute terms to the underestimation we obtained in a previous study [33].

5.3. Sex estimation performance

The sex estimation performance is set out on both the right side of Table 8 and in the red graph in Fig. 4. The accuracy was 90% when the aggregation of the per-tooth distributions was conducted using a simple average of the estimated probabilities ($n = 0$). The recall was 94.54% for females and 85.45% for males, while the precision was 86.67% and 94%, respectively. The F1 measure was around 90% in both classes. When a linear uncertainty penalisation was added ($n = 1$), the overall accuracy dropped to 88.18%. The female recall and the male precision decreased noticeably, while the female precision and the male recall increased to a lesser extent, resulting in a more balanced level of performance across these classes. The F1 scores were noticeably lower, with a difference of about 2% with respect to the aggregation with no uncertainty ($n = 0$). The results were much worse when only the tooth with the least uncertainty was used for the sex classification. Specifically, the accuracy fell below 87% (86.82%) and none of the other metrics were above 89%.

The sex classification accuracy followed a different pattern to the age estimation performance when the value of n was increased. As shown in Fig. 4 (red line), every value under 9 yielded an accuracy below that obtained with the simplest aggregation policy. The results improved the baseline ($n = 0$) from $n = 9$ to $n = 14$, and then the performance started to drop. The greatest accuracy was achieved with $n = 11$, reaching 91.82%. In this optimal setup, the recall for the female class and the precision for the male class remained very similar to these values obtained without uncertainty-aware aggregation. However, the precision for the female class, the recall for the male class, and the F1 score improved by more than 2%.

An example of the aggregation of per-tooth sex distributions for a female subject is presented in Fig. 7. In this case, the best per-tooth estimations were produced for teeth 46, 44, 43, 33, and 34, and the worst for 48, 45, and 35. The three aggregation policies described in Section 3 were analysed, with the penalisation parameter n in the exponential penalisation policy set to 11 based on the optimal value reported in the red graph in Fig. 4. In every case, the returned probability classified the sex of the subject using the threshold of 0.5. However, when the most uncertain estimations – those that assign similar probabilities to the male and female classes – were penalised, the output probability, and hence the confidence of the classification, was higher for the female class.

Table 9 contains our comparison of the results of our proposed sex estimation method to those of the other automatic sex estimation approaches. In general terms, the former was outperformed by the latter in relation to the classification accuracy achieved, with the differences between 2.5% (vs Ilić et al. [38]) and 5.1% (vs Milošević et al. [39]).

6. Discussion and conclusions

Estimating chronological age and biological sex is a crucial task in a number of clinical or forensic processes, and has been widely tackled by the research community [7,8]. Nevertheless, the majority of studies have described manual or semiautomatic approaches, which are very time-consuming and have a high degree of subjectivity. In recent years,

Table 8

Impact of the aggregation policy of sex head on the estimation performance.

Aggregation policy	Acc.	Female			Male		
		Recall	Precision	F1	Recall	Precision	F1
No uncertainty ($n = 0$)	90.00	**94.54**	86.67	90.43	85.45	94.00	89.52
Linear ($n = 1$)	88.18	90.00	88.64	88.39	86.36	89.62	87.96
Exp ($n = n^{*}, n > 1$)	**91.82**	**94.54**	**89.65**	**92.03**	**89.09**	**94.23**	**91.58**
Only best tooth ($n \to \infty$)	86.82	88.18	85.84	87.00	85.45	87.85	86.63

Table 9

Comparison of the available automatic sex estimation systems and the proposed approach.

