Deep Neural Networks for Chronological Age Estimation From OPG Images

Nicolás Vila-Blanco, María J. Carreira, *Member, IEEE*, Paulina Varas-Quintana, Carlos Balsa-Castro, and Inmaculada Tomás

***Abstract*— Chronological age estimation is crucial labour in many clinical procedures, where the teeth have proven to be one of the best estimators. Although some methods to estimate the age from tooth measurements in orthopantomogram (OPG) images have been developed, they rely on time-consuming manual processes whose results are affected by the observer subjectivity. Furthermore, all those approaches have been tested only on OPG image sets of good radiological quality without any conditioning dental characteristic. In this work, two fully automatic methods to estimate the chronological age of a subject from the OPG image are proposed. The first (DANet) consists of a sequential Convolutional Neural Network (CNN) path to predict the age, while the second (DASNet) adds a second CNN path to predict the sex and uses sex-specific features with the aim of improving the age prediction performance. Both methods were tested on a set of 2289 OPG images of subjects from 4.5 to 89.2 years old, where both bad radiological quality images and images showing conditioning dental characteristics were not discarded. The results showed that the DASNet outperforms the DANet in every aspect, reducing the median Error (E) and the median Absolute Error (AE) by about 4 months in the entire database. When evaluating the DASNet in the reduced datasets, the AE values decrease as the real age of the subjects decreases, until reaching a median of about 8 months in the subjects younger than 15. The DASNet method was also compared to the state-of-the-art manual age estimation methods, showing significantly less over- or under-estimation problems. Consequently, we conclude that the DASNet can be used to automatically predict the chronological age of a subject accurately, especially in young subjects with developing dentitions.**

***Index Terms*— Deep learning, panoramic images, chronological age, dental age, forensic age.**

I. Introduction

CHRONOLOGICAL age estimation is a key task in a variety of scenarios: in forensics to estimate the age of a corpse; in trials where the age of the defendant is unclear because of a suspicion that he/she is lying or there is an absence of clarifying documentation; in migration management to assess if a migrant is, or is not, an adult; and in adoption processes to determine the estimated age of undocumented children. A skeletal analysis is one of the best methods for estimating age, especially in children and teenagers, with the most reliable body indicators being the hand-wrist area and the teeth. The hand-wrist as a manual tool for estimating age has been studied for many years and is described in a number of publications [1]–[3], as are some automated methods [4], [5]. Meanwhile, although the development of the teeth is reportedly dependent on race, this is affected less by environmental factors. This means that any variability is reduced during growth, as highlighted in established research [6]–[9].

The relevant literature describes several methods for conducting chronological age estimations by analysing dental maturation in orthopantomogram (OPG) images, which provide a panoramic view of the entire dentition and surrounding bones in the oral cavity. One of the first such methods was presented by Nolla [6], who proposed a 10-stage model of individual tooth development, with the first stage being the presence of the crypt and the last stage occurring when the roots have fully developed and the apices are closed. Each stage provides a score for an individual tooth. These are then added together and the result is translated into a chronological age value using a specific table. Another well-known method was proposed by Haavikko [10], who identified a link between the alveolar and clinical tooth eruption (as defined in [11] and [12], respectively) in relation to the dental development stages proposed by Gleiser & Hunt [11].

Perhaps the most studied and tested method is that developed by Demirjian et al. [13], who also identified a set of dental development stages (eight in this case) and produced a score for each of them. These stages are also translated

Manuscript received November 26, 2019; revised January 3, 2020; accepted January 19, 2020. Date of publication January 31, 2020; date of current version June 30, 2020. This work was supported in part by the Consellería de Cultura, Educación e Ordenación Universitaria under Grant ED431G/08, in part by the Potential Growth Group ED431B 2017/029, in part by the Competitive Reference Group ED431C 2017/69, in part by the N Vila-Blanco Support ED481A-2017, and in part by the European Regional Development Fund (ERDF). *(Corresponding author: María J. Carreira.)*

Nicolás Vila-Blanco and María J. Carreira are with the Centro Singular de Investigación en Tecnoloxías Intelixentes, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain, and also with the Health Research Institute Foundation of Santiago (FIDIS), 15706 Santiago de Compostela, Spain (e-mail: nicolas.vila@usc.es; mariajose.carreira@usc.es).

Paulina Varas-Quintana, Carlos Balsa-Castro, and Inmaculada Tomás are with the Oral Sciences Research Group, Department of Surgery and Medical Surgical Specialities, Special Needs Unit, School of Medicine and Dentistry, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain, and also with the Health Research Institute Foundation of Santiago (FIDIS), 15706 Santiago de Compostela, Spain (e-mail: paulina.varas@rai.usc.es; cbalsa@coitt.es; inmaculada.tomas@usc.es).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2020.2968765

into scores via a specific table, with the score of each tooth depending on its stage of development and the sex of the subject. The total of all the scores is known as *maturity index*. This can be represented with centiles where the chronological age is known, meaning that the distribution table can be used to estimate age. In later work [14], the authors revisited their model and proposed three new scoring systems, with the main changes being the addition of new development stages for some teeth and the use of different teeth to predict age.

More recently, Cameriere *et al.* [8] studied the relationship between open apices in tooth roots and chronological age. They relied on the left-hand side mandibular teeth to measure the distance between the inner side of the open apex in single-rooted teeth and the sum of the distances between the inner side of the open apices in multirooted teeth, with a regression formula proposed to translate all the measurements into a chronological age value.

All these methods have been compared in a variety of populations [15]–[30], with helpful conclusions reached about the most suitable teeth for age estimation and dental development differences depending on sex and race.

It is worth noting that gender influences the pattern of dental development. Indeed, it has been proved that girls reach a particular development stage before boys [31], which is especially noticeable with permanent teeth [32]. Consequently, the methods mentioned above take gender into account to improve the performance of these age-estimation approaches. In some cases, the tables mapping specific indices to numeric ages were compiled independently for boys and girls [6], [13]; in other cases, the formulae to regress the age include sex as an input variable [8], [33].

Although these age-estimation methods perform reasonably well and are widely used in the scientific community [34], [35], all of them rely on manual measurements or classifications, which require much time and effort. Indeed, in an example described in [36], the estimated mean time for calculating an individual's chronological age using the Demirjian-Chaillet method was 10 minutes. The high estimated time-demands of the classical manual methods is a significant limitation given the large sample sizes required methodologically if optimal results are to be obtained [26], [37], [38]. The inconvenience of applying these manual methods in routine clinical activity is also a major factor.

There is also a high degree of subjectivity (both intra- and inter-observer) in the dental evaluations obtained from some classical methods. An example is the approaches of Nolla [6] and Dermijian *et al.* [13], which are not based on numerical quantification, but the assignment of an ordinal stage associated with dental maturity [6], [13].

Furthermore, the datasets used to initially develop and subsequently validate these methods had strict exclusion criteria, including the absence of teeth, endodontics or caries. The outcome is techniques that do not work with subjects who have any of these conditions.

In the last few years, Deep Learning (DL) techniques have been used to automate tasks involving OPG images. All of these techniques are able to extract rich features from the images automatically and for a variety of purposes, including: the segmentation of the dental region [39]; the detection and classification of individual teeth [40]–[42]; the detection of previous treatment, e.g., endodontics [43]; the reconstruction of OPG images where a patient was badly positioned [44]; and the diagnosis of osteoporosis [45] and jaw tumours [46].

In the current paper, a novel DL approach to chronological age estimation based on dental age is proposed as a way to overcome the issues of the classical methods mentioned previously. Our approach is completely automatic, uses raw OPG images without any preprocessing, and does not require dental measurements or other manual annotations to obtain accurate results.

