TESE DE DOUTORAMENTO

DEEP LEARNING FOR CHRONOLOGICAL AGE AND SEX PREDICTION FROM DENTAL PANORAMIC RADIOGRAPHS

Nicolás Vila Blanco

ESCOLA DE DOUTORAMENTO INTERNACIONAL DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOUTORAMENTO EN INVESTIGACIÓN EN TECNOLOXÍAS DA INFORMACIÓN

SANTIAGO DE COMPOSTELA
2022

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Declaración do autor da tese

Don Nicolás Vila Blanco

Título da tese: **Deep learning for chronological age and sex prediction from dental panoramic radiographs**

Presento a miña tese, seguindo o procedemento adecuado ao Regulamento, e declaro que:

1. *A tese abarca os resultados da elaboración do meu traballo.*
2. *De ser o caso, na tese faise referencia ás colaboracións que tivo este traballo.*
3. *Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.*
4. *A tese é a versión definitiva presentada para a súa defensa e coincide a versión impresa coa presentada en formato electrónico.*

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

En Santiago de Compostela, 10 de decembro de 2021

Asdo. Nicolás Vila Blanco

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Autorización do/a director/a/titor/a da tese

Deep learning for chronological age and sex prediction from dental panoramic radiographs

Dona María José Carreira Nouche, Profesora Titular da Área de Ciencia da Computación e Intelixencia Artificial da Universidade de Santiago de Compostela

Dona Inmaculada Tomás Carmona, Profesora Titular da Área de Estomatoloxía da Universidade de Santiago de Compostela

INFORMAN:

*Que a presente tese correspóndese co traballo realizado por **Don Nicolás Vila Blanco**, baixo a nosa dirección/titorización, e autorizamos a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como directoras/titoras desta non incorre nas causas de abstención establecidas na Lei 40/2015.*

*De acordo co indicado no Regulamento de Estudos de Doutoramento, declara tamén que a presente tese de doutoramento é idónea para ser defendida en base á modalidade de **Monográfica con reproducción de publicacións**, na que a participación do doutorando foi decisiva para a súa elaboración e as publicacións se axustan ao Plan de Investigación.*

En Santiago de Compostela, 10 de decembro de 2021

Asdo. María José Carreira Nouche
Directora tese

Asdo. Inmaculada Tomás Carmona
Directora tese

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

A loita nobre e leal das ideas é o que asegura o progreso

Castelao

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Agradecementos

En primeiro lugar, gustaríame amosar o meu sincero agradecemento ás miñas directoras de tese, as profesoras María José Carreira e Inmaculada Tomás, pola confianza que depositaron en min e os valiosos consellos que fixeron que esta investigación fose posible. Grazas tamén a Carlos polo seu interese e disposición para botar unha man desde o primeiro momento.

Tamén me gustaría agradecer ao Centro de Investigación en Tecnoloxías Intelixentes da USC, polos recursos postos a disposición dos investigadores e a formación ofertada, así como ás unidades de apoio do centro pola súa dispoñibilidade e disposición para resolver calquera dúbida. E en particular, grazas aos meus compañeiros de laboratorio, por ensinarme a desconectar do traballo en momentos de afogo e reconectar con el en momentos de distracción.

Grazas tamén á Facultade de Odontoloxía da USC, por proporcionar as radiografías necesarias para esta investigación, e a todas as compañeiras que dedicaron o seu tempo a colectar ditas imaxes e etiquetalas. Gustaríame facer unha mención especial a Paulina, unha gran amiga e de quen aprendín moito ao longo deste período.

O meu maior agradecemento é sen lugar a dúbidas para a miña familia. A Jenni, a miña compañeira de vida, por facerme tan feliz e afortunado e por estar ao meu carón tódolos días apoiándome nos éxitos e nos tropezos. Tamén aos meus pais e á miña irmá, a quen lles debo ser quen son e chegar a onde cheguei. Déronme forzas cando as precisei e un tirón de orellas cando o merecía, e nunca poderei agradecerllo o suficiente.

Por último, gustaríame agradecer ás entidades que fixeron posible esta tese co seu financiamento, en especial á Consellería de Educación, Universidades e Formación Profesional da Xunta de Galicia, a través do seu programa de axudas de apoio á etapa predoutoral e as axudas de consolidación e estruturación de unidades de investigación de referencia competitiva e de potencial de crecemento.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10 de decembro de 2021

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Resumo

A estimación da idade cronolóxica e o sexo constitúen tarefas fundamentais nunha ampla variedade de disciplinas. A primeira consiste en predicir o tempo transcorrido entre o nacemento dunha persoa ata o momento actual —no caso de persoas vivas— ou ata o momento da súa morte —en persoas falecidas—, mentres que a segunda trata de clasificar un individuo como home ou muller. Ámbalas tarefas son esenciais no eido da antropoloxía forense para identificar restos humanos ou recolectar información acerca de poboacións pasadas. Ademais, a estimación da idade cronolóxica resulta crucial para determinar se unha persoa acadou a maioría de idade, validar partidas de nacemento dubidosas ou controlar fluxos de migración, entre outros.

Os métodos empregados para estimar a idade cronolóxica varían de acordo a varios aspectos. Por unha banda, o procedemento é diferente dependendo de se a persoa está viva ou falecida. No primeiro caso, a totalidade do corpo pode ser analizada e é posible realizar avaliacións que requiran da interacción do suxeito, como as capacidades mentais, mentres que no segundo caso a información adoita ser máis limitada, usualmente só o esqueleto, pero pódense utilizar métodos destrutivos que achegan información adicional. O tramo de idade do suxeito tamén condiciona o proceso de estimación da idade, xa que os indicadores biolóxicos se correlacionan coa idade cronolóxica en maior grado nas persoas máis novas, e dita relación vaise difuminando a medida que a persoa medra. Este fenómeno fai que a estimación da idade sexa máis complicada en persoas adultas. Ademais, a raza, a nutrición, a xenética ou o nivel socioeconómico, entre outros, tamén condicionan a composición do corpo e o seu desenvolvemento, o que provoca que os métodos de estimación da idade, desenvolvidos orixinalmente para unha poboación específica, teñan que ser adaptados antes de ser aplicados a outras poboacións.

USC UNIVERSIDADE

De entre todos os indicadores biolóxicos dispoñibles para a estimación da idade cronolóxica, o estado da dentición é o máis amplamente utilizado por diversas características. A composición dos dentes fai que estes sexan moi resistentes ante desastres masivos, de modo que poden ser preservados non só despois da morte, senón despois de miles de anos. Ademais, os factores xenéticos e ambientais afectan á dentición en moito menor medida que ao resto de partes do esqueleto, polo que o seu desenvolvemento é menos variable. A estimación da idade cronolóxica a partir da dentición constitúe ademais unha ferramenta valiosa que permite aos ortodoncistas e odontopediatras avaliar as fases de desenvolvemento, principalmente en nenos/as e adolescentes, desde o punto de vista do diagnóstico e o tratamento. Desta maneira, pódese comprobar se o desenvolvemento está avanzado ou atrasado con respecto ao esperado a unha idade determinada, o cal ten enorme relevancia no caso dalgunhas enfermidades sistémicas como a parálise cerebral ou a síndrome de Down, que adoitan asociarse a un atraso na maduración psicolóxica e esquelética.

Ademais dos métodos baseados na avaliación da dentición, outras metodoloxías reportaron a idoneidade de estruturas óseas adxacentes para a estimación da idade cronolóxica, sendo a mandíbula unha das máis relevantes. Os cambios deste óso ao longo da vida dun individuo, sobre todo na etapa da nenez e a adolescencia, demostraron levar a unha predición da idade cronolóxica moi precisa en persoas non adultas mediante o estudo de medidas lineais e angulares entre puntos anatómicos característicos. Nos individuos adultos, porén, os cambios mandibulares están altamente influenciados pola perda de dentes e a estimación da idade pode ser errónea.

Por outra banda, a estimación do sexo xoga un rol fundamental na análise forense, especialmente na identificación de individuos cando o corpo está danado ou incluso nun proceso de descomposición. É importante remarcar que a determinación do sexo é a etapa máis discriminativa de cantas se realizan durante o proceso de identificación, xa que descarta a un gran número de individuos. Ademais, é unha tarefa case obrigada antes de realizar a estimación da idade cronolóxica, xa que esta está baseada en estándares específicos dependendo do sexo do individuo. De entre todas as metodoloxías existentes para determinar o sexo, os métodos bioquímicos baseados na análise de ADN son os máis precisos. Porén, na práctica son custosos, consomen moito tempo e a súa dispoñibilidade é limitada, e é por isto que a determinación do sexo adoita ser levada a cabo mediante avaliación morfolóxica e morfométrica do esqueleto.

A precisión obtida cos métodos de estimación do sexo ten un comportamento totalmente oposto ao dos métodos de estimación da idade cronolóxica, é dicir, os mellores resultados

USC UNIVERSIDADE

obtéñense en individuos adultos. Isto é debido principalmente a que o dimorfismo sexual nas estruturas do corpo humano empeza a desenvolverse durante a puberdade, xunto con outras características sexuais secundarias como o crecemento de pelo corporal e unha xeración muscular máis acusada nos homes, ou o crecemento dos peitos e o ciclo menstrual nas mulleres. Moitos autores reportaron que a determinación do sexo a través do esqueleto en nenos e nenas é tamén complicado pola falta de mostras osteolóxicas de calidade, o cal leva irremediablemente a modelos de estimación máis pobres.

De entre todas as metodoloxías propostas para a determinación do sexo en adultos, os expertos coinciden en que a pelve é o mellor indicador dado o seu desenvolvemento nas mulleres para poder dar a luz, o cal leva a modelos con unha precisión próxima ao 100%. Así e todo, a descomposición do corpo humano despois da súa morte pode dificultar a recuperación dalgunhas partes do esqueleto, e por iso é preciso establecer metodoloxías que fagan uso das estruturas máis resistentes. Este é o caso da dentición, cuxo dimorfismo sexual foi estudado de cara ao desenvolvemento de métodos de estimación do sexo. Tamén foi estudada neste sentido a mandíbula, que demostrou ser aínda mellor indicador debido ás acusadas diferenzas na súa forma entre homes e mulleres.