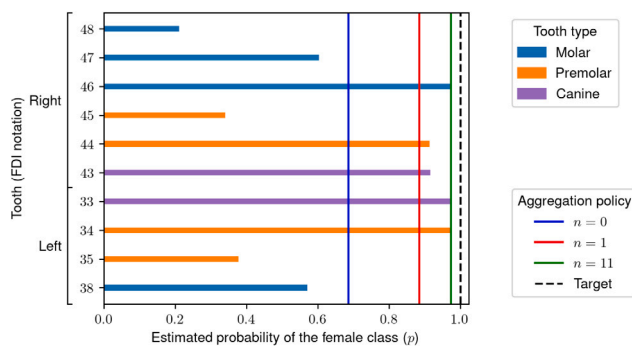| Model	Age range	Accuracy (%)
Ilić (2019) [38]	–	94.3
Milošević (2019) [39]	19–85	**96.9**
Ke (2020) [40]	16–70	94.6
Vila-Blanco (2020) [41]	5–90	75.0–96.7
Ours	16–60	91.8

Fig. 7. Example of the estimated per-tooth sex distributions for a female subject, and the final prediction according to the different aggregation policies ($n \in \{0, 1, 11\}$).

fully automatic methods have been proposed for processing a dental X-ray image and using it to establish the subject's chronological age and sex [32,33,39,41]. These studies report high performance levels that enable rapid and confident estimations, suggesting that the subjectivity problems of manual processes can be greatly reduced. However, and despite efforts to improve the interpretability of these methods [33], they still lack mechanisms to explain the reasoning behind their findings.

This study proposes a two-step process for estimating chronological age and biological sex using an OPG. In the first step, a CNN is used to detect a set of mandibular teeth, outputting oriented bounding boxes where each tooth is located. The features learned by this network are evaluated in the boxes and fed into a second CNN, which applies a set of convolutional layers and a fully connected layer to produce a per-tooth estimation of age or sex. Instead of giving a single point estimation for each tooth, it yields a parameterised probability distribution, which is Gaussian for the age estimation and a Bernoulli for the sex classification. These per-tooth distributions are then aggregated by taking into account their uncertainty, that is, through a weighted sum which gives to the mean of each distribution a weight that is inversely proportional to the distribution's variance.

The performance of the tooth detection was comparable to the other automatic methodologies, outperforming them in the mAP@[0.5:0.95] metric and staying in second position in the other two. This is quite remarkable, since the proposed approach was evaluated in mixed dentitions (children and adults), while the other methods described have only been tested on fully developed adult dentitions. We would expect a noticeable drop-off in the performance of these other methodologies if child subjects were included in their samples, as this would then involve very small objects (when tooth development is beginning), as well as great variability in terms of the tooth aspect ratio (the teeth are wider when they start to develop and become taller as they grow).

In addition, and to the best of our knowledge, this is the first attempt to detect the teeth using oriented bounding boxes. This type of box fits the tooth better than the horizontally aligned versions, as any overlapping with the neighbour teeth is avoided. Consequently, it also enables the extraction of data on tooth orientation, which is crucial for detecting conditions such as impacted wisdom teeth, as well as other measurements like height and width.

Our proposed approach was evaluated in relation to the age and sex estimation subsystem using according to four different ways of aggregating the per-tooth probability distributions: the average of the distribution means; the weighted sum of the distribution means through a linear penalisation of the uncertainty; the weighted sum of the distribution means via an exponential penalisation of the uncertainty; and using the least uncertain tooth as the only predictive indicator.

The age estimation experiments were performed with subjects younger than 25, as results are reportedly inaccurate in older subjects due to their fully developed teeth [54]. The outcomes confirmed that the uncertainty-aware aggregation policies increased the predictive capabilities of the model and decreased the errors in the estimations. The overestimation of 0.2 years (2.4 months) produced by the first policy was reduced to 0.17 (2.04 months) years by the second and 0.06 years (0.72 months) by the third; the fourth policy, meanwhile, yielded a negligible underestimation of 0.02 (0.24 months). Similarly, the AE-based metrics were also improved: 0.97 years for the mean AE, while the model was able to catch 94% of the data variability, with both of these outcomes occurring when an exponential penalisation of the uncertainty was applied. It has also been confirmed that the performance was significantly lower for subjects older than 20 due to the end of the tooth development period.