TABLE I
AGE AND SEX DISTRIBUTION OF THE OPG DATASET AFTER DISCARDING IMAGES LABELLED WITH THE WRONG ACQUISITION OR BIRTH DATES

Age (years)	Male	Female	Total
4.5-9	210	206	416
10-19	474	489	963
20-29	172	332	504
30-39	64	81	145
40-49	35	51	86
50-59	35	49	84
60-69	28	30	58
70-89.2	12	19	31
Total	1030	1257	2289

II. MATERIALS AND METHODS

A set of 2289 OPG images from Spanish Caucasian subjects aged between 4.5 and 89.2 were provided to us by the School of Medicine and Dentistry, Universidade de Santiago de Compostela (Spain). The image collection was performed in accordance with the ethical standards of our institution's research committee and the 1964 Declaration of Helsinki and its later amendments [47]. The images were labelled with the subject's date of birth and the date the image was acquired, meaning that the chronological age was calculated in days. Table I shows the range of ages in the dataset, with most of the images corresponding to patients younger than 30. This is partly because clinical research has highlighted that the final developmental milestone for teeth is reached between the ages of 17 and 21, meaning that the maximum predictive capability is expected to be within that range. However, additional images from older subjects were included in the dataset to enable us to perform a number of experiments to confirm this belief.

The dataset was reviewed to remove images with incorrect acquisition or birth dates. Each image was analysed by a dental expert to determine if the appearance of the dentition matched the calculated chronological age. As the development of teeth varies considerably across the population, 38 of the most obvious cases (those where the calculated chronological ages were particularly discordant with expectations) were discarded.

We also considered the exclusion criteria applied to OPG images in earlier studies that used traditional chronological age-calculation methods [48]. Of our entire dataset,

Fig. 1. Example of an OPG image from a 9-year old boy. A rapid palatal expander is used to widen the upper jaw.

1752 images portrayed the presence of one or more of the following conditions: orthodontics, prostheses, implants, external elements like earrings, fillings, endodontics, caries, missing teeth, dental root remains, and distorted or blurred radiographs (see an example in Fig. 1). However, unlike the classical studies, we did not exclude these images, but instead labelled them as *imperfect*. The remaining 537 images are referred to as *perfect* in what follows.

The database consisted of high-resolution 8-bit images. Their height was always 1552 pixels, while the width varied from 2700 to 3200 pixels. The images were, however, downsized to 256x128 pixels, which is a reasonable resolution that enables dental experts to distinguish small structures in the oral cavity and provide an estimation of the age and sex of the subject. In this way, the images are somewhat smaller in size and the processing procedure is, consequently, more efficient.

The chronological age predictions were carried out using Convolutional Neural Networks (CNNs), which have demonstrated a convincing capacity to manage medical images [49]. They also represent an evolution of the classic multilayer perceptron that was developed to exploit the spatial relationships between the input values [50]. The architecture of a CNN for image processing varies greatly depending on the purpose, but does consist of the sequential application of 2D convolution kernels, each of which is followed by a non-linear activation layer. The kernels (also called filters) are learnable and represent specific features of the image. The output of applying a convolutional layer over an input image is referred to as a *feature map*, which represents the location of the learned feature in the image as a *heatmap*.

The convolutional layers are often alternated with downsizing (also called *pooling*) layers, which enables the network to learn image features at different scales. So, the first time a convolution kernel is applied, low-scale features like corners or points are learned. Then, in successive steps, these features are combined into higher-scale features to manage complex objects in the final layers. If the network is developed to perform classification or regression tasks, a fully connected network is appended [50].

In the present paper, two different network architectures were developed using the Keras toolkit [51]. The first of these, known as the DANet (Dental Age Net) in what follows (red items in Fig. 2), comprised a single path that worked at five different image scales. Two blocks composed of a 3×3 convolution layer and a ReLU non-linear activation function [52] were applied sequentially at each scale. The application of two consecutive 3×3 convolutions is widely thought to be a less complex replacement for a single 5×5 convolution. Although the former is not as powerful, the receptive field of the output neurons is the same and the performance is comparable. To move to the next scale, we applied a 2×2 *Max Pooling* layer to halve the resolution of the output by only preserving the maximum pixel in each 2×2 output block. No padding was added to the images in order to preserve the size of the feature maps after the convolutions, which cause a 2-pixel loss in every convolution step. As precise feature localisation was not the objective of this research, both training and inference speeds were improved without having an impact on overall performance. For the final scale, the second 3×3 convolution was replaced with a 2×2 version and the final *Max Pooling* layer by a *Global Average Pooling* layer (GAP). This produced the average value of each feature map. The values obtained were then combined in a weighted sum to predict a subject's age in days.

The second approach is known as DASNet (Dental Age and Sex Net), and has a more complex architecture and a main aim of improving age-estimation performances by incorporating sex-specific features. As the dental development pattern is different in boys and girls, a second path identical to DANet was added to the network in order to classify the images according to the gender of the subject (blue items in Fig. 2). The predictions of this path can be seen as the probability that a given image belongs to the positive class[1], and so the final weighted sum was followed by a sigmoid activation to force the output to lie within the range [0,1]. The goal of this second path was not the sex classification output itself, but the learning of sex-specific features. This is why the sex classification performance was not evaluated independently for the sex path, but for the entire DASNet model.

In the DASNet configuration (Fig. 2), the sex path shared information with the age path at intermediate points, forcing the latter to utilise the sex features. Specifically, the output of every pooling layer in the sex path was concatenated depth-wise to the output of the same pooling layer in the age path. Thereafter, the age path applied a convolutional layer to combine both age and sex information. This was also the case for the GAP layer in the sex path, the output of which was concatenated to the same layer of the age path and then combined in the final weighted sum. In this way, every time the sex path sent information to the age path, this path decided the degree of integration of the sex features.

Batch Normalisation was used after every non-linearity to increase training speed and global performance of the method [53] (see Fig. 2). This normalisation approach was applied to each training mini-batch and consisted of two steps. First, a standard normalisation mapped the input, meaning that the data distribution had zero mean and unit standard deviation. Second, as this can affect the expressive power

[1]In this work, it is assumed that the male sex is the negative class (encoded as 0s) and female sex the positive class (encoded as 1s).