Como no caso da estimación da idade cronolóxica, os cambios relacionados co sexo varían dunha poboación a outra debido ás compoñentes xenéticas e factores como a nutrición ou o nivel socioeconómico. Isto provoca que os modelos de determinación do sexo precisen ser modificados mediante a inclusión de estándares poboacionais específicos para obter unha maior precisión.

Aínda que existen métodos para avaliar o estado e desenvolvemento dos dentes e outros ósos circundantes baseados en análises histolóxicas ou cambios químicos, as radiografías dentais supoñen unha alternativa rápida, pouco custosa e non destrutiva. Existen distintos tipos dependendo do método de adquisición, sendo as radiografías panorámicas ou ortopantomografías (OPG) as máis utilizadas xa que cobren a totalidade da dentición e outras moitas estruturas óseas. Deste modo, unha OPG pode axudar no diagnóstico de enfermidade periodontal, quistes mandibulares ou cancro oral, entre outros.

Nas últimas décadas propuxéronse diversas metodoloxías que utilizan radiografías para a estimación da idade. Algunhas delas utilizan o criterio da erupción clínica ou alveolar, como o método de Haavikko, aínda que a erupción dental ocorre nun período tan curto que a aplicabilidade é moi limitada. Os atlas constitúen outra ferramenta para a estimación da idade, xa que exemplifican as fases de desenvolvemento dentario mediante táboas e gráficas.

USC UNIVERSIDADE

Outra gran familia de métodos son aqueles baseados nos sistemas de puntuación, nos que unha serie de dentes son avaliados individualmente en base a unhas fases de desenvolvemento predefinidas. Cada dente obtén unha puntuación concreta, que logo se emprega para calcular a idade. Os exemplos de sistemas de puntuación máis coñecidos son os métodos de Nolla e Demirjian. En 2006, Cameriere propuxo utilizar unha aproximación novidosa a este campo, xa que atopou unha correlación inversa entre o grao de apertura dos ápices e a idade. A través dunha fórmula sinxela conseguiu acadar resultados moi precisos.

En canto á estimación radiolóxica do sexo, a vía máis explorada é a que relaciona a idade coa evolución de diversas medidas lineais e angulares da mandíbula, como a altura e a flexión da rama mandibular ou a altura dos cóndilos. Tamén se ten observado un dimorfismo sexual acusado en outras estruturas da cavidade oral observables radioloxicamente, como os senos maxilares ou o foramen mental. A determinación do sexo mediante a análise da dentición tamén foi analizada, aínda que a maioría de estudos traballaron con restos do esqueleto ou con moldes dentais e os resultados non son tan prometedores. Os traballos puramente radiolóxicos son escasos, sendo Capitaneanu o autor do único traballo realizado sobre radiografías dentais panorámicas.

Os últimos avances no campo do procesamento de imaxe tiveron unha gran repercusión no modo no que as imaxes médicas foron utilizadas, sendo especialmente relevante o caso das radiografías. O punto de inflexión veu marcado pola combinación de modelos de redes neuronais moi precisos e o desenvolvemento de tecnoloxías computacionais extremadamente potentes, o que levou ao auxe de unha nova familia de técnicas coñecida como aprendizaxe profunda. Especificamente, as redes neuronais convolucionais —*convolutional neural networks* ou CNNs— destacan pola súa capacidade para o procesamento de datos espaciais —como as imaxes— dunha maneira precisa e escalable.

A radioloxía dental beneficiouse destes desenvolvementos, dende o proceso de adquisición das imaxes ata os métodos de diagnóstico máis complexos. Algunhas metodoloxías centráronse en mellorar as imaxes para paliar problemas na toma das radiografías, como artefactos ou fallos de calibración nos dispositivos de adquisición. Outros traballos propuxeron métodos automatizados para detectar estruturas específicas nas radiografías, como os dentes, a mandíbula ou o foramen mental, entre outros. Isto levou aos/ás expertos/as a ir un paso máis aló, e desenvolver estratexias automáticas para un diagnóstico asistido, como por exemplo a detección de fracturas, enfermidade periodontal, osteoporosis, quistes e tumores, caries den-

tais ou lesións apicais, entre outras moitas patoloxías. A precisión destes métodos chega a ser comparable á dun/ha experto/a, reforzando a súa validez como sistemas de apoio.

Estes avances tamén chegaron aos métodos de estimación da idade cronolóxica e o sexo. Un dos primeiros exemplos é o traballo de Čular, que combinou un método de detección do terceiro molar en radiografías panorámicas capaz de parametrizar a súa forma e textura, con unha rede neuronal que converte dita información nunha estimación numérica da idade. Ambas compoñentes traballan dun xeito totalmente automático, e polo tanto non é necesaria a intervención humana.

Como en outros moitos campos, as CNNs axudaron non só a automatizar certas tarefas, senón tamén a mellorar a súa precisión. En canto á estimación da idade cronolóxica, De Tobel desenvolveu un sistema de detección de fases de desenvolvemento dentario —en concreto as propostas por Demirjian— a partir de recortes de imaxes de terceiros molares. Merdietio mellorou dita proposta engadindo unha CNN capaz de obter automaticamente ditos recortes en radiografías panorámicas, facendo que o sistema funcionase dun xeito totalmente automático. Outros autores, como Banar e Guo, propuxeron CNNs centradas en producir estimacións da idade cronolóxica nun único paso (sistemas *one-stage*), sen detectar dentes concretos ou outras estruturas anatómicas. Esta aproximación foi seguida por outros/as investigadores/as para a determinación do sexo, como Ilić, Milošević ou Ke, con resultados moi prometedores.

En xeral, as metodoloxías mencionadas anteriormente demostraron un alto potencial —especialmente as baseadas en aprendizaxe profunda— non só para detectar e medir estruturas observables nas radiografías, senón tamén para avaliar características de alto nivel, como a súa forma e aparencia, e inferir un diagnostico clínico ou un perfil biolóxico do suxeito ao que pertence a imaxe. Aínda que a precisión obtida é salientable, a complexidade destes métodos está asociada a un decrecemento na súa interpretabilidade, o cal é crucial para a súa aplicación nun entorno real. Neste sentido, existen dúas posibles vías: enfocarse en metodoloxías simples e interpretables asumindo unha perda na precisión, ou desenvolver modelos altamente precisos e provelos de mecanismos que axuden a interpretar os resultados obtidos.

O propósito xeral desta tese de doutoramento é o desenvolvemento dunha metodoloxía totalmente automatizada baseada en técnicas de aprendizaxe profunda, con vistas a obter resultados precisos pero interpretables. Ademais, a experimentación levada a cabo pretende demostrar non só a idoneidade da radiografía dental panorámica para a estimación da idade e do sexo, senón tamén a adecuación de estruturas esqueléticas individuais como a mandíbula e

USC UNIVERSIDADE

a dentición para o mesmo fin. Especificamente, propuxéronse tres obxectivos diferentes neste sentido.

1. **Estimación da idade cronolóxica e o sexo en radiografías dentais panorámicas completas.** Deseño dunha arquitectura de CNN *one-stage* que produza as predicións directamente dende a imaxe de entrada, sen resultados intermedios. As hipóteses principais, probadas xa no ámbito clínico, son que o desenvolvemento das estruturas esqueléticas presentes na cavidade oral segue distintas temporizacións en homes e mulleres, e que a aparición de características sexuais secundarias non acontece até a puberdade. Isto significa que idade e sexo dependen a unha da outra, e polo tanto a arquitectura desenvolvida ten que detectar características da imaxe tanto sexuais como relativas ao desenvolvemento para mellorar as predicións de idade e sexo.

2. **Estimación da idade cronolóxica e o sexo a partir da forma mandibular en radiografías dentais panorámicas.** Baseándose na investigación forense que ten probado a idoneidade da forma da mandíbula como bo indicador para a estimación da idade e o sexo, é preciso deseñar unha metodoloxía que automatice a extracción do contorno mandibular nas radiografías, caracterice cuantitativamente dita forma e obteña precisións de idade e sexo fiables utilizando esta información. Ademais, débense proporcionar explicacións visuais de como a forma da mandíbula afecta a ditas predicións.

3. **Estimación da idade cronolóxica e o sexo a partir dos dentes presentes nunha radiografía dental panorámica.** Aínda que o primeiro obxectivo xa contempla o uso da radiografía dental completa, é preciso cuantificar a contribución de cada dente en ditas estimacións dado que son as partes do esqueleto máis empregadas para a estimación da idade cronolóxica. Neste obxectivo preténdese desenvolver unha arquitectura baseada en aprendizaxe profunda que detecte os dentes presentes na imaxe automaticamente e utilice as características visuais dos mesmos para predicir a idade e o sexo. De novo, a saída do sistema deberá comprender non só estimacións cuantitativas, senón tamén métodos para explicar como se obteñen ditas estimacións, o cal é crucial para os/as expertos/as clínicos/as.

Na memoria desta tese de doutoramento detállase o traballo realizado para cumprir cos obxectivos propostos, especificando a metodoloxía seguida, os resultados obtidos e unha argumentación da relevancia que ten cada contribución.

USC UNIVERSIDADE

Dado que actualmente non existe un conxunto de radiografías dentais panorámicas público anotado coa idade e o sexo do/a paciente, foi preciso construír un banco de imaxes propio co que desenvolver e validar as metodoloxías propostas. No capítulo 3 descríbese dito conxunto de datos, denominado OPG-AS (*OPG database for Age and Sex determination*), incluíndo o proceso mediante o cal se recopilaron as imaxes, as características das mesmas e a distribución demográfica dos individuos. Ademais, amósanse exemplos de imaxes incluídas en dita base de datos.