These results demonstrate that the proposed methodology outperforms the other automatic age estimation systems, in terms of both under- or overestimations and absolute errors. This is remarkable, as the new approach only uses image information retrieved from inside the tooth bounding boxes, unlike the studies by De Back et al. [32], Vila-Blanco et al. [33], Wallraff et al. [34], and Hou et al. [35], which used the entire OPG to produce their age estimations. This suggests that the teeth are the structures in the oral cavity which correlates best with the chronological age, at least in OPGs. It is also notable that the performance is still comparable or even better than the other methods in cases where only the best tooth is used. The achieved performance make this approach suitable for being used as the primary method in any required age estimation task.

The results for sex estimation were obtained from subjects older than 16, as puberty is the point at which dimorphic skeletal indicators become noticeable [6]. The performance of the same four aggregation policies yielded a different pattern than in the age estimation experiments. Although the classification accuracy dropped from 90% ($n = 0$) to 88.18% ($n = 1$) when applying the linear uncertainty penalisation, the gap between the female and male outcomes was reduced. The classification performance with the optimal n value ($n = 11$) was better than that achieved with both the first and second aggregation policies for every single metric, with an overall accuracy of almost 92%. Finally, the results obtained with the least uncertain tooth were, by far, the worst performers in the comparison, indicating that multiple teeth must be included to get the best outcomes.

The degree of accuracy achieved with our approach was remarkable and implies that the teeth have great potential for detecting biological

sex. However, the alternative automatic methods for estimating sex from OPGs have produced better outcomes [38–41], suggesting that other structures present are highly dimorphic. This is in line with other studies that report a greater degree of dimorphism in the mandible than in the teeth [11]. However, it is noticeable that the proposed approach outperformed the other method aimed at estimating the sex from the radiological analysis of the teeth [55] by a very large margin (91.8% vs 79.5% of accuracy, respectively). In this regard, we believe that this method could be complemented with the addition of other highly-dimorphic bone structures in the radiological image, such as the mandible, to build a more accurate yet explainable sex estimation pipeline.

The visual representations of the predicted per-tooth distributions have proved to be consistent with the findings in the clinical literature. As an example, the most confident age estimations for the case presented in Fig. 5 (a subject aged 18 years and 3 months) were achieved with the third molars (38 and 48), which are expected to be the only teeth in development at that age [6]. It can be seen in Fig. 7 that the canines (33 and 43) yielded the most confident sex estimations, followed by the first premolars (34 and 44). This is in line with the study by Zorba et al. [56].

Overall, the proposed age and sex estimation models have some advantages over earlier automatic approaches. First, they do not require the presence of any specific tooth to work, unlike other methods [28–31,36], ensuring their applicability in a wider range of scenarios. Moreover, although our model is able to produce age and sex estimations with a single tooth, its performance is expected to improve when multiple teeth are available. In addition, the output of our age estimation methodology is a numeric value, which is more precise than that obtained with the N-stage classification system [22,28–30,37]. It is worth noting that the performance reported for both the age and sex estimation methods was obtained in an OPG database collected with no exclusion criteria, which entails a very realistic test environment.

The interpretability of automatic estimation methods, particularly those related to the medical field, is already required by the main competent authorities to approve their application in real environments [57]. In this regard, it is crucial to know how the estimation methods work and which piece of information they rely on the most. This can be useful not only for validating these methods against the clinical findings already reported in the literature, but also to discover new indicators that were not explored yet. In the particular case of the proposed age and sex prediction method, the estimations can be easily traced back to build understandable explanations. The fact that an individual probability distribution is produced for each detected tooth gives experts a powerful tool to with which to conduct further studies regarding: the teeth that make the best contributions; the most suitable teeth for a specific age cohort; the behaviour of the proposed approach when the most relevant teeth are missing; and the minimum set of teeth required to produce the optimum performance. Moreover, the age estimation methodology could also be used to detect issues relating to the development of specific teeth, such as advanced or delayed growth.

In conclusion, the methodology described in this paper is valuable for both accurately estimating chronological age and biological sex in a variety of scenarios using OPGs, and producing interpretable per-tooth estimations that are easy to analyse.