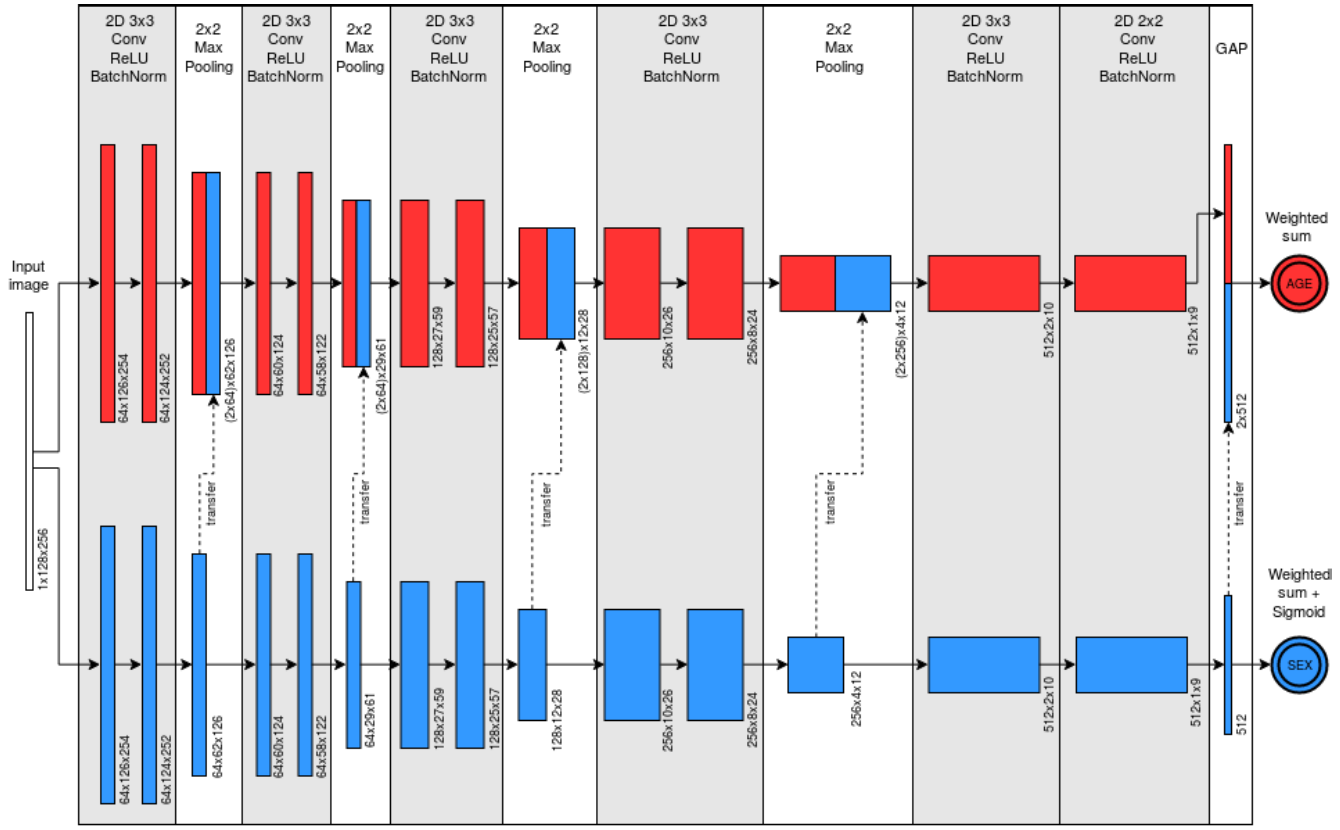


Fig. 2. DASNet architecture. In red, the age-prediction path representing the DANet (Dental Age Net) approach. This consists of a set of convolutional and pooling layers with which to extract image features at different scales and, ultimately, determine a subject's age in days. In blue, the gender classification path, which complements the DANet approach and leads to the development of the DASNet (Dental Age and Sex Net) approach. This imitates the architecture of the age path to classify the sex of the subject. At different points, the age path receives information from the sex path and integrates it before the next layer.

of the network, the distribution was shifted and scaled by the learnable parameters α and β, respectively, ensuring that it was always possible to recover the original distribution if necessary. The application of this normalisation method added more parameters to the network, and so the model became more complex. Nevertheless, as the network had to manage less variability in the images, it was able to learn faster and produced better results.

III. Experiments and Performance Evaluation

In order to evaluate the performance of the DANet and DASNet architectures, the training approach needed to be defined.

The learnable network parameters were initialised with the Xavier uniform initialiser [54], which uses a uniform distribution within custom limits to improve the convergence of the training process. In every training iteration, the network was fed a batch of images, and the prediction error was calculated through the loss function $\mathcal{L}$. The bigger the batch size, the faster the convergence, although there is a potential risk of becoming stuck in a loss function local minimum. On the other hand, the smaller batch sizes are noisier and have greater variability, meaning that the network learns more slowly. It can, however, manage extreme images and the results will, ultimately, be better. In our research, the accuracy of the network was prioritised over the training speed, and a small batch size of four images was chosen for this reason.

Every time a network was fed with a batch of images, the network weights had to be updated. Using the so-called backpropagation method, the loss function $\mathcal{L}$ was propagated back through the network to calculate the contribution of each part of the network to that loss. The network weights of each layer were then updated according to that contribution and the learning rate, which can vary during the training process. Although there are many ways to update the weights, we chose the Adadelta optimiser [55] because it adapts the learning rate automatically, meaning that the network can learn faster at the start. It can then make smaller steps forwards to fine-tune the model and avoid local minima, without the need to set this behaviour change manually. The decay factor was set at 0.95.

The number of times the network could traverse the whole dataset (also called *epochs*) was set at 200, although the training was configured to stop earlier if the results did not improve for 40 successive epochs. The 8-fold Cross-Validation (CV) method was used to prevent the overfitting of the network to the input dataset (Fig. 3). First, the dataset was split into eight parts and two different datasets were built in each CV iteration: the training/validation dataset, which was composed of seven parts; and the test dataset, comprising the eighth part. The first of these parts was in turn divided into two other parts:

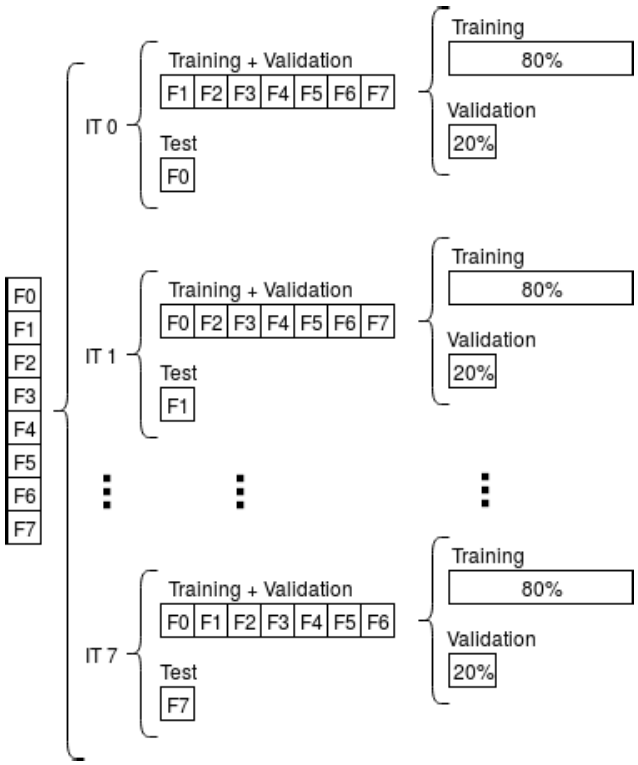


Fig. 3. Cross-validation approach defined to assess the performance of DANet and DASNet. First, the dataset is split into eight folds (from *F0* to *F7*). Then, for each cross-validation iteration (from *IT0* to *IT7*), a fold is omitted for testing of the model and the rest of the folds as a whole are divided into training (80%) and validation (20%) sets.

training (80%) and validation (20%). The training set was then fed into the network in batches composed of four images, as explained previously in this section. Finally, the network performance could be assessed in the union of the eight test datasets obtained by CV, which corresponded to the entire dataset. It is worth noting that all of the dataset divisions were conducted by preserving the distribution of the ages and the proportion of males and females, meaning that the training, evaluation and test datasets had the same characteristics.

Every batch was augmented to increase the variability of the dataset and, therefore, improve the generalisation capabilities of the network. As shown in Table II, the images were slightly perturbed in relation to: translation in both axes, rotation, contrast and brightness changing, and zooming. The brightness changing was performed according to the Power Law Transformation, where each pixel value in an image is raised to a given factor. The contrast changing was conducted according to the formula $f \cdot (I - 0.5) + 0.5$, where f is the changing factor and I the input image. After the transformation, pixel values under 0 or over 1 were clipped to preserve the [0,1] range. In the zooming operation, the image was scaled up or down by a factor and placed in the centre of the image frame. As a consequence, zoomed-in images were cropped at the borders and zoomed-out images had to be padded with 0s to preserve the original size. The factors used in each transformation were randomly sampled according to a uniform distribution with the limits and probabilities shown in Table II.

The loss function varied depending on the output of the network. The root mean squared error (RMSE) loss function

TABLE II
DATA AUGMENTATION TRANSFORMATIONS USED TO IMPROVE THE TRAINING SET. THE FACTOR COLUMN GIVES THE LIMITS OF THE DISTRIBUTION USED TO SAMPLE THE TRANSFORMATION PARAMETERS. THE PROBABILITY COLUMN REPRESENTS THE PROPORTION OF THE BATCH SIZE THAT IS AFFECTED BY A PARTICULAR TRANSFORMATION

Transformation	Factor	Probability
Horizontal flip	-	0.5
Translation X	(-10,10) pixels	0.9
Translation Y	(-8,8) pixels	0.9
Rotation	(-1,1) degrees	0.9
Contrast	(0.8,1.2)	0.9
Brightness	(0.8,1.2)	0.9
Zoom	(0.9,1.1)	0.2

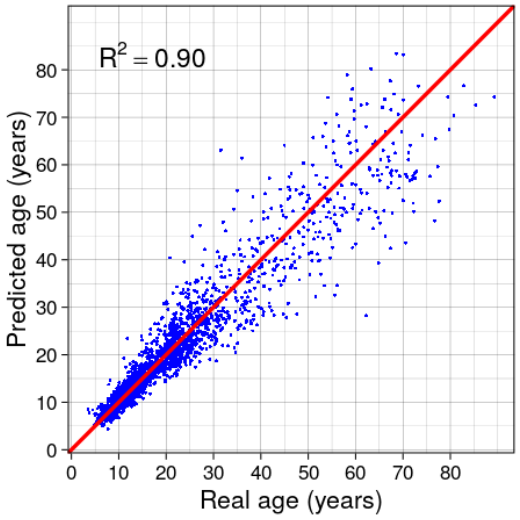


Fig. 4. Real age versus the DASNet predictions evaluated for the entire dataset, S_{all}. The closer the points are to the red line, the better the performance.