O capítulo 4 describe dúas arquitecturas de CNN *one-stage* para a estimación da idade cronolóxica e unha para a estimación do sexo. A arquitectura máis simple, DANet, consiste nunha única vía onde capas convolucionais e totalmente conectadas son aplicadas secuencialmente para obter a predición da idade. Por outra banda, DASNet engade unha vía secundaria para extraer en paralelo as características que presentan maior dimorfismo sexual e propágaas á vía principal en puntos intermedios, co obxectivo de reducir o erro cometido na estimación da idade. Ámbalas arquitecturas mostraron resultados ao nivel do estado da arte, especialmente nos rangos de idades máis novos, con erros absolutos medios de aproximadamente oito meses nos individuos menores de 15 anos. Porén, DASNet obtivo estimacións máis fiables, o cal permitiu confirmar a hipótese clínica de que o desenvolvemento das estruturas esqueléticas segue distintas temporizacións en homes e mulleres, e que esa información debe ser tida en conta para desenvolver métodos de estimación máis precisos. Ademais, os roles das dúas vías de DASNet foron intercambiados para validar esta aproximación no problema da estimación do sexo, sendo a vía principal a que estima o sexo e a secundaria a que estima a idade. Dito enfoque acadou resultados á altura do estado da arte, sendo especialmente salientable a precisión obtida nos menores de idade.

No capítulo 5 detállase un método para caracterizar a forma da mandíbula en imaxes panorámicas dun xeito totalmente automático e repetible. Para isto proponse o uso dunha CNN para localizar puntos característicos no contorno mandibular mediante a regresión dos seus mapas de calor. Ademais, propóñense catro descritores de forma para caracterizar cuantitativamente a mandíbula, incluíndo un conxunto de medidas lineais e angulares calculadas a partir de oito puntos característicos, o tamaño do centroide, as variacións de forma con respecto á forma media e un grupo de parámetros de forma obtidos a través dun modelo de distribución de puntos. O rendemento da CNN para a detección do contorno mandibular foi comparado coa precisión obtida por unha persoa experta adestrada, sendo máis robusta e fiable a CNN. O uso dos descritores mandibulares foi validado no problema de estimación

USC UNIVERSIDADE

da idade cronolóxica e o sexo mediante métodos simples de regresión e clasificación. O rendemento acadado demostrou que os descritores propostos son útiles para este obxectivo. A combinación dos parámetros de forma e o tamaño do centroide demostrou ser a máis efectiva, acadando un erro absoluto medio de pouco máis de ano e medio na estimación de idade e unha precisión de case 88% na determinación do sexo. Ademais, o modelo de distribución de puntos resulta de grande utilidade para visualizar os cambios na forma da mandíbula e dotar de máis explicabilidade ao método proposto.

O capítulo 6 describe un método totalmente automático para predicir a idade e o sexo a partir da aparencia dos dentes en imaxes panorámicas. Dito método consta de dous compoñentes principais. Primeiro aplícase un detector de obxectos para extraer as caixas delimitadoras (*bounding boxes*) dos dentes, tendo en conta a súa orientación. Nun segundo paso, as características extraídas polo detector en cada dente son propagadas a unha CNN para a estimación da idade e o sexo que, para cada dente, obtén unha estimación da distribución de probabilidade tanto da idade —mediante unha distribución normal— como do sexo —mediante unha distribución de Bernoulli—. Finalmente, o método proposto agrega as distribucións estimadas tendo en conta a súa incerteza. Os resultados amosaron que esta metodoloxía acadou un erro menor que calquera outro método previo de estimación da idade en radiografías panorámicas, sendo de menos de un ano en menores de 25 anos, e unha precisión de estimación do sexo moi por riba de calquera método que empregase a dentición, acadando case un 92% de precisión. Para dotar ao método dunha maior interpretabilidade, tamén se propón un método baseado no método Grad-CAM con modificacións para ter en conta a incerteza das estimacións.

Finalmente, no capítulo 7 resúmense as contribucións desta tese de doutoramento e os posibles campos de aplicación, mentres que no capítulo 8 propóñense distintas vías para continuar o traballo, co obxectivo tanto de mellorar os resultados como de potenciar unha futura aplicabilidade dos mesmos en escenarios reais.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Contents

Glossary of clinical terms **1**

List of acronyms **3**

1 Introduction **5**
1.1 Chronological age estimation 6
1.2 Sex estimation 9
1.3 Dental radiology for chronological age and sex estimation 11

2 Justification and objectives **23**

3 OPG-AS: a database of dental panoramic radiographs **27**

4 One-stage CNN for age and sex prediction from oral cavity structures **33**
4.1 Introduction 35
4.2 Materials and methods 37
4.3 Experiments and performance evaluation 40
4.4 Results 45
4.5 Discussion and conclusions 51
4.6 Appendix A 53
4.7 Appendix B 60

5 CNN and shape modelling for age and sex prediction from the mandible **67**
5.1 Introduction 68
5.2 Materials and methods 69
5.3 Experiments and performance evaluation 76

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.4 Results 77
5.5 Discussion and conclusions 86

6 Two-stage CNN for age and sex prediction from tooth appearance 89
6.1 Introduction 90
6.2 Related work 91
6.3 Materials and methods 95
6.4 Experiments and performance evaluation 101
6.5 Results 103
6.6 Discussion and conclusions 111

7 Conclusions 117

8 Future work 121

Bibliography 125

List of Figures 149

List of Tables 153

Publications 157

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Glossary of clinical terms

alveolar eruption Emergence of the teeth through the alveolar bone.

aminoacid racemisation Conversion of an L-aminoacid compound to a D-amino acid or vice versa, which allows dating of organic samples.

attrition Tooth substance loss due to the contact between adjacent teeth.

calcification Accumulation of calcium salts in body tissue, facilitated by parathyroid hormone and vitamin D.

cementum Calcified substance that covers the root of a tooth. It is part of the periodontium, and it is composed of 45-50% of minerals and 50-55% of both organic and water.

clinical eruption Emergence of the teeth through the gingiva.

dentine Calcified tissue of the tooth, covered with enamel in the tooth crown and with cementum in the root. Its composition is 45% mineral, 33% organic, and 22% water.

enamel One of the four major tissues that make up the normally visible part of the tooth, covering the crown, which is known by its extremely hardness due to the high mineral composition (90%).

epiphysis Rounded end of a long bone, at its joint with the adjacent bone.

odontogenesis Process through which teeth are developed.

periodontal Related to the periodontium, a tissue that provides the teeth with the support they need to carry out several functions.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

proximal surface The surface of a tooth contacting with its neighbouring teeth. It can be mesial (toward the midline) or distal (toward the posterior area).

pulp The most internal layer of the tooth. It is composed of connective tissue and special cells called odontoblasts.

root resorption Progressive loss of dentine and cementum as a consequence of the action of the osteoclasts.

taphonomic process Stage through which organisms decay and become fossilised, representing the transition from the biosphere to the lithosphere.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

List of acronyms

BCE binary cross entropy. 42, 73, 102

CBCT cone beam computed tomography. 15

CCD charge-coupled device. 11

CNN convolutional neural network. xx, xxi, 20, 21, 25, 33, 35, 37–39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59–61, 63, 65, 67, 69–71, 73, 75–79, 81, 83, 85–89, 91, 93–97, 99, 101, 103, 105, 107, 109, 111–113, 115, 117, 118, 151, 155

CPM convolutional pose machine. 72, 76, 78, 79

CPN cascade pyramid network. 72, 76, 78, 79

CT computed tomography. 15, 19, 85, 87, 155

CV cross validation. 40, 41, 60, 85, 150, 155

DANet Dental Age Network. 33, 38, 39, 42–46, 51–54, 56–58, 117, 150, 153, 154

DASNet Dental Age and Sex Network. 33, 34, 38, 39, 42–54, 56–58, 60, 62, 63, 65, 117, 118, 150, 153, 154

DSANet Dental Sex and Age Network. 34, 60, 62–65

ELU exponential linear unit. 98

ERFNet efficient residual factorized convnet. 92

USC

FCN fully convolutional network. 71, 72, 92

FDI *Fédération Dentaire Internationale*. 94, 96

FPN feature pyramid network. 72, 96, 97, 101, 152

GAP global average pooling. 39

Grad-CAM gradient-weighted class activation mapping. 44, 49, 51, 99, 150

HRNet high-resolution network. 72, 76, 78, 79

IoU intersection over union. 101

MCI mandibular canine index. 10, 61

OPG orthopantomogram. 12, 25, 27, 31–33, 35–37, 49, 51, 52, 60, 61, 63, 67, 69, 70, 85, 86, 88, 90–95, 97, 99, 103, 112–115, 117, 118, 123, 149, 152–156

OPG-AS OPG database for Age and Sex determination. xx, 27–31, 37, 70, 95, 117, 149, 153

PANet path aggregation network. 92

PDM point distribution model. 67, 69, 70, 74, 75, 80, 84–87, 151, 155

R-CNN region-based convolutional neural network. 91, 92, 96, 97, 152

ReLU rectified linear unit. 38

ResNet residual network. 76, 78, 92, 94

RMSE root mean squared error. 42, 73

ROI region of interest. 92, 96, 97, 101, 152

RPN region proposal network. 96, 97, 101, 152

SHN stacked hourglass network. 72, 76, 78, 79, 83, 86, 155

SiZer significant zero-crossing of the derivative. 53, 56, 59, 151

YOLO you only look once. 94

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 1

Introduction

Chronological age estimation and biological sex classification are two relevant tasks in a variety of clinical procedures. They consist, respectively, in determining the time elapsed from a subject's birth and the chance to be a male or a female. Both age and sex estimation are useful tools for forensic anthropologists to perform the identification of human remains, or to collect information about the population demographics of a past era. Age estimation has also proven to be crucial for legal age determination, validation of birth certificates, or control of migration flows, among others.

Many body indicators have been used for both age and sex estimation. However, experts agree that the oral cavity hosts anatomical structures whose development correlates with chronological age to a large degree, and these structures are different enough across males and females to enable accurate sex determination. Specifically, the teeth and mandible are considered good candidates due to their resistance and the ease of being radiologically observed.

In the last decades, researchers and clinical experts developed numerous numeric models to convert dry-bone or radiologic measurements into an estimation of chronological age or sex. However, the application of these models has a series of disadvantages, such as the inherent subjectivity caused by the expert's measurements or the time required to conduct the process.