CRediT authorship contribution statement

Nicolás Vila-Blanco: Conceptualization, Methodology, Software, Writing – original draft. **Paulina Varas-Quintana:** Resources, Data curation. **Ángela Aneiros-Ardao:** Resources, Data curation. **Inmaculada Tomás:** Supervision, Validation, Writing – review & editing. **María J. Carreira:** Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. Franklin, Forensic age estimation in human skeletal remains: current concepts and future directions, Leg Med 12 (1) (2010) 1–7.
[2] H.M. Garvin, N.V. Passalacqua, Current practices by forensic anthropologists in adult skeletal age estimation, J Forensic Sci 57 (2) (2012) 427–433.
[3] J. Bruzek, A method for visual determination of sex, using the human hip bone, Am J Phys Anthropol 117 (2) (2002) 157–168.
[4] D.T. Case, A.H. Ross, Sex determination from hand and foot bone lengths, J Forensic Sci 52 (2) (2007) 264–270.
[5] I. Özer, K. Katayama, M. Sahgir, E. Güleç, Sex determination using the scapula in medieval skeletons from East Anatolia, Coll Anthropol 30 (2) (2006) 415–419.
[6] A.M. Christensen, N.V. Passalacqua, E.J. Bartelink, Chapter 10 - age estimation, in: Forensic Anthropology, second ed., Academic Press, 2019, pp. 307–349.
[7] G. Willems, A review of the most commonly used dental age estimation techniques, J Forensic Odontostomatol 19 (1) (2001) 9–17.
[8] A.P. Joseph, R. Harish, P.K.R. Mohammed, R. Vinod Kumar, How reliable is sex differentiation from teeth measurements, Oral Maxillofac Pathol J 4 (1) (2013) 289–292.
[9] R. Kapila, K. Nagesh, A.R. Iyengar, S. Mehkri, Sexual dimorphism in human mandibular canines: a radiomorphometric study in South Indian population, J Dent Res Dent Clin Dent Prospects 5 (2) (2011) 51.
[10] G.T. Schwartz, M.C. Dean, Sexual dimorphism in modern human permanent teeth, Am J Phys Anthropol 128 (2) (2005) 312–317.
[11] D. Franklin, P. O'Higgins, C.E. Oxnard, I. Dadour, Sexual dimorphism and population variation in the adult mandible, Forensic Sci Med Pathol 3 (1) (2007) 15–22.
[12] N. Vila-Blanco, P. Varas-Quintana, Á. Aneiros-Ardao, I. Tomás, M.J. Carreira, Automated description of the mandible shape by deep learning, Int J Comput Ass Rad (2021) 1–10.
[13] A. Panchbhai, Dental radiographic indicators, a key to age estimation, Dentomaxillofac Radiol 40 (4) (2011) 199–212.
[14] A.K. Jain, H. Chen, Matching of dental X-ray images for human identification, Pattern Recognit 37 (7) (2004) 1519–1532.
[15] A.B. Oktay, Tooth detection with convolutional neural networks, in: Medical Technologies National Congress (TIPTEKNO), IEEE, 2017, pp. 1–4.
[16] G. Silva, L. Oliveira, M. Pithon, Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives, Expert Syst Appl 107 (2018) 15–31.
[17] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, L. Oliveira, Deep instance segmentation of teeth in panoramic X-ray images, in: 31st Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE, 2018, pp. 400–407.
[18] D.V. Tuzoff, L.N. Tuzova, M.M. Bornstein, A.S. Krasnov, M.A. Kharchenko, S.I. Nikolenko, M.M. Sveshnikov, G.B. Bednenko, Tooth detection and numbering in panoramic radiographs using convolutional neural networks, Dentomaxillofac Radiol 48 (4) (2019) 20180051.
[19] T.L. Koch, M. Perslev, C. Igel, S.S. Brandt, Accurate segmentation of dental panoramic radiographs with U-nets, in: IEEE 16th International Symposium on Biomedical Imaging (ISBI), IEEE, 2019, pp. 15–19.
[20] A.F. Leite, A. Van Gerven, H. Willems, T. Beznik, P. Lahoud, H. Gaêta-Araujo, M. Vranckx, R. Jacobs, Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs, Clin Oral Investig 25 (4) (2021) 2257–2267.
[21] B. Silva, L. Pinheiro, L. Oliveira, M. Pithon, A study on tooth segmentation and numbering using end-to-end deep neural networks, in: 33rd Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE, 2020, pp. 164–171.
[22] C. Kim, D. Kim, H. Jeong, S.-J. Yoon, S. Youm, Automatic tooth detection and numbering using a combination of a CNN and heuristic algorithm, Appl Sci-Basel 10 (16) (2020) 5624.
[23] C. Muramatsu, T. Morishita, R. Takahashi, T. Hayashi, W. Nishiyama, Y. Ariji, X. Zhou, T. Hara, A. Katsumata, E. Ariji, et al., Tooth detection and classification on panoramic radiographs for automatic dental chart filing: improved classification by multi-sized input data, Oral Radiol 37 (1) (2021) 13–19.
[24] J.-H. Lee, S.-S. Han, Y.H. Kim, C. Lee, I. Kim, Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs, Oral Surg Oral Med Oral Pathol Oral Radiol 129 (6) (2020) 635–642.
[25] M.P. Muresan, A.R. Barbura, S. Nedevschi, Teeth detection and dental problem classification in panoramic X-Ray images using deep learning and image processing techniques, in: IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), IEEE, 2020, pp. 457–463.