$\mathcal{L}_{DANet}$ was used in the DANet, as shown in Eq. (1), where θ are the network parameters, a is the true age, $\hat{a}$ is the predicted age, and n is the number of examples in the training batch (four in this case). In the DASNet loss function $\mathcal{L}_{DASNet}$ (Eq. (2)), the RMSE loss was used for the age regression and the log-loss for the sex classification, where s is the true sex encoded as 1 or 0 and $\hat{s}$ is the predicted probability of belonging to the positive class. As age and sex were in different scales, the age RMSE was multiplied by a factor of $w = 10^{-3}$ to balance out the impact of both losses in the $\mathcal{L}_{DASNet}$.

$$\mathcal{L}_{DANet}(\theta) = \sqrt{\frac{\sum_{i=1}^{n}(a_i^2 - \hat{a}_i^2)}{n}} \tag{1}$$

$$\mathcal{L}_{DASNet}(\theta) = w \cdot \sqrt{\frac{\sum_{i=1}^{n}(a_i^2 - \hat{a}_i^2)}{n}} - \frac{1}{n}\sum_{i=1}^{n}[s_i \cdot \log(\hat{s}_i) + (1 - s_i) \cdot \log(1 - \hat{s}_i)] \tag{2}$$

A set of regression metrics were computed to assess the performance of both the DANet and DASNet. The median of

TABLE III
PERFORMANCE COMPARISON BETWEEN DANET (AGE PATH) AND DASNET (AGE AND SEX PATHS). E: ERROR; AE: ABSOLUTE ERROR; MED.: MEDIAN; IQR: INTERQUARTILE RANGE; P99: 99TH PERCENTILE; R^2: COEFFICIENT OF DETERMINATION; ACC: ACCURACY; AUC: AREA UNDER THE ROC CURVE. THE ERROR METRICS ARE GIVEN IN YEARS

		DANet						**DASNet**									
		Age regression						**Age regression**						**Sex classification**			
		E	**AE**				R^2	**E**	**AE**				R^2				
Dataset	**#Img**	**med.**	$\mu \pm \sigma$	**med.**	**IQR**	**p99**		**med.**	$\mu \pm \sigma$	**med.**	**IQR**	**p99**		**Acc**	**Sens.**	**Spec.**	**AUC**
S_{all}	2289	0.44	3.19 ± 4.32	1.66	3.01	21.09	0.87	0.12	2.84 ± 3.75	1.48	2.70	18.16	0.90	0.854	0.878	0.823	0.925
$S_{<40}$	2030	0.15	1.97 ± 2.26	1.19	1.97	10.80	0.85	0.04	1.80 ± 2.12	1.10	1.75	10.48	0.87	0.848	0.851	0.845	0.926
$S_{<30}$	1885	0.08	1.52 ± 1.54	1.04	1.57	6.85	0.88	0.06	1.43 ± 1.44	0.96	1.50	6.51	0.89	0.826	0.847	0.800	0.908
$S_{<25}$	1718	0.09	1.21 ± 1.17	0.87	1.20	5.53	0.89	0.07	1.17 ± 1.11	0.85	1.18	5.34	0.90	0.825	0.830	0.820	0.904
$S_{<20}$	1381	-0.04	0.93 ± 0.80	0.72	1.02	3.78	0.88	0.02	0.89 ± 0.77	0.69	0.95	3.45	0.89	0.800	0.800	0.801	0.888
$S_{<15}$	1087	-0.07	0.78 ± 0.65	0.64	0.81	2.91	0.82	-0.07	0.75 ± 0.57	0.64	0.72	2.58	0.84	0.750	0.722	0.773	0.829

the prediction error ($E(med.)$) was calculated by subtracting the predicted values from the real ones, and computing their median value. It was beneficial to determine if the network was prone to over- (values lower than 0) or under-estimation (values greater than 0). It should be noted that the use of the median as a measure for prediction error was justified by the asymmetry and unimodality features found in our observed and predicted data. These features were assessed via the tests proposed in [56] and [57], jointly with some exploratory analysis, as further described in the Supplementary Material. First, the asymmetry observed in the real data was preserved in the predictions made by the two networks (according to the MGG symmetry test [56]). The mean error was therefore not a suitable assessment approach. Second, permutation tests were carried out to assess whether there were significant differences between the median values of the original data and the outputs of the networks. P-values with larger than usual significance levels were obtained, meaning that there were no relevant differences. However, the median error can be closer to 0 if there is a similar number of negative and positive prediction errors. Accordingly, the absolute error (AE) was computed to measure the magnitude of the errors, which was summarised with the mean and standard deviation ($AE(\mu \pm \sigma)$), the median ($AE(med.)$), the interquartile range ($AE(IQR)$) and the 99th percentile ($AE(p99)$). While the mean and median provide a general idea of the age-estimation performance, the standard deviation, interquartile range and 99th percentile are useful for assessing the impact of the poorest estimated images. The coefficient of determination (R^2) was also calculated to measure the correlation between the real and predicted values and, therefore, the proportion of the age variance explained by the model.

The sex classification performance was evaluated to assess if the DASNet was learning the sex-specific features effectively. As the output of the sex path was a continuous output between 0 and 1, a classification threshold of 0.5 was used to determine the output class and calculate a set of binary classification metrics. The accuracy concerns the proportion of well-classified images, while the sensitivity and specificity represent the ratio of correctly classified images from each class. This value was changed to the range [0,1] and the sensitivity and specificity values were calculated for each value to assess the sex-classification performance without depending on a specific classification threshold. It was then possible to produce a curve between these two metrics (ROC curve). The area under this curve (AUC) was added to the performance metrics, as this approach is widely used to test the robustness of classification methods.

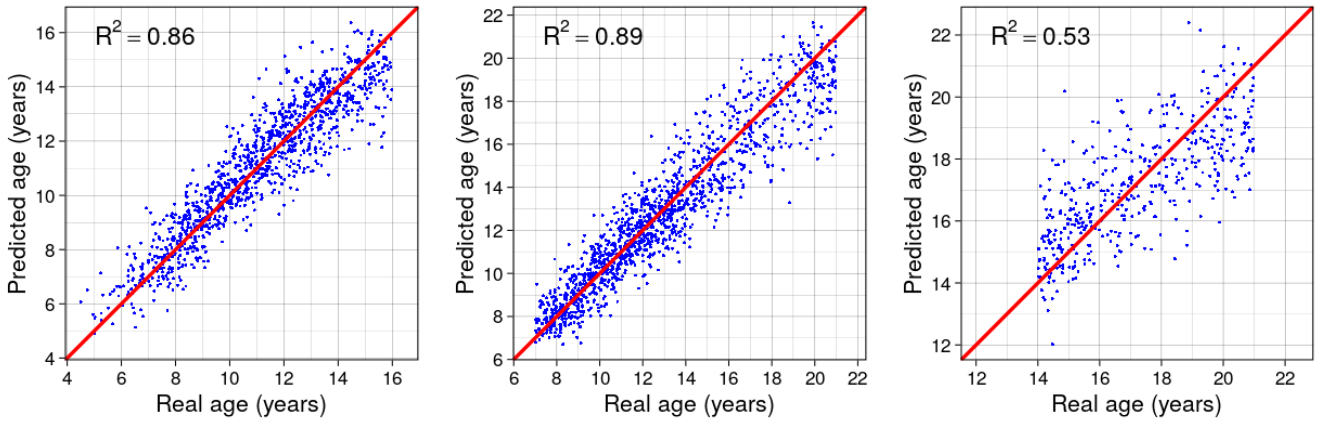
In addition to the original dataset S_{all}, reduced datasets $S_{<40}$, $S_{<30}$, $S_{<25}$, $S_{<20}$ and $S_{<15}$ were generated for subjects younger than 40, 30, 25, 20 and 15, respectively. The DANet and DASNet were trained and tested on each dataset, meaning that the age-prediction performance could be assessed at different levels.