To alleviate these problems, this PhD Thesis introduces three different models for automatic age and sex estimation on dental panoramic radiographs to assess the suitability not only of the whole X-ray image but also of specific skeletal structures present in the image, such as the mandible and the teeth.

USC UNIVERSITAT DE VALÈNCIA

1.1 Chronological age estimation

It is widely agreed that different ages can be measured in an individual [75]. On the one hand, the chronological age represents the lifetime of the individual, given by the time that has elapsed since the birthdate. On the other hand, biologic age —also called physiologic and developmental age— is a longevity measurement in terms of body development, and it is assessed differently for each body part or system [60, 152]. Skeletal age, for example, represents the degree of bone maturity in relation to a pattern obtained from a normal population of the same sex. Morphological age measures the growth and development of several body segments and the weight of the individual. Sexual age assesses all the puberty features that allow distinguishing between males and females. Dental age is determined from both deciduous and permanent tooth development. Other ages, such as mental age, have also been used as a development mark.

The estimation of the chronological age by assessing one or several biologic ages has been used for a variety of purposes. One example is forensic medicine, specifically human identification. This process does not only have administrative and ethical implications but also serious legal consequences, due to the changing of civil status of the deceased human and their relatives. The number of missing and unidentified people has been growing worldwide every year [188] for numerous reasons: mass disasters, murder cases, missing indigenous peoples, or migration catastrophes, among others [69]. Thus, the use of new techniques to accelerate and improve the results is of crucial importance.

Legal medicine also needs processes to estimate the age of an individual. In the last years, growing migration to or between European countries increased the number of refugees and asylum seekers, who do not have documented evidence of their birthdate or are suspected of providing false documentation [194]. This is especially noticeable when those people come from undeveloped regions, where almost 30% of births are unregistered [67], or from countries at war, where there are discrepancies in the recognition of birth certificates between opponents. In those cases, the main challenge is to determine the legal age attainment, as the level of protection is generally higher for minors. Moreover, the implication in legal processes is different for people under and over the legal age. Depending on the country, this threshold is set between 14 and 21, and the legal limit of age to get into a criminal process is set between 0 —some countries set this value by default— and 16 [47].

Adoption processes also require age estimation methods, as the law in force in some countries establishes strict criteria in this regard —Italian law requires a minimum of ten years

USC UNIVERSIDADE

and a maximum of 45 years of difference between the child and the new parents, while Indian law makes it mandatory for children to be aged between one and 12 years to be adopted [183]. These criteria are also applied for marriage processes, where some countries require that the contracting parties be over 18 years of age.

The methods used for chronological age estimation vary according to some aspects. First, the procedure is quite different depending on whether the person is alive or dead. In the former, every body part can be assessed, some anthropometric measurements such as the patient height are easily obtained and evaluations requiring the interaction of the subject —like the previously mentioned mental age— are an option. However, every analysis has to be carried out in a non-destructive way, which limits the completeness of the studies. As a counterpart, the information obtained from a dead person is scarcer and usually restricted to a skeleton, which may be incomplete in some cases, but its assessment can be performed more thoroughly because destructive methods such are histological and biochemical analyses are possible [67].

The age cohort to which the subject belongs also conditions the age estimation pipeline. It is widely agreed that biological and chronological ages are highly correlated in younger ages, and become more distant as the subject grows. This is known as the *trajectory effect* and makes age estimation more difficult the older the subject is [161]. Moreover, some body indicators are no longer useful when the patient achieves a certain development milestone. In the case of skeletal age, most authors consider that the development ends when all epiphyses have fused [44]. Thereafter, morphological and metric studies of bones —size, shape, and degree of ossification— have to be substituted by the assessment of bone degradation.

Although the age estimation methods are very heterogeneous, everyone relies on the same base approach. Biologic indicators are evaluated either quantitatively and/or qualitatively, and translated into a chronological age value or age cohort by using prior information collected in a set of samples. The main issue comes from the fact that many factors influence the link between biological and chronological ages, and so the development of these population standards may be biased. Although male and female bodies are completely different in a variety of aspects and their development pattern follows different timings, ethnics, nutrition, genetics, or socioeconomic level, among others, also condition the body composition and development. This causes that the methods for chronological age estimation, which are originally developed for a specific population, need to be modified before being applied to another population. In the long term, this entails a lot of work in the research community.

USC UNIVERSIDADE

Among the available biologic indicators, dental age is the most commonly used for several reasons. The composition of the teeth provides them with an extreme resistance against mass disasters, so the dentition can be preserved not only after death but also after thousands of years. Moreover, dental mineralisation is less affected by genetic and environmental factors than any other skeletal part, so it is considered a better chronological age indicator [229]. The specific characteristics of the teeth and adjacent bones were used since Roman times to identify dead people [210]. Although the identification can be straightforward if antemortem dental records are available, the estimation of some features such as age, sex, or race leads to more accurate results, and let experts carry out this task even when there are no previous records. In legal medicine, there is an agreement to use dental radiography combined with hand-wrist radiography and a set of anthropometric measurements for the best accuracy [120].

Chronological age estimation from dental age and surrounding bones is also a valuable tool that lets orthodontists and paediatric dentists assess development stages, mainly in children and teenagers, from the point of view of diagnosis and treatment. Thus, experts can check if the development is advanced or delayed concerning the expected development at a certain age. This has special relevance in the case of some systemic conditions, such as cerebral palsy or Down syndrome, that are associated with a delay in physiological, skeletal, and dental maturation [63].

Various approaches have been used to perform dental age estimation, namely morphological, biochemical, and radiological [181]. Morphological methods analyse microscopic structure changes during dental development. For example, Gustafson [91] noted six dental tissue changes related to age, namely the attrition of the incisal or occlusal surfaces due to mastication, periodontitis, secondary dentine, the cementum apposition, root resorption, and transparency of the root. Biochemical methods are based on aminoacid racemisation, a reaction that occurs in some aminoacids that have two isomers. Although it is a very long process that is not yet usually noticed at the death of a living being, there are some aminoacids with a very high racemisation rate such as the aspartic acid, whose L/D ratio can reach 8% at the age of 60 in a human being [96]. If the racemisation rate is known, experts can estimate the age of living tissue and therefore the age of a living human with reasonable accuracy [193]. Finally, radiologic methods have been probably the most used approach for dental age estimation, as they allow a wider range of applications —both living and dead subjects—, a faster evaluation, and a cheaper cost, which is crucial in many scenarios. They will be analysed more in detail in Section 1.3.1.

In addition to methods based on teeth assessment, other approaches reported the suitability of adjacent bones to carry out chronological age estimation, being the mandible one of the most relevant examples. The age-related changes of this bone have been studied for decades, either through the whole bone [11] or specific parts [112, 189]. According to Franklin *et al.* [73, 74], mandibular morphology can lead to highly accurate age predictions in subadults both by analysing its shape and linear distances and angles between anatomical landmarks. In adult subjects, however, the mandible changes are highly influenced by tooth loss and thus the age estimation can be misleading [172].

1.2 Sex estimation

The estimation of sex plays a crucial role in forensic analysis, especially in human identification procedures where the body is seriously damaged or decomposed [56]. Although biochemical sexing methods based on DNA analysis are very accurate, in practice they are expensive, time-consuming, and their availability is limited, and so sex determination is usually carried out through a morphological and morphometric skeletal analysis [196]. It is worth noting that sex estimation is the most discriminative step in the human identification pipeline, and so it greatly simplifies the search for antemortem records. Moreover, it is a mandatory task before chronological age estimation, as these methods are almost always based on sex-specific standards [31, 105]. Sex estimation is also of great importance in forensic archaeology, as it provides valuable information on the demographic patterns of past populations and their relation to pathological conditions, growth and development, and diseases, among others [220].

The accuracy of sex estimation methods has the opposite behaviour than chronological age estimation in terms of age cohorts, that is, the better results are obtained in adult people. This is mainly because the sexual dimorphism in body structures starts to develop during puberty, together with other secondary sexual characteristics such as the increased body hair or the higher muscle generation rate in males, and the breasts growth and menstrual cycle in females [187]. Some authors reported that skeletal sex determination in children is also difficult due to the scarcity of high-quality osteological samples, which leads irretrievably to poor estimation models [105].

Among the proposed sex estimation approaches in adults, all experts agree that the pelvis is the best indicator as a result of the development related to childbirth function in females [217],

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

yielding sex estimation accuracies of 100% or close to it [89, 219]. Other bones characterised by high dimorphism are the skull, some bones in hands and feet, the patella, or the sternum [119]. However, the taphonomic processes happening in the body when it is exposed outdoors may hinder the recovery of some skeletal parts [207], and thus sex estimation methods using other bones have been developed for a wider range of options in the most difficult cases.

Because of their extreme resistance, dentition has been studied as an indicator of sexual dimorphism [221]. Given that the applicability of biochemical methods is limited due to cost and availability issues, and the morphological traits of teeth are reportedly useless for performing sex estimation [177], much of the research effort has been focused on odontometric methods. These methods are based on linear or volumetric measurements both on skeletal remains or dental casts, or through radiographic imaging [31]. Lots of different dental indicators have been used for estimating sex, namely root lengths [238], crown measurements [173], or intercusp distances [94]. Related to the latter, the mandibular canine index (MCI) proposed by Rao *et al.* [184] was applied to a variety of populations with successful results [104, 185], although some studies pointed to several limitations [3].

Besides the utility of teeth as a reliable indicator for sex estimation, other adjacent bone structures have also proven to be highly dimorphic and so suitable for determining sex. For example, the palate in males is deeper and more elongated [43], while the maxillary sinuses are significantly narrower in females [170]. The analysis of mandible's dimorphism has also led to numerous studies. Besides the methods based on the morphology of the mandible [79, 99], metric approaches have relied on linear measurements such as the bigonial width, bicondylar breadth, ramus height and breadth, and gonial angle, among others [169, 207, 211].