[26] F.P. Mahdi, K. Motoki, S. Kobashi, Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs, Sci Rep 10 (1) (2020) 1–12.

[27] M. Chung, J. Lee, S. Park, M. Lee, C.E. Lee, J. Lee, Y.-G. Shin, Individual tooth detection and identification from dental panoramic x-ray images via point-wise localization and distance regularization, Artif Intell Med 111 (2021) 101996.

[28] J. De Tobel, P. Radesh, D. Vandermeulen, P.W. Thevissen, An automated technique to stage lower third molar development on panoramic radiographs for age estimation: a pilot study, J Forensic Odontostomatol 35 (2) (2017) 42.

[29] R. Merdietio Boedi, N. Banar, J. De Tobel, J. Bertels, D. Vandermeulen, P.W. Thevissen, Effect of lower third molar segmentations on automated tooth development staging using a convolutional neural network, J Forensic Sci 65 (2) (2020) 481–486.

[30] N. Banar, J. Bertels, F. Laurent, R.M. Boedi, J. De Tobel, P. Thevissen, D. Vandermeulen, Towards fully automated third molar development staging in panoramic radiographs, Int J Legal Med (2020) 1–11.

[31] L. Čular, M. Tomaić, M. Subašić, T. Šarić, V. Sajković, M. Vodanović, Dental age estimation from panoramic X-ray images using statistical models, in: Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis, IEEE, 2017, pp. 25–30.

[32] W. De Back, S. Seurig, S. Wagner, B. Marré, I. Roeder, N. Scherf, Forensic age estimation with Bayesian convolutional neural networks based on panoramic dental X-ray imaging, in: International Conference on Medical Imaging with Deep Learning (MIDL), 2019.

[33] N. Vila-Blanco, M.J. Carreira, P. Varas-Quintana, C. Balsa-Castro, I. Tomás, Deep neural networks for chronological age estimation from OPG images, IEEE Trans Med Imaging (2020).

[34] S. Wallraff, S. Vesal, C. Syben, R. Lutz, A. Maier, Age estimation on panoramic dental X-ray images using deep learning, in: Bildverarbeitung Für Die Medizin 2021, Springer, 2021, pp. 186–191.

[35] W. Hou, L. Liu, J. Gao, A. Zhu, K. Pan, H. Sun, N. Zheng, Exploring effective DNN models for forensic age estimation based on panoramic radiograph images, in: 2021 International Joint Conference on Neural Networks, IJCNN, IEEE, 2021, pp. 1–8.