In order to compare the age-estimation performance of the DASNet to other methods, we examined several studies that used a manual approach in a similar population (in racial terms). Reduced versions of the entire dataset were generated to match the age range of the subjects in each study, and different metrics such as E means and medians, and their respective dispersion measures, and AE values were determined.

Finally, the behaviour of the network was analysed to determine the regions of the image that it focused on to estimate chronological age, and this was then checked to see if it matched the opinions of the dental experts in the clinical field. The Grad-CAM method [58] was used to build image attention maps. The objective of the method is to feed a network with an image, calculate the gradients of the target (in this case, a linear layer to predict age), and flow them into a previous convolutional layer to provide a heatmap of the image regions that contribute the most to increasing or decreasing the output.

IV. RESULTS

As shown in Table III, there was generally a high correlation between the ages estimated by the DANet and DASNet and the real ages ($R^2 = 0.87$ and 0.90, respectively). The DASNet provided better results than the DANet in the S_{all} dataset, as the mean AE was about 2 years and 10 months, giving a



(a) Between 4 and 16 years old (b) Between 7 and 21 years old (c) Between 14 and 21 years old

Fig. 5. Real age versus the DASNet predictions for some reduced datasets generated for the subsequent comparison with manual methods. The red line shows the ideal performance. Ranges [4-17], [3.9-15.4], [6-13] and [5-15] are not displayed due to their great similarity with range [4-16].

TABLE IV

COMPARISON OF MANUAL METHODS FOR DENTAL AGE-ESTIMATION (WHITE CELLS) WITH THE DASNET APPROACH (GREY CELLS). E: ERROR (REAL - PREDICTED); IQR: INTERQUARTILE RANGE; AE: ABSOLUTE ERROR. THE ERROR METRICS ARE GIVEN IN YEARS. ALL THE PRESENTED METHODS RELY ON MANUAL MEASUREMENTS OR CLASSIFICATIONS EVERY TIME THE AGE OF A SUBJECT NEEDED TO BE ESTIMATED

						Metrics									
Work	**Method**	**Age**	**Population**	**#Images**		**E(μ)**		**E(σ)**		**E(median)**		**E(IQR)**		**AE(μ)**	
[60]	Demirjian	14-21	Spanish	1054	446	-0.1 / -0.07 (1)	0.04	1.23 / 1.22 (1)	1.49	-	0.02	-	1.93	-	1.19
[37]	Demirjian	7-21	Spanish	2641	1381	0.853	0.12	0.389 / 0.564 (2)	1.20	-	-0.07	-	1.38	-	0.91
	Nolla					-0.213		0.496 / 0.230 (2)		-		-		-	
[62]	Demirjian	4-17	Spanish	308	1223	0.76 / 0.88 (2)	0.10	1.01 / 1.09 (2)	1.00	-	-0.07	-	1.25	-	0.78
	Chaillet					0.37 / 0.21 (2)		1.04 / 1.07 (2)		-		-		-	
[38]	Cameriere	4-16	European Caucasian	2652	1160	-	-0.02	-	0.97	-0.114	-0.06	1.22	1.21	-	0.75
[61]	Häävikko	3.9-15.4	Italian	500	1119	-0.35	0.06	0.72	0.94	-	0.008	-	1.21	-	0.73
	Cameriere					-0.02 / 0.1 (2)		0.71 / 0.71 (2)		-0.04 / 0.09 (2)		-		0.55 / 0.53 (2)	
[26]	Häävikko	6-13	Italian	1089	835	-0.09 / -0.23 (2)	-0.01	0.79 / 0.73 (2)	0.80	-0.14 / -0.23 (2)	-0.04	-	1.05	0.62 / 0.59 (2)	0.63
	Willems					0.42 / 0.25 (2)		0.77 / 0.89 (2)		0.34 / 0.15 (2)		-		0.67 / 0.69 (2)	
	Cameriere		Italian,			-		-		0.059		0.709		0.488	
[59]	Demirjian	5-15	Spanish	756	1082	-	0.02	-	0.92	-0.7	-0.02	1.5	1.18	1.076	0.72
	Willems		& Croatian			-		-		-0.072		1.590		0.933	

(1) Depending on the specific tooth used for age estimation (left third molar / right third molar)

(2) Male / female

median E of +0.12 years. When evaluating and comparing both networks in relation to the reduced datasets, the AE values decreased as the subjects' real ages fell, with these values lower in the DASNet than the DANet in all cases. In the $S_{<15}$ dataset, the mean AE was 0.78 years for the DANet and 0.75 years for the DASNet, with median AE values of 0.64 years in both networks. The AE standard deviation, IQR and p99 for this dataset also systematically decreased as the real age fell, reaching values of 0.65, 0.81 and 2.91, respectively, in the DANet and 0.57, 0.72 and 2.58 in the DASNet. As better performance results were obtained from the DASNet, the following analyses focus on this model.

In Fig. 4, which was generated for the DASNet predictions, the level of correlation was much higher in subjects younger than 20. In the reduced datasets generated for the subsequent comparison with other classical methods, the performance of the automatic approach was better for subjects younger than 14 or 15 (as seen in Fig. 5). This started to worsen with older subjects, with this remarkable result displayed in Fig. 5c. In this final plot, the network did not seem to have enough of the information it required to estimate the age.

The sex-classification results of the DASNet are also presented in Table III. The top accuracy (Acc) of 85.4% was achieved in the S_{all} dataset, with AUC above 0.92, while the top error was in the $S_{<15}$ dataset, where the accuracy remained below 80%. The values of the sensitivity and specificity metrics showed that the images of females were classified better in S_{all}, $S_{<40}$, $S_{<30}$ and $S_{<25}$ (higher sensitivity), while

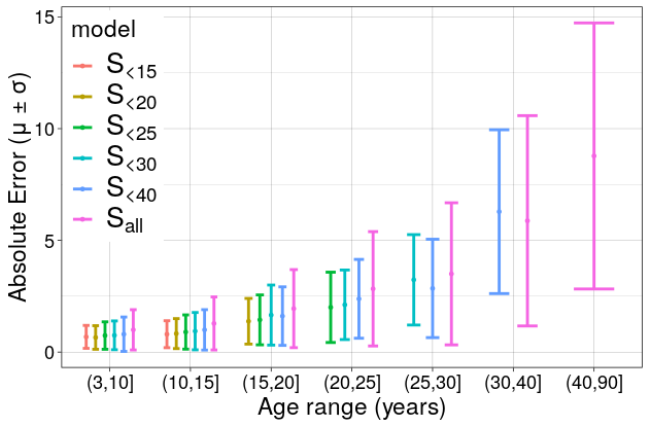


Fig. 6. DASNet absolute error in the chronological age estimation when training with reduced datasets and evaluating performance for different age ranges.