As in the case of age estimation, the sex-related changes of the human body vary across different regions due to genetic and environmental factors such as nutrition or socioeconomics [218], and so the methods developed for specific population standards need to be adapted in accordance before applied to other population samples. It is also remarkable that some traits that are sexually dimorphic in one population may be less in another [225]. In this regard, an effort is being made to develop methods associated with worldwide databases including multiple populations, to simplify the sex determination step for any given unknown sample [155].

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1.3 Dental radiology for chronological age and sex estimation

Dental radiographs —especially dental panoramic radiographs— have been used since the discovery of the X-rays, with many application fields: detection of dental abnormalities and treatment planning, prostheses designing and building, assessment of child tooth development, human identification by matching post-mortem and ante-mortem images, and many more. They provide much more information than a simple clinical observation due to the revealing of hidden tooth parts and other supporting structures. The acquisition setup consists of two main parts: an X-ray generator, which converts electrical input power into X-rays, and an X-ray detector —traditionally a photographic film— which records the amount of radiation passing through the object or body to capture [16]. Depending on the amount of radiation absorbed by the body structures, they will be reflected as brighter or darker areas in the film. In the last decades, electronic innovations allowed acquiring radiologic images without films. The most widely used technology in this regard is the charge-coupled device (CCD), which substitutes the film as the sensing element, as it can acquire light information and convert it into thousands of light intensity values that finally compose the digital image.

Although the only possible modifications in film image acquisition are the ones that control the density of the X-rays (kilovoltage, exposure time, miliamperage, and target-film distance), the images could be processed in a further step by scanning and converting them to a digital format. On the other hand, the digital acquisition is carried out in one step, and it is less prone to errors or image artefacts. Some digital X-ray machine manufacturers even started to provide acquisition devices with integrated processing algorithms to improve and slightly modify the images [97, 131]. Nevertheless, a bad calibration can lead to unrecoverable image artefacts, noise, and under- or overexposed areas, either through analogue or digital acquisition devices.

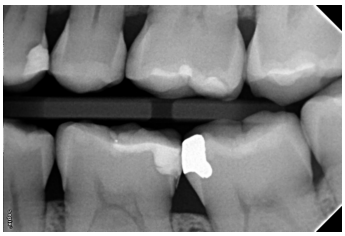
Depending on the desired dental capturing area, various types of dental images can be recorded. In intraoral image acquisition, the film or sensor is placed inside the mouth, while in extraoral images it is placed outside the mouth, on the rear side of the head. There are three types of intraoral images, which differ in the projection. The periapical view reveals both anterior and posterior teeth (Figure 1.1a), and it is useful to assess apical infection or inflammation, periodontal status, or the presence and position of unerupted teeth, among others. As shown in Figure 1.1b, the bite-wing view covers the crown of posterior teeth, so monitoring of caries, existing restorations, and periodontal status can be performed. Finally, the occlusal view shows either the floor of the mouth or the palate —which is the case of Figure 1.1c—, depending on the arch where the film is placed. These views aid in the assessment of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

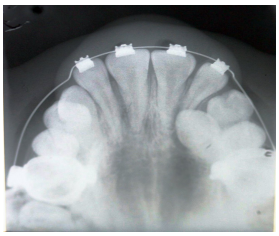
the periapical status of upper anterior teeth, detection of unerupted canines and odontomes, evaluation of cysts or tumours in the anterior maxilla, and assessment of fractures of the anterior teeth and alveolar bone.

The best-known extraoral view is the dental panoramic radiograph (Figure 1.1d), also called orthopantomogram (OPG) or simply panoramic image. As shown in Figure 1.2, it covers the whole dentition, as well as many other structures, so it can reveal periodontal disease, cysts in the jaw bones, jaw tumours and oral cancer, impacted teeth, temporomandibular joint disorders, or sinusitis, among other conditions. There are noticeable differences with respect to the acquisition of intraoral images. In the panoramic view, the X-ray tube and the film or sensor carrier rotate synchronously but in the opposite direction about a pivot point, as shown in Figure 1.1e. Some sub-images are taken during this movement and then combined. Each sub-image is taken in an orthogonal projection of the imaged area, so the final image is not expected to present a significant distortion. However, it is very difficult to obtain a completely undistorted image, as the curvature of the arches can be different in each person.

(a) Periapical.

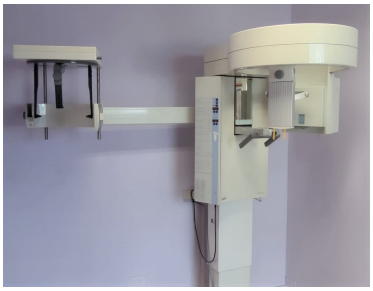


(b) Bite-wing.



(c) Occlusal.

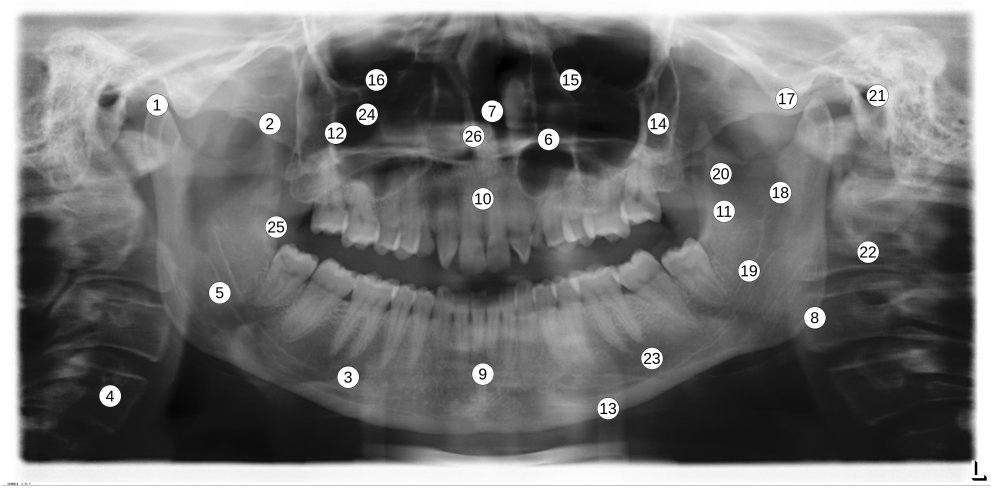
(d) Panoramic or OPG.



(e) OPG acquisition device.

Figure 1.1: Different views of dental radiographs.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



1. Mandibular condyle
2. Coronoid process
3. Mental foramen
4. Cervical vertebra
5. Mandibular canal
6. Hard palate
7. Nasal fossa
8. Angle of the mandible
9. Genial tubercles
10. Incisive foramen
11. Soft palate
12. Zygomatic process
13. Inferior border of the mandible
14. Zygomatic bone
15. Infraorbital canal
16. Rim of the orbit
17. Articular eminence
18. Nasopharyngeal airway
19. Oropharyngeal airway
20. Medial pterygoid plate
21. External auditory meatus
22. Ear lobe
23. Submandibular fossa
24. Maxillary sinus
25. External oblique ridge
26. Nasal septum

Figure 1.2: Anatomic structures observable in a dental panoramic radiograph.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Due to the difference in the acquisition process, the different views do not have the same quality. Intraoral images are, in general, sharper, better enlightened, and less noisy than panoramic images, which are usually more blurry and lower in contrast. Even all dental images are sensitive to the acquisition procedure, this is especially remarkable in panoramic acquisitions, where subject positioning is the most common error. In an ideal setup, the subject is placed centred in a specific focus plane, with his chin pointed slightly down. If the chin is pointed front or pointed up, the arches are flattened or even reversely curved, the apical area of upper incisors can be out of focus, and the condyles are moved outwards. On the contrary, if the chin is tilted down, the arches become too curved, the apices of the upper incisors are also out of focus, and overlapping of hyoid bone and mandible can be noticed. If the patient tilts the head sidewards, both image halves become asymmetric, and if there is a turning sideward, proximal surfaces appear overlapped. If the tongue is not aligned with the palate, a dark shadow can be seen over the maxillary teeth, between the palate and the tongue dorsum, and if the patient is positioned in a slumped position, a white pyramid-shaped shadow centred at the lower part of the picture can appear [62, 146, 190].

Patient movements are also a big problem in the acquisition of an image of any kind because the light is recorded over a period during which the subject is expected to remain immobile. They are even more problematic in panoramic radiographs because several sub-images have to be taken to finally form an image, so the final exposure time is greater. Logically, the larger the exposure time, the more noticeable the movements of the patients are. The causes of these movements comprise respiratory motion, swallowing, and discomfort. Although scanning devices have some effective features to avoid patient movements, such as headrests, temporal fixations, or chin supports, they cannot be completely avoided [93].

Panoramic imaging produces lower quality images, and it is prone to errors, though it has many advantages. To cover the whole dentition without a panoramic radiograph, experts would need to do a full-mouth series, that is, a set of 18 intraoral images —4 bite-wings, 8 posterior periapicals, and 6 anterior periapicals. This practice would obviously imply higher storage requirements and lead to a more difficult assessment and a higher radiation dose than a single acquisition [227]. Because panoramic images cover not only the whole dentition area but also adjacent bones, they are also useful to the assessment and monitoring of other conditions. Moreover, technology innovations in recent years led to sharper panoramic radiographs, so that the differences in quality between intraoral and extraoral views have become smaller and smaller [191].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

It is worth noting that other radiological techniques for dental imaging have been used in the last few decades. One example is the cone beam computed tomography (CBCT), which is a particular case of computed tomography (CT) imaging where the X-rays follow a divergent pattern resulting in a cone. Due to the high quality, it has made it possible to improve numerous dentistry procedures such as endodontics, implantology, or orthodontics. However, the implementation of CBCT is relatively new when compared to panoramic imaging, for example, which leads to a lack of population-specific CBCT datasets and thus hinders its application in forensic procedures. Moreover, a CBCT scan reportedly leads to a higher radiation dose and cost than a dental panoramic radiograph [139, 175], and its availability is also more limited. Thus, the methods presented in this PhD thesis were focused on dealing with panoramic images.