[36] S. Kim, Y.-H. Lee, Y.-K. Noh, F.C. Park, Q.-S. Auh, Age-group determination of living individuals using first molar images based on artificial intelligence, Sci Rep 11 (2021).

[37] Y.-C. Guo, M. Han, Y. Chi, H. Long, D. Zhang, J. Yang, Y. Yang, T. Chen, S. Du, Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images, Int J Legal Med (2021) 1–9.

[38] I. Ilić, M. Vodanović, M. Subašić, Gender estimation from panoramic dental X-ray images using deep convolutional networks, in: IEEE EUROCON 18th International Conference on Smart Technologies, IEEE, 2019, pp. 1–5.

[39] D. Milošević, M. Vodanović, I. Galić, M. Subašić, Estimating biological gender from panoramic dental X-Ray images, in: 11th International Symposium on Image and Signal Processing and Analysis (ISPA), IEEE, 2019, pp. 105–110.

[40] W. Ke, F. Fan, P. Liao, Y. Lai, Q. Wu, W. Du, H. Chen, Z. Deng, Y. Zhang, Biological gender estimation from panoramic dental X-ray images based on multiple feature fusion model, Sens Imaging 21 (1) (2020) 1–11.

[41] N. Vila-Blanco, R.R. Vilas, M.J. Carreira, I. Tomás, Towards deep learning reliable gender estimation from dental panoramic radiographs, in: 9th Starting AI Researcher Symposium (STAIRS). ECAI 2020 (European Conference on Artificial Intelligence), 2655, CEURS Workshop Proceedings, 2020.

[42] W.M. Association, World medical association declaration of helsinki: Ethical principles for medical research involving human subjects, JAMA 310 (20) (2013) 2191–2194, http://dx.doi.org/10.1001/jama.2013.281053, arXiv:https://jamanetwork.com/journals/jama/articlepdf/1760318/jsc130006.pdf.

[43] Labelbox, Labelbox, , 2021, https://labelbox.com/.

[44] P. Hölttä, M. Nyström, M. Evälahti, S. Alaluusua, Root–crown ratios of permanent teeth in a healthy finnish population assessed from panoramic radiographs, Eur J Orthodont 26 (5) (2004) 491–497.

[45] S. Perschbacher, Interpretation of panoramic radiographs, Aust Dent J 57 (2012) 40–45.

[46] Dentistry — Designation system for teeth and areas of the oral cavity, 2016, International Organization for Standardization, Geneva, CH, 2016.

[47] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, in: C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems, 28, Curran Associates, Inc., 2015.

[48] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125.

[49] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.

[50] C.M. Bishop, Mixture density networks, Aston University, 1994.

[51] A. Brando Guillaumes, Mixture density networks for distribution and uncertainty estimation, Universitat Politècnica de Catalunya, 2017.

[52] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, R. Girshick, Detectron2, , 2019, https://github.com/facebookresearch/detectron2.

[53] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (VOC) challenge, Int J Comput Vis 88 (2) (2010) 303–338.

[54] J.L. Prieto, E. Barbería, R. Ortega, C. Magaña, Evaluation of chronological age based on third molar development in the spanish population, Int J Legal Med 119 (6) (2005) 349–354.

[55] C. Capitaneanu, G. Willems, R. Jacobs, S. Fieuws, P. Thevissen, Sex estimation based on tooth measurements using panoramic radiographs, Int J Leg Med 131 (3) (2017) 813–821.

[56] E. Zorba, K. Moraitis, S.K. Manolis, Sexual dimorphism in permanent teeth of modern greeks, Forensic Sci Int 210 (1–3) (2011) 74–81.

[57] J. Amann, D. Vetter, S.N. Blomberg, H.C. Christensen, M. Coffee, S. Gerke, T.K. Gilbert, T. Hagendorff, S. Holm, M. Livne, et al., To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems, PLOS Digit Health 1 (2) (2022) e0000016.