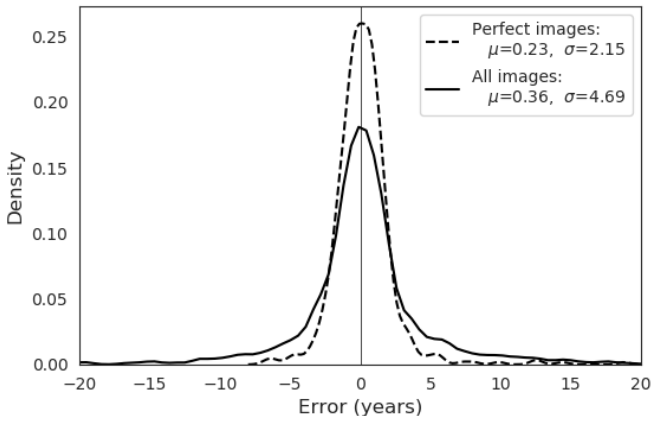
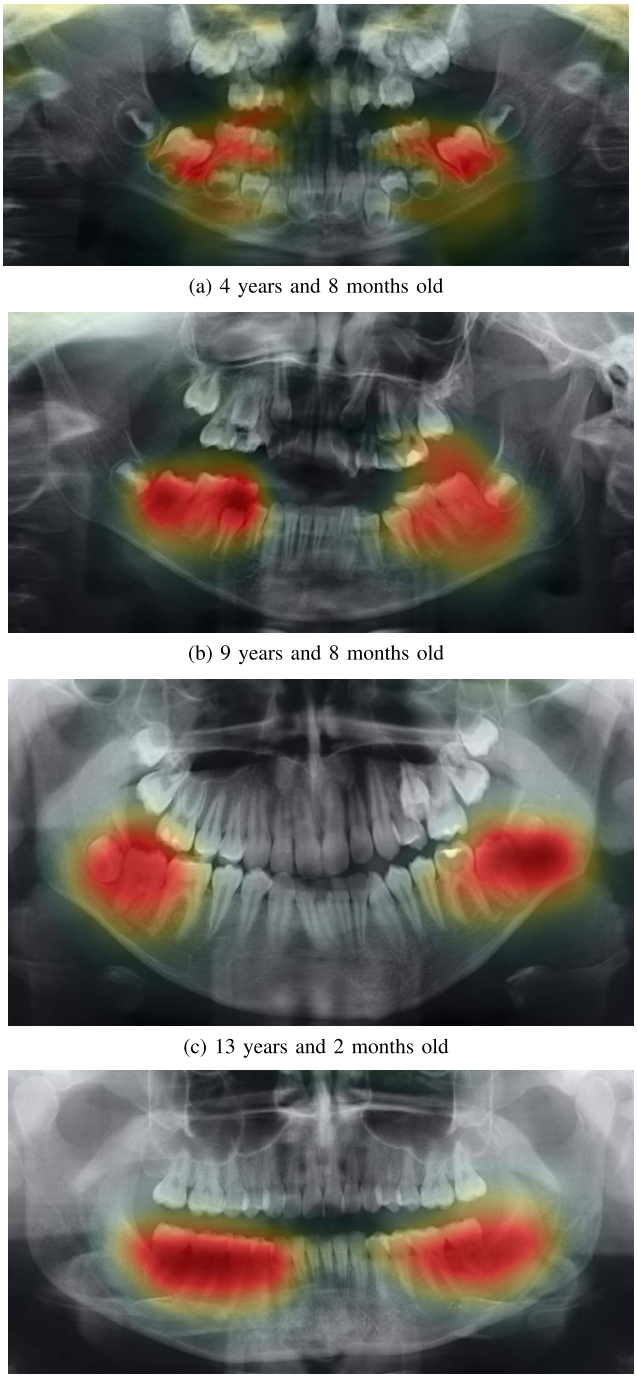


Fig. 7. Distribution of the DASNet errors (true age - predicted age), shown as the density estimation, in all the images versus the *perfect* images (those which do not present any conditioning radiological or dental characteristic, as explained in section II).

those of males were categorised better in $S_{<20}$ and $S_{<15}$ (higher specificity).

As highlighted in Fig. 6, at the intra-dataset level for any training age-range, the models derived from the groups of subjects that showed a greater age amplitude produced higher mean AE values than the models where the age amplitude was narrower. For example, using the S_{all} and $S_{<15}$ models to evaluate the results in the age range [10,15] produced mean AE values of 1.28 ± 1.18 years and 0.80 ± 0.60 years, respectively. At the inter-dataset level, the DASNet's age-estimation performance started to worsen for all the models when older subjects were taken into account in the training dataset, with the AE values rising and the distribution of the errors widening. For example, using the S_{all} model to evaluate the results in the range [10,15] produced a mean AE of 1.28 ± 1.18 years, with these values rising up to 5.87 ± 4.70 years in the age range [30,40].

In relation to the DASNet's performance on *perfect* OPG images with respect to the entire dataset (Fig. 7), the mean E in the first was closer to 0 and the distribution was concentrated more around the average than for the values calculated for the latter ($E = 0.23 \pm 2.15$ years, in contrast to 0.36 ± 4.69 years).



(a) 4 years and 8 months old

(b) 9 years and 8 months old

(c) 13 years and 2 months old

(d) 25 years and 2 months old

Fig. 8. Example of Grad-CAM attention maps at different ages. The redder the region, the more attention paid to it by the network.

Table IV details the comparisons of the manual dental age-estimation methods to the DASNet approach, with the results expressed in years. In total, thirteen comparisons were made for seven series of subjects within an age range of 3.9 to 21. The most commonly used metrics were the E means and the corresponding standard deviations. As can be seen in the table, and considering at least one of the metrics analysed, the DASNet approach produced better results than the manual methods in all the comparisons. In terms of the AE values, six comparisons were possible and, in half of them, the DASNet

approach produced values that were lower than those found in the previous series (the mean AE of DASNet = 0.63 vs 0.67/0.69 for the Willems method [26]; the mean AE of the DASNet = 0.72 vs 1.076 and 0.933 for the Dermijian and Willems methods, respectively [59]). In terms of the E means, in the nine comparisons carried out, the DASNet approach surpassed all the manual methods, with its values lower than those detected in the different series (mean E range of the DASNet vs. the manual methods = −0.01, 0.12 vs. −0.35, 0.88) [26], [60]–[62]. These observations were corroborated in the comparisons in which E medians were used [26], [38], [59]. Focusing on the dispersion measurements, and specifically the standard deviations, six of the nine comparisons revealed that the DASNet was associated with higher values than those detected by the manual methods, with these ranging from 0.80 to 1.49 compared to 0.23 to 1.23.

With regard to the network behaviour, the attention regions observed with Grad-CAM (see Fig. 8) varied significantly across all the images, but it can be seen that the posterior teeth of the inferior arch usually contribute more to age prediction. This behaviour was observed when feeding the network with images of subjects younger than 25.

V. Discussion and Conclusions

Establishing someone's chronological age using their dental age is a topic of great importance in the clinical dental field, and is reflected in the high number of studies published on the topic in the literature [34], [35]. However, expert dentists are well aware of the limitations of the classical methods, such as the high time-demands of the procedure and the considerable subjectivity of some non-quantitative approaches. Consequently, the development of an automated method through the application of DL techniques is of great interest in the clinical dental field, as it could bring about substantial improvements in terms of time and subjectivity.

This paper presents two novel DL approaches to estimating the age of a subject from an OPG image. The first, DANet, consists of a single path where convolution and pooling layers are interleaved to learn image features at different scales and, ultimately, estimate age. The second method, DASNet, adds a second path, nearly identical to the first, to estimate sex. The sex path shares information with the age path to force it to use sex-specific features at intermediate points. In this way, sex information is taken into account to improve the age estimation process. To the best of our knowledge, this work is the first to use DL techniques for the automatic calculation of chronological age from OPG images.

In all the examined circumstances, the DASNet outperformed the DANet, with the results being especially remarkable when the former was trained with the $S_{<15}$ dataset: the mean and median AE remained at 0.75 and 0.64 years (about 9 and 8 months, respectively) and the standard deviation stayed below 0.6 years (about 7 months), producing a very narrow error distribution. As expected, the performance of both networks worsened when older people were introduced into the experiment, confirming the clinical fact that, for those older than 25, which is when the permanent teeth are completely formed, it is difficult to assess age accurately [60]. Our results also revealed that, as an extreme case, a model trained with the entire dataset was less accurate when estimating the age of a child than one trained with subjects younger than 15. Although this may seem obvious, it reinforces the need to adapt the model to be as accurate as possible depending on the target population.

Focusing on the DASNet, the sex-classification task was evaluated using the same datasets. The AUC values, which went above 0.92 in the best case, also demonstrated that the sex estimations were not biased towards one sex or another. The accuracy was around 83-85% in the datasets S_{all}, $S_{<40}$, $S_{<30}$, while the top error was in the $S_{<15}$ dataset, where the accuracy was 75%. These results are in line with [63], [64], and improve those obtained in [65], [66].

Interestingly, the sex classification behaved in the opposite way to the age estimation. While the age regressor performed worse in older people, because it could not identify relevant age features, the sex classification tended to achieve better results when older people were included. These findings have a clear biological explanation: from the age of 25, there are no revealing anatomical variations in an image that are particularly associated with the process of dental maturation, enabling us to reliably predict chronological age using an automated procedure; that is, the maturational dental age is no longer correlated with the chronological age. In contrast, the images of subjects from the age of 25 upwards contain relevant anatomical characteristics related to size in the mature state (anthropometric measurements of maxillary bones and teeth), which makes it possible to reliably predict the gender of a subject using an automated procedure. These findings are consistent with those obtained by other authors, who used different odontological sex-estimation methods [67].