1.3.1 Dental radiology for chronological age estimation

Dental radiographs provide a simple, non-invasive, and reproducible way to assess the development status of the dentition and adjacent bones, both in living individuals or deceased people. The state of the art of radiological approach to dental age estimation includes different methods, which vary depending on the assessed dental development stage. Therefore, there are three well-differenced cohorts: the first covers the fetus, newborns, and individuals in the postnatal stage; the second includes children and teenagers; and the third includes the adult population. The absence of developmental markers not only in the adult teeth and skull but the whole skeleton makes the degeneration-based methods the only solution for chronological age assessment in the adult population. However, their accuracy is limited as they are aimed at quantifying highly variable processes.

To make an accurate dental age estimation in subadults and young adults, the use of proper indicators for each age range is necessary. For example, in the pre- and postnatal stage, calcification of the deciduous teeth at crown formation time is used, knowing that this period starts more or less at the 14^{th} week of the intrauterine life. Krauss and Jordan [118] built several tables where they proposed ten development stages based on the mineralisation of some deciduous teeth and the permanent first molar. In both childhood and adolescence, there are a lot of developmental markers that can be used for age estimation: development and eruption of deciduous teeth, development of permanent teeth, loss of deciduous teeth, and eruption of permanent teeth. Finally, between the age of 16-18 and the age of 25, the way to

USC UNIVERSIDADE

assess dentition development is through the third molar development, specifically its eruption and root formation.

The clinical eruption, as the most easily observable indicator of tooth development, is the first reference to dental age estimation, and it corresponds to the time at which the teeth are visible in the oral cavity [34]. Age estimation based on the dental eruption pattern has been introduced by Haavikko [92], who compiled a table with the relationship between the alveolar eruption timings —based on the Lewis & Garn criteria [133]—, clinical eruption timings, and a modified version of the stages of development proposed by Gleiser & Hunt [87], all of them evaluated through panoramic imaging. However, the dental eruption is a relatively short period, both in deciduous and permanent dentition, so it is a very limited technique for age estimation.

The first attempt to study dental development in a longer period was made through the atlases [229]. In these tables and graphs, the authors divided a given period of the life of an individual into different sequential stages of development by using several factors such as dental mineralisation. These studies assess, tooth by tooth, some important events, such as the odontogenesis, enamel and dentine apposition, crown completeness, both alveolar and clinical eruption, or root completeness. Although generic times for a given population are presented, they are useful to estimate the age of an individual and determine the development of each tooth, showing if it is advanced or delayed concerning the average. The most recent example of an atlas is the London atlas [6], a very extensive description of human tooth development from 28 weeks of the intrauterine period to the age of 24. It reports the median stage of tooth development —according to the atlas of Moorrees *et al.* [152]— and eruption —relative to the alveolar bone level— for all age values. The London Atlas has been widely used for chronological age estimation from dental radiographs, mostly in subadults aged from five to 16 [85, 147].

Besides atlas methods for age estimation, other authors established scoring systems, being the Nolla method [163] one of the most well-known. In her work, the researcher proposed a system to divide the dental development of children and adolescents into 11 stages and 41 substages, where the first one (stage 0) refers to the absence of dental crypt and the last stage (stage 10) represents the full formation of the roots and the full closing of the apices. Each stage was given a numeric score, so the sum of the scores of all teeth gives a measure of dental maturity. This value can be translated into a numeric age estimation —with a year precision— through a table specifically built by the researcher. This method has been tested in

several studies [17, 20, 141], showing good results in general. Nevertheless, the large number of substages may make the method hard to follow and lead to a low intra- and inter-observer agreement.

Another commonly used scoring method for age estimation was proposed by Demirjian *et al.* [60]. Also based on a tooth development ranking, the authors proposed 8 alphabetical stages —from A to H— for each of 7 left-mandibular teeth —all but the third molar— associated with specific criteria. These stages can be transformed into ratings through sex-specific tables and added together to obtain the so-called *maturity score*. This score is represented through percentile charts concerning the age of the individuals, so they can be used to convert from *maturity score* to chronological age. In a posterior work [59], the authors revised their method and proposed three new methods based on the original, with a view to wider applicability.

In 2006, Cameriere *et al.* [27] proposed a novel approach for age estimation. They found that the correlation between age and the openness of dental apices is significant and negative. The degree of apex openness of the second premolar, together with the number of teeth with fully closed apices and the sex, was used to develop a regression formula that computes the age estimation.

Given the high availability of teeth when finding unknown remains and the high accuracy obtained when they are used in subadult age estimation, the use of other adjacent bones has been little explored and mostly focused on adult age estimation through morphological and histological methods. Even so, there are some works in this regard reporting a noticeable correlation of some radiological indicators with the chronological age, such as the mandibular ramus length [57], the gonial angle [198], or the condylar growth [22].

Although morphological age estimation methods often require an ex-vivo tooth analysis, some morphological features are radiologically observable, such as the apposition of secondary dentine in teeth. In this regard, Kvaal *et al.* [124] and Cameriere *et al.* [26] proposed to use the relationship between the pulp chamber and the whole tooth —in terms of height and area, respectively— as an indicator to estimate the age. Cameriere *et al.* [25] also used two different projections (labio-lingual X-rays and lateral X-rays) to add volumetric information and so improve the results, reporting that the best tooth to take both measurements was the canine.

Due to the vast literature on the use of dental radiography for chronological age estimation, the age estimation errors and differences give an idea of which populations are more advanced or delayed in terms of dental development. In general, populations in which some age

estimation methods incur an overestimation, show a more advanced dental development than the reference sample of the method. Equally, an underestimation implies that the dental development of the population is delayed with respect to the reference sample. For example, in the knowledge that Demirjian's method usually overestimates age, the French-Canadian community (which has Caucasian and Amerindian origin) seems to show a delayed dental development in relation to populations of most other races, like Caucasian [8], Asian [234], Asian + Caucasian [35], and African [176].

Some authors have also been able to assess not only the ethnic variability but also the impact of socio-geographic factors through the application of radiographic methods. Martin-de las Heras *et al.* [145] pointed out that the latter influences even to a greater extent. Chaillet [36] used a large database of radiographs from eight different countries to propose international scores to estimate the age of the children whose ethnic origin is unknown. These authors found three main groups in order from the fastest to the slowest dental maturity: Australia (which has the fastest maturity), France and Finland; Belgium and Sweden; and French-Canada and Korea. All these findings suggest that there is a need for population-specific tables and/or population-specific formulas.

Due to the misleading results yielded by the radiographic methods when applied to different samples than those for which they were developed, great efforts have been done to adapt them to new populations. For example, Demirjian's scores were adapted to specific populations by Willems [230], Teivens [213], and Chaillet [37], among others, resulting in improvements also in other populations [8]. Cameriere's formula was also adapted to increase the age estimation performance, being the Indian formula one of the best-known examples [183].

1.3.2 Dental radiology for sex estimation

Dental radiology plays an important role in the sex determination task, as it allows to assess both qualitatively and quantitatively the dimorphic features not only of the teeth but also of adjacent bones. As a counterpart to the estimation of chronological age, where teeth are the main development indicators, the research regarding sex determination is not so focused on the teeth but more evenly spread across many radiologically observable structures.

As mentioned previously, researchers have reported a great level of sexual dimorphism in the mandible. The most simple approach to assess it radiologically is through linear measurements in dental panoramic radiographs involving well-known anatomic landmarks and regions. Regarding the mandibular ramus, a set of measurements have been used. The

USC UNIVERSIDADE

minimum ramus breath and the condylar height have proven to be the most sexually dimorphic measurements [101]. The maximum ramus breadth and the ramus height with respect both to the condyle and the coronoid process have also been studied in this regard [101, 192]. The ramus flexure as a sex determination indicator, initially proposed by Loth & Henneberg [140], was applied to dental panoramic images with successful results, although the researchers agree that it should not be used as the only sex determination indicator [12]. More complex methods such as mandible shape analysis have also been applied to study the sexual dimorphism of the human mandible [165]. However, their application is more difficult as the available methods do not rely on radiographic imaging, and thus manual dry bone measurements or the 3D digitisation of a set of mandible landmarks are required [49, 76, 79].

Other highly dimorphic bone structures observable in dental panoramic radiographs are the maxillary sinuses, which are reportedly wider and higher in males than in females, and thus are useful for sex determination [182]. Measurements involving the mental foramen —specifically the distance to the lower border of the mandible— have also been confirmed as good radiologic indicators for sex estimation [38].

The potential usefulness of teeth to carry out sex estimation has been analysed in numerous studies, although most of them rely on measurements performed on skeletal remains or dental casts [31]. The only study which applies an odontometric approach in dental panoramic radiographs was performed by Capitaneanu *et al.* [30], who analysed a set of linear measurements and ratios in all permanent teeth in left quadrants, reporting that the mandibular and maxillary canines show the greatest dimorphism, especially in terms of tooth length. Paknahad *et al.* [171] studied the predictive capabilities of the deciduous dentition to estimate the sex through radiologic measurements in periapical images, reporting a moderate accuracy which is in line with the previously mentioned difficulty to establish the sex in subadults. Volumetric measurements of the teeth have also been assessed to estimate the biological sex, relying all these studies on CT-based scans [54, 212].

1.3.3 Automatic methods for age and sex estimation from dental radiographs

The last advances in the image processing field have had a great impact on the way medical images are used [14], which is especially noticeable in the case of radiography given the long time it has been used in clinical procedures. The turning point has probably been marked by the advances in neural networks combined with the development of more powerful computing

USC UNIVERSIDADE

technologies, which led to the rise of a new family of techniques known as deep learning [127]. Within deep learning, convolutional neural networks (CNNs) stand out as a group of methods that have demonstrated a convincing capacity to manage medical images in an accurate yet scalable manner.[137]. They also represent an evolution of the classic multilayer perceptron that was developed to exploit the spatial relationships between the input values [128].

The architecture of a CNN for image processing varies greatly depending on the purpose, but it usually consists 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. Therefore, 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 [128].

Dental radiography has benefited from these developments, from the acquisition process improvement to the most complex methods for automatic diagnosis. The simplest example is the inclusion of image enhancement algorithms to overcome acquisition issues, such as a lack of brightness or contrast [84], with two main objectives: allowing an easier and more reliable diagnosis by a clinician and preparing the images for the application of further image processing methods.

Other methods focused on detecting specific structures or tissues in the radiographs. In this regard, lots of studies developed automatic approaches for detecting teeth [107, 216, 231], the mandible [1], or mental foramen [113], among others. The possibility of detecting relevant image regions with high precision made many works go one step further by proposing computer-aided diagnosis systems. This includes methods for diagnosing mandible and tooth fractures [81, 206], periodontal disease [214], osteoporosis [130], cysts and tumours [125], dental caries [166], apical lesions [66], and many more. The accuracy of those methods is reportedly comparable to the diagnosis performance of a clinician, enforcing their suitability as a decision support tool.

The recent advances in image processing have also made it possible to automate dental age estimation methods to a higher degree. Čular *et al.* [52] described a method to perform age

estimation with a numerical and continuous output, which also worked on the third molars. Specifically, the authors proposed the use of an Active Appearance Model to localise the third molar and parameterise its shape and texture. In a second step, these parameters were fed into a neural network to estimate the chronological age. As both steps work in a fully automatic way, there is no need for human intervention.

As in most topics involving image processing, CNNs have helped not only to automate some tasks but also to improve their performance. Regarding age estimation, De Tobel *et al.* [58] developed a staging system for the third molar based on modified Demirjian's stages and used a deep neural network to classify third molar image crops into one of those stages. This method only required a minimum intervention of the expert to crop the region of interest where the third molar is placed. This approach was refined by Merdietio *et al.* [149] by substituting the manual crop step with the DenseNet CNN, which allows running the estimation in a fully automatic way. Banar *et al.* [13] developed a similar method, with a slightly more complex third molar segmentation, in which the tooth is first located and then segmented. Kim *et al.* [116] followed a similar approach to De Tobel *et al.*'s. The authors also developed a two-step approach which firstly requires a manual crop of the third molar, although in this case the four third molars are required. In the second step, each of the four teeth is classified into different age groups, and the classifications are merged through a majority voting system. The authors established two different age group divisions: the first grouped subjects younger than 20, the subjects aged 20 to 49 and those over 50; the second split the middle group into three subgroups, namely the subjects aged 20 to 29, the subjects aged 30 to 39 and those aged 40 to 49.

Although deep learning methods had already been introduced in the previously mentioned studies, De Back *et al.* [55] proposed the use of a CNN —specifically a Bayesian CNN— as the only step to estimate the chronological age. Therefore, the expert does not need to specify which features of the image should be taken into account, as the network focuses automatically on those regions which contribute the most to age estimation. Furthermore, the estimation process can be executed even if one or several teeth are missing.

In the same way as age estimation problems, modern techniques in image processing have enabled the development of completely automatised procedures for sex classification. For example, Ilić *et al.* [100] relied on a widely used CNN architecture —known as VGG16— to automatically estimate the gender from a dental panoramic radiograph. Moreover, Milošević *et al.* [151] compared several state-of-the-art CNN architectures, reporting that VGG16 was

the most suitable for performing sex estimation. Ke *et al.* [114] proposed a three-step pipeline to combine the image features learned by a VGG16 architecture at different scales and so overcome the problem of shallow information loss.

All these studies demonstrated the high potential of computational methods —especially those based on deep learning techniques— not only to detect and measure the structures observable in dental radiographs but also to assess higher-level characteristics such as their shape and appearance and infer a clinical diagnosis or a biological profile of the subject to which the image belongs. Although the reported performance of the proposed methods is remarkable, their complexity is often associated with an interpretability decrease, which is crucial for their applicability in a real scenario [33].

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CHAPTER 2

JUSTIFICATION AND OBJECTIVES

The problem of chronological age and sex prediction on radiographic samples has been studied for decades with successful results. Currently, there are two main groups of methods in this regard. The manual methods are easy to apply and understand, as they often consist in taking well-known measurements and applying a simple formula to predict age and/or sex. However, they are time-consuming —in the order of several minutes for analysing each sample—, which may be a handicap when conducting large-scale studies. On the other hand, automatic methods —especially those based on deep learning— require minimal or no expert involvement, which speeds up the whole process, and usually show a greater accuracy. Though, they are more difficult to explain as the image features used by the method are unknown to the expert.

This lack of explainability is hindering the deployment and application of automatic methods, as it is a priority for regulatory agencies [122]. In this regard, there are two main solutions: focusing on simple and interpretable manual models and assuming a loss in the accuracy achieved, or developing very accurate automatic models and providing them with mechanisms to help interpret the results obtained.

The main purpose of this PhD Thesis is to develop a fully automatic method based on deep learning techniques for estimating both chronological age and sex from dental panoramic images to obtain accurate yet explainable results. Furthermore, the experimentation to be carried out should serve to prove not only the usefulness of dental panoramic radiographs as a whole for age and sex estimation, but also the suitability of individual skeletal structures such as the mandible and the dentition. Specifically, three different objectives have been defined in this regard:

USC UNIVERSIDADE

O1. A convolutional neural network for chronological age and sex prediction from full dental panoramic radiographs

As this is the initial proof of concept, the proposed convolutional neural network will keep a one-stage architecture, that is, it will produce the age and sex estimations directly from the input image, without any intermediate results. The main hypothesis, which has been already proved in the clinical domain, is that the development of the tooth and the adjacent bones follows different timings for males and females, and that the development of secondary sexual features does not happen until puberty. This means that both age and sex depend notably on each other. Thus, the proposed architecture has to detect both developments- and sexual-related features to improve the age and sex estimation results.

O2. An automatic methodology to perform chronological age and sex estimation from the mandible shape in dental panoramic radiographs

Based on forensic research reporting that mandible shape is a good indicator of both age and sex, a methodology for automatic mandible shape detection and its subsequent use for chronological age and sex estimation will be developed. It will be based on deep learning techniques and statistical shape analysis, respectively. The proposed method has to yield not only accurate age and sex predictions, but also a visual explanation of how the mandible shape varies according to these two variables.

O3. A convolutional neural network to estimate chronological age and sex estimation from teeth in dental panoramic radiographs

Although in the first objective the whole image will be used for the age and sex estimations, there is a need to quantify the specific contribution of each tooth in those estimations. The method developed in this objective will be based on deep learning techniques, and it has to automatically detect the teeth in the image and use their features to estimate both age and sex. Again, the output of the proposed methodology will be comprised not only of the quantitative estimations, but also an explanation of how these estimations are obtained.

The structure of this PhD dissertation is organised as follows: on top of Chapters 1 and 2, this document covers the description of the employed database, the three main contributions of this PhD Thesis and additional chapters for conclusions and future work.

- **Chapter 3** describes the OPG database built during this PhD Thesis to develop and evaluate the proposed methods for age and sex prediction, reporting the process followed to collect the images and the image characteristics.
- **Chapter 4** describes two one-stage CNN architectures for chronological age estimation and one one-stage CNN architectures for sex estimation, all of them working directly on dental panoramic radiographs.
- **Chapter 5** details a pipeline for an automatic description of the mandible shape composed of a CNN-based method for detecting anatomical landmarks all over the mandible contour and four different mandible descriptors for automatic shape characterisation.
- In **Chapter 6**, a methodology for predicting the chronological age and sex of a subject from the tooth appearance is proposed, which combines uncertain per-tooth predictions to improve the results.
- **Chapter 7** summarises the contributions of this PhD Thesis, the results, and the potential applicability in real scenarios.
- **Chapter 8** concludes with a discussion of the possible lines of future research.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 3

OPG-AS: A DATABASE OF DENTAL PANORAMIC RADIOGRAPHS

The success of any supervised machine learning approach —i. e., that which needs to learn from a set of given examples— has proven to depend not only on its capabilities, but also on the quality of the examples used both to train and evaluate it. Obtaining a high-quality database requires two main features. First, the examples of the dataset need to be curated in a way that noisy and unrepresentative data is avoided as much as possible. Second, the variability of the dataset should be such that as many cases as possible are represented. Compliance with these two precepts is key to ensure that the suitability of the proposed methodology is assessed in the most realistic scenario. This chapter describes the construction of the OPG database for Age and Sex determination (OPG-AS), a set of dental panoramic radiographs used during this doctoral thesis.

Every single OPG used in this PhD was provided by the Radiology Unit of the School of Medicine and Dentistry, Universidade de Santiago de Compostela (Spain), and acquired with the direct digital panoramic unit shown in Figure 1.1e (Orthophos Plus DS; Sirona USA, Charlotte, NC). The device was configured through the included P1 preset, which corresponds to a total acquisition duration of 24 seconds, of which 14.1 seconds are used for the radiation exposure. The greyscale images obtained with this program were 1,552 pixels high, while their width varied between 2,400 and 3,200 pixels.

USC UNIVERSIDADE

A careful anonymisation process was conducted to isolate any subject identification data from the image metadata. The preserved information for each image consists of an automat-

ically generated image identifier, the sex and the date of birth of the subject, and the date the image was acquired. This information made it possible to calculate the chronological age of the subject at the acquisition time, with a precision of one day. Moreover, the whole image collection process was performed in accordance with the approval of the ethical committee of the Universidade de Santiago de Compostela (USC), Spain (Figure 3.1), and following the 1964 Declaration of Helsinki and its later amendments [232].

The dataset was reviewed to remove images with an 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.

It is important to note that the OPG-AS database used in the experiments described in the following chapters has not been available from the outset of this PhD, but has been built up in parallel to the conducted research, and is still growing to this day. This means that the number of images available in the first experiments was lower than in the last ones. Furthermore, the great effort needed to label the images for some described research also led to the use of reduced versions of the database. As a result, the database size varied substantially across each experiment.

At the time of writing, the database is composed of 3,890 images of subjects aged 3.5 to 94.7 years, corresponding to the demographic distribution shown in Table 3.1. It can be seen that more than 60% of the cases are concentrated in the 5-25 age bracket, whereas the images belonging to subjects younger than five and older than 80 barely account for 1% of the total. This is mainly because most of the subjects are grouped into two big clusters. The first one corresponds to the patients of the clinic located within the School of Dentistry, which specialises in the care of children, and thus it is composed of patients aged between six and 16 years, approximately. The second cluster comprises the under- and post-graduate students who are required to carry out a full study of their dentition through the dental panoramic radiograph as part of their studies. Therefore, it contains images of subjects aged 20 to 25 years. Moreover, the dataset is highly balanced in terms of sex distribution, except for the age cohort 20-25, where the number of female images is almost twice as high as the number of male images. Regarding the race of the subjects, all images belong to Spanish Caucasian subjects.

USC UNIVERSIDADE

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

VICERREITORÍA DE INVESTIGACIÓN
E INNOVACIÓN
Oficina de Investigación e Tecnoloxía
Servizo de Convocatorias e Recursos Humanos de I+S
Edificio CACTUS – Campus Vida
15782 Santiago de Compostela
Tel. 981 547 040 - Fax 981 547 077
Correo electrónico: cittinfo@usc.es
http://imaisd.usc.es

JOSÉ MANUEL CIFUENTES MARTÍNEZ, PRESIDENTE DO COMITÉ DE BIOÉTICA DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA,

INFORMA:

Que o proxecto de investigación para: **"Extraer características relevantes de radiografías dentais panorámicas con varios obxectivos, entre os que destacan a predición da idade cronolóxica, a detección de lesións, o diagnóstico de enfermidades bucodentais ou o seguimento de procesos clínicos",** do que son investigadoras responsables Dona **Inmaculada Tomʹñas Carmona** e Dona **María José Carreira Nouche,** ten sido examinado polo Comité de Bioética de esta Universidade, cumprindo o seu protocolo experimental os requisitos éticos esixidos.

Este documento non exime da obtención de permisos ou autorizacións e o cumprimento de outras normativas de aplicación.

Lugo, 18 de xullo de 2019.

Figure 3.1: Approval of the ethical committee of the Universidade de Santiago de Compostela (USC) to use the OPG-AS database.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

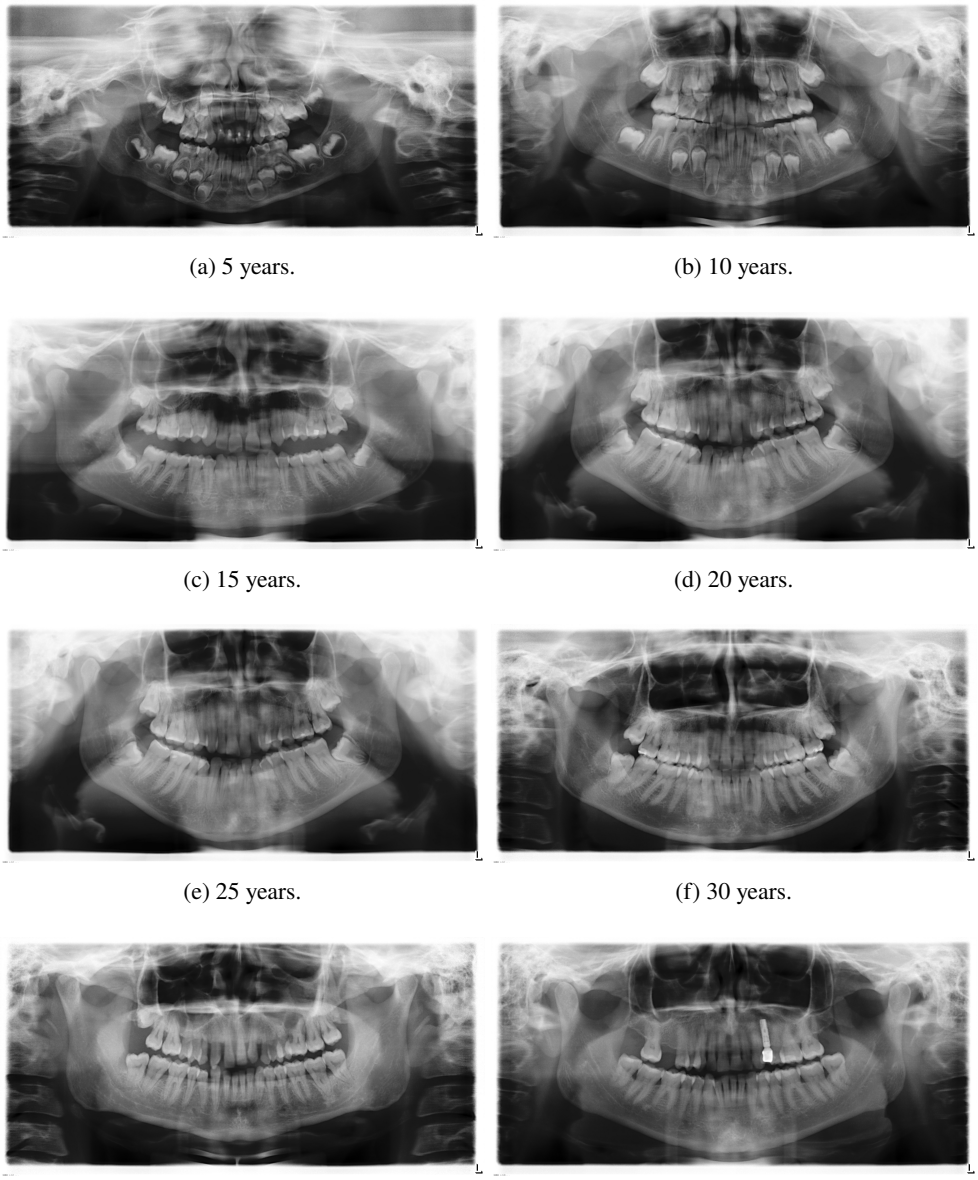
Table 3.1: Demographic characteristics of the OPG-AS database at the end of this PhD Thesis.

Age group	Male	Female	Total	Percentage (%)
[0, 5)	4	2	6	0.15
[5, 10)	325	297	622	16.00
[10, 15)	447	439	886	22.79
[15, 20)	202	201	403	10.37
[20, 25)	157	323	480	12.35
[25, 30)	103	159	262	6.74
[30, 35)	83	112	195	5.02
[35, 40)	57	101	158	4.06
[40, 45)	65	90	155	3.99
[45, 50)	66	78	144	3.70
[50, 55)	65	86	151	3.88
[55, 60)	54	71	125	3.22
[60, 65)	52	44	96	2.47
[65, 70)	43	52	95	2.44
[70, 75)	26	36	62	1.59
[75, 80)	8	22	30	0.77
[80, 85)	7	6	13	0.33
[85, 90)	2	2	4	0.10
[90, 95)	0	1	1	0.03
Total	1,766	2,122	3,890	100.00
Percentage (%)	45.42	54.58	100.00	

It is worth remarking that this set of images contains some examples revealing observable conditions, namely, orthodontics, prostheses, implants, external elements like earrings, fillings, endodontics, caries, missing teeth, or dental root remains, among others. Also, the quality of the images varies depending not only on the exposure parameters, but also on patient's movement and positioning.

In the following, a set of examples extracted from the database are shown. Specifically, some images belonging to patients of different age cohorts are presented in Figure 3.2. Similarly, Figure 3.3 depicts some examples of the radiologically observable conditions mentioned above.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) 5 years. (b) 10 years.

(c) 15 years. (d) 20 years.

(e) 25 years. (f) 30 years.

(g) 40 years. (h) 50 years.

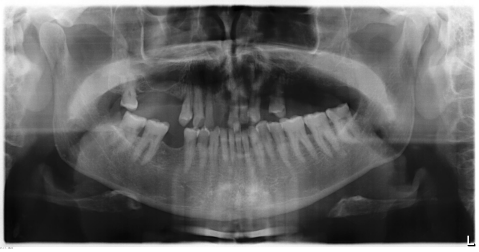
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figure 3.2: OPG examples by age.

(a) Orthodontics.

(b) Prostheses.

(c) Caries.



(d) Missing teeth.

(e) Endodontics.

(f) Implants.

(g) Tooth fillings.

(h) Dental root remains.

Figure 3.3: OPG examples by condition.

CHAPTER 4

ONE-STAGE CNN FOR AGE AND SEX PREDICTION FROM ORAL CAVITY STRUCTURES

As mentioned in Chapter 1, 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 orthopantomograms 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 sets of good radiological quality without any conditioning dental characteristic. In this chapter, two one-stage fully automatic methods to estimate the chronological age of a subject from the OPG are proposed. The first —Dental Age Network (DANet)— consists of a sequential CNN path to predict the age, while the second —Dental Age and Sex Network (DASNet)— adds a second CNN path to predict the sex and uses sex-specific features to improve the age prediction performance. Both methods were tested on a set of 2,289 OPGs 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 DASNet outperforms 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 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

USC UNIVERSIDAD

manual age estimation methods, showing significantly fewer problems of over- or under-estimation. Consequently, we conclude that DASNet can be used to automatically predict the chronological age of a subject accurately, especially in young subjects with developing dentition. Moreover, additional experimentation is conducted in Appendix B (Section 4.7) to assess the validity of the proposed approach in the problem of sex classification. In this regard, Dental Sex and Age Network (DSANet) architecture is proposed as a modification of DASNet through which the main is set to predict the sex and the secondary path to predict the age. This setup was evaluated in an extended version of the dataset, comprising 3,400 images.

The contents of this chapter are extracted from the following publications, the reproduction rights of which have been granted by the CC BY licence (Figures 4.1 and 4.2):

> Vila-Blanco, N.[*,†], Carreira, M. J.[*,†], Varas-Quintana, P.[‡], Balsa-Castro, C.[†,‡], and Tomás, I.[†,‡] (2020). Deep neural networks for chronological age estimation from OPG images. IEEE transactions on medical imaging, 39(7), 2374-2384. 103615, 2020. DOI:10.1109/TMI.2020.2968765

> Vila-Blanco, N.[*,†], Vilas, R. R.[*], Carreira, M. J.[*,†], and Tomás, I.[†,‡] (2