Compared to the state of the art [26], [37], [38], [59]–[62], and by taking the mean and median errors closer to zero, the DASNet approach reduces the problem of under- or over-estimation, which is fairly common in the classical manual methods [26], [37], [60]–[62]. Furthermore, the mean absolute error $AE(\mu)$ was lower than that obtained with the Demirjian and Willems methods in [59] and with the Willems approach in [26]. In contrast, the standard deviations of the errors $E(\sigma)$ of the DASNet were greater than in several series that used different manual methods [26], [37], [60], [61]. This is because poor quality radiological images, as well as images that included conditioning dental characteristics, were taken into account in our research, which affected the distribution of the errors. However, it is notable that a fully automatic method produced a prediction error that was, in most cases, better than the manual methods, even when working with images with technical issues.

The behaviour of the DASNet was clinically consistent and in line with the manual age-estimation methods that rely on the posterior mandibular teeth to build up their prediction models [8], [13].

These results are very satisfactory, although there is still room for improvement, particularly in terms of the origin of the image set, which came from a single acquisition device with a specific configuration. Another consideration to bear

in mind is that the comparison with manual methods for estimating chronological age based on the state of the art was a first approximation. Our next objective is to confirm the findings derived from this initial assessment in a subsequent clinical study. This study will use the same set of OPG images, with the same technical and clinical characteristics, to evaluate the diagnostic accuracy of the chronological age estimation by simultaneously applying traditional manual methods and the network we have developed. In this sense, a multisource image set would not only enable a fairer comparison with other dental age-estimation methods, but also the design of a more complex network architecture that can handle greater image variability and will, therefore, produce even more reliable results.

In conclusion, the DASNet method can be used to automatically and accurately predict someone's chronological age, especially in young subjects with developing dentitions.

ACKNOWLEDGMENT

The authors would like to thank R. L. Gibson for the initial English proofreading and R. M. Crujeiras for her help with the statistical proofs. The authors have no relevant conflicts of interest to disclose.

REFERENCES

[1] W. W. Greulich, S. I. Pyle, and T. W. Todd, *Radiographic Atlas of Skeletal Development of the Hand and Wrist*. Palo Alto, CA, USA: Stanford Univ. Press, 1959, vol. 2.

[2] J. M. Tanner *et al.*, *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*, vol. 16. London, U.K.: Academic, 1975.

[3] A. F. Roche, D. Thissen, and W. Chumlea, *Assessing the Skeletal Maturity of the Hand-Wrist: Fels Method*. Springfield, IL, USA: Thomas, 1988.

[4] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi, "Deep learning for automated skeletal bone age assessment in X-ray images," *Med. Image Anal.*, vol. 36, pp. 41–51, Feb. 2017.

[5] B. Liang *et al.*, "A deep automated skeletal bone age assessment model via region-based convolutional neural network," *Future Gener. Comput. Syst.*, vol. 98, pp. 54–59, Sep. 2019.

[6] C. Nolla, "The development of the human dentition," *ASDC J. Dent Child*, vol. 27, no. 27, pp. 66–254, 1960.

[7] C. F. Moorrees, E. A. Fanning, and E. E. Hunt, "Age variation of formation stages for ten permanent teeth," *J. Dent Res.*, vol. 42, no. 6, pp. 1490–1502, Nov. 1963.

[8] R. Cameriere, L. Ferrante, and M. Cingolani, "Age estimation in children by measurement of open apices in teeth," *Int. J. Legal Med.*, vol. 120, no. 1, pp. 49–52, Jan. 2006.

[9] K. Haavikko, "Tooth formation age estimated on a few selected teeth. A simple method for clinical use," *Proc. Finn Dent Soc.*, vol. 70, no. 1, pp. 9–15, 1974.

[10] K. Haavikko, "The formation and the alveolar and clinical eruption of the permanent teeth. An orthopantomographic study," *Suom Hammaslaak Toim*, vol. 66, no. 3, p. 103, 1970.

[11] I. Gleiser and E. E. Hunt, Jr., "The permanent mandibular first molar: Its calcification, eruption and decay," *Amer. J. Phys. Anthropol.*, vol. 13, no. 2, pp. 253–283, 1955.

[12] A. B. Lewis and S. M. Garn, "The relationship between tooth formation and other maturational factors," *Angle Orthodontist*, vol. 30, no. 2, pp. 70–77, 1960.

[13] A. Demirjian, H. Goldstein, and J. Tanner, "A new system of dental age assessment," *Hum. Biol.*, vol. 45, no. 2, pp. 211–227, May 1973.

[14] A. Demirjian and H. Goldstein, "New systems for dental maturity based on seven and four teeth," *Ann. Hum. Biol.*, vol. 3, no. 5, pp. 411–421, Jan. 1976.

[15] U. H Gg and L. Matsson, "Dental maturity as an indicator of chronological age: The accuracy and precision of three methods," *Eur. J. Orthodontics*, vol. 7, no. 1, pp. 25–34, Feb. 1985.

[16] M. Nyström, J. Haataja, M. Kataja, M. Evalahti, L. Peck, and E. Kleemola-Kujala, "Dental maturity in Finnish children, estimated from the development of seven permanent mandibular teeth," *Acta Odontologica Scandinavica*, vol. 44, no. 4, pp. 193–198, Jan. 1986.

[17] V. Staaf, H. Mörnstad, and U. Welander, "Age estimation based on tooth development: A test of reliability and validity," *Eur. J. Oral. Sci.*, vol. 99, no. 4, pp. 281–286, Aug. 1991.

[18] S. Koshy and S. Tandon, "Dental age assessment: The applicability of Demirjian's method in South Indian children," *Forensic Sci. Int.*, vol. 94, nos. 1–2, pp. 73–85, Jun. 1998.

[19] R. Nykänen, L. Espeland, S. I. Kvaal, and O. Krogstad, "Validity of the Demirjian method for dental age estimation when applied to Norwegian children," *Acta Odontologica Scandinavica*, vol. 56, no. 4, pp. 238–244, Jan. 1998.

[20] C. Farah, D. Booth, and S. Knott, "Dental maturity of children in Perth, Western Australia, and its application in forensic age estimation," *J. Clin. Forensic Med.*, vol. 6, no. 1, pp. 14–18, Mar. 1999.

[21] M. Bolaños, M. Manrique, M. Bolaños, and M. Briones, "Approaches to chronological age assessment based on dental calcification," *Forensic Sci. Int.*, vol. 110, no. 2, pp. 97–106, May 2000.

[22] C. Mckenna, H. James, J. Taylor, and G. Townsend, "Tooth development standards for South Australia," *Austral. Dental J.*, vol. 47, no. 3, pp. 223–227, Sep. 2002.

[23] A. Olze, D. Bilang, S. Schmidt, K.-D. Wernecke, G. Geserick, and A. Schmeling, "Validation of common classification systems for assessing the mineralization of third molars," *Int. J. Legal Med.*, vol. 119, no. 1, pp. 22–26, Jan. 2005.

[24] M. Maber, H. Liversidge, and M. Hector, "Accuracy of age estimation of radiographic methods using developing teeth," *Forensic Sci. Int.*, vol. 159, pp. S68–S73, May 2006.

[25] S. Martin-de las Heras, P. García-Fortea, A. Ortega, S. Zodocovich, and A. Valenzuela, "Third molar development according to chronological age in populations from Spanish and Magrebian origin," *Forensic Sci. Int.*, vol. 174, no. 1, pp. 47–53, Jan. 2008.

[26] I. Galić *et al.*, "Accuracy of Cameriere, Haavikko, and Willems radiographic methods on age estimation on Bosnian–Herzegovian children age groups 6–13," *Int. J. Legal Med.*, vol. 125, no. 2, pp. 315–321, Mar. 2011.

[27] S. Hegde, A. Patodia, and U. Dixit, "Staging of third molar development in relation to chronological age of 5–16 year old Indian children," *Forensic Sci. Int.*, vol. 269, pp. 63–69, Dec. 2016.

[28] Y. Zhai, H. Park, J. Han, H. Wang, F. Ji, and J. Tao, "Dental age assessment in a northern Chinese population," *J. Forensic Legal Med.*, vol. 38, pp. 43–49, Feb. 2016.

[29] M. Rivera, S. De Luca, L. Aguilar, L. A. V. Palacio, I. Galić, and R. Cameriere, "Measurement of open apices in tooth roots in Colombian children as a tool for human identification in asylum and criminal proceedings," *J. Forensic Legal Med.*, vol. 48, pp. 9–14, May 2017.

[30] M. E. Berkvens, S. I. Fairgrieve, and S. Keenan, "A comparison of techniques in age estimation using the third molar," *Can. Soc. Forensic Sci. J.*, vol. 50, no. 2, pp. 74–83, Apr. 2017.

[31] S. M. Garn, A. B. Lewis, and D. L. Polacheck, "Variability of tooth formation," *J. Dent Res.*, vol. 38, no. 1, pp. 135–148, Jan. 1959.

[32] H. Liversidge, "Variation in modern human dental development," in *Cambridge Studies in Biological and Evolutionary Anthropology*. Cambridge, U.K.: Cambridge Univ. Press, 2003, pp. 73–113.

[33] R. Cameriere, L. Ferrante, D. De Angelis, F. Scarpino, and F. Galli, "The comparison between measurement of open apices of third molars and Demirjian stages to test chronological age of over 18 year olds in living subjects," *Int. J. Legal Med.*, vol. 122, no. 6, pp. 493–497, Nov. 2008.

[34] M. A. Bittencourt, G. O. Cericato, A. Franco, R. S. Giràò, A. P. Lima, and L. R. Paranhos, "Accuracy of dental development for estimating the pubertal growth spurt in comparison to skeletal development: A systematic review and meta-analysis," *Dentomaxillofacial Radiol.*, Dec. 2017, Art. no. 20170362.

[35] M. Y. P. M. Yusof, I. W. Mokhtar, S. Rajasekharan, R. Overholser, and L. Martens, "Performance of Willem's dental age estimation method in children: A systematic review and meta-analysis," *Forensic Sci. Int.*, vol. 280, pp. 245.e1–245.e10, Nov. 2017.

[36] P. Kapoor and V. Jain, "Comprehensive chart for dental age estimation (DAEcc8) based on Demirjian 8-teeth method: Simplified for operator ease," *J. Forensic Legal Med.*, vol. 59, pp. 45–49, Oct. 2018.

[37] M. Melo and J. Ata-Ali, "Accuracy of the estimation of dental age in comparison with chronological age in a Spanish sample of 2641 living subjects using the Demirjian and Nolla methods," *Forensic Sci. Int.*, vol. 270, pp. 276.e1–276.e7, Jan. 2017.

[38] R. Cameriere, D. De Angelis, L. Ferrante, F. Scarpino, and M. Cingolani, "Age estimation in children by measurement of open apices in teeth: A European formula," *Int. J. Legal Med.*, vol. 121, no. 6, pp. 449–453, Oct. 2007.

[39] G. Silva, L. Oliveira, and M. Pithon, "Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives," *Expert Syst. Appl.*, vol. 107, pp. 15–31, Oct. 2018.

[40] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, "Deep instance segmentation of teeth in panoramic X-ray images," in *Proc. 31st SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Oct. 2018, pp. 400–407.

[41] A. B. Oktay, "Tooth detection with convolutional neural networks," in *Proc. Med. Technol. Nat. Congr. (TIPTEKNO)*, Oct. 2017, pp. 1–4.

[42] D. V. Tuzoff *et al.*, "Tooth detection and numbering in panoramic radiographs using convolutional neural networks," *Dentomaxillofacial Radiol.*, vol. 48, no. 4, May 2019, Art. no. 20180051.

[43] Y.-F. Kuo *et al.*, "A convolutional neural network approach for dental panoramic radiographs classification," *J. Med. Imag. Health Inform.*, vol. 7, no. 8, pp. 1693–1704, Dec. 2017.

[44] X. Du, Y. Chen, J. Zhao, and Y. Xi, "A convolutional neural network based auto-positioning method for dental arch in rotational panoramic radiography," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 2615–2618.

[45] P. Chu *et al.*, "Using octuplet siamese network for osteoporosis analysis on dental panoramic radiographs," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 2579–2582.

[46] W. Poedjiastoeti and S. Suebnukarn, "Application of convolutional neural network in the diagnosis of jaw tumors," *Healthcare Inform. Res.*, vol. 24, no. 3, p. 236, 2018.

[47] W. M. Association, "World medical association declaration of Helsinki: Ethical principles for medical research involving human subjects," *JAMA*, vol. 310, no. 20, pp. 2191–2194, Nov. 2013, doi: 10.1001/jama.2013.281053.

[48] L. F. Tomás, L. S. M. Mónico, I. Tomás, P. Varela-Patiño, and Benjamín Martin-Biedma, "The accuracy of estimating chronological age from Demirjian and Nolla methods in a Portuguese and Spanish sample," *BMC Oral Health*, vol. 14, no. 1, p. 160, 2014.

[49] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[50] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[51] F. Chollet *et al.* (2015). *Keras*. [Online]. Available: https://keras.io

[52] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[53] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167

[54] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[55] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," 2012, *arXiv:1212.5701*. [Online]. Available: https://arxiv.org/abs/1212.5701

[56] W. Miao, Y. R. Gel, and J. L. Gastwirth, "A new test of symmetry about an unknown median," in *Random Walk, Sequential Analysis and Related Topics: A Festschrift in Honor of Yuan-Shih Chow*. Singapore: World Scientific, Dec. 2006, pp. 199–214.

[57] J. Ameijeiras-Alonso, R. M. Crujeiras, and A. Rodríguez-Casal, "Mode testing, critical bandwidth and excess mass," *TEST*, vol. 28, no. 3, pp. 900–919, Sep. 2019.

[58] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

[59] R. Cameriere, L. Ferrante, H. Liversidge, J. Prieto, and H. Brkic, "Accuracy of age estimation in children using radiograph of developing teeth," *Forensic Sci. Int.*, vol. 176, nos. 2–3, pp. 173–177, Apr. 2008.

[60] J. L. Prieto, E. Barbería, R. Ortega, and C. Magaña, "Evaluation of chronological age based on third molar development in the Spanish population," *Int. J. Legal Med.*, vol. 119, no. 6, pp. 349–354, Nov. 2005.

[61] A. C. Butti, A. Clivio, M. Ferraroni, E. Spada, A. Testa, and A. Salvato, "Häävikko's method to assess dental age in Italian children," *Eur. J. Orthodontics*, vol. 31, no. 2, pp. 150–155, 2008.

[62] A. Cruz-Landeira, J. Linares-Argote, M. Martínez-Rodríguez, M. S. Rodríguez-Calvo, X. L. Otero, and L. Concheiro, "Dental age estimation in Spanish and Venezuelan children. Comparison of Demirjian and Chaillet's scores," *Int. J. Legal Med.*, vol. 124, no. 2, pp. 105–112, 2010.

[63] A. B. Acharya and S. Mainali, "Sex discrimination potential of Buccolingual and mesiodistal tooth dimensions," *J. Forensic Sci.*, vol. 53, no. 4, pp. 790–792, Jul. 2008.

[64] D. Franklin, P. O'Higgins, C. E. Oxnard, and I. Dadour, "Discriminant function sexing of the mandible of Indigenous South Africans," *Forensic Sci. Int.*, vol. 179, no. 1, pp. 84.e1–84.e5, Jul. 2008.

[65] D. L. Anderson and G. W. Thompson, "Interrelationships and sex differences of dental and skeletal measurements," *J. Dental Res.*, vol. 52, no. 3, pp. 431–438, May 1973.

[66] T. R. Peckmann, S. Meek, N. Dilkie, and M. Mussett, "Sex estimation using diagonal diameter measurements of molar teeth in African American populations," *J. Forensic Legal Med.*, vol. 36, pp. 70–80, Nov. 2015.

[67] C. Capitaneanu, G. Willems, and P. Thevissen, "A systematic review of odontological sex estimation methods," *J. Forensic Odonto-Stomatol.*, vol. 35, no. 2, p. 1, 2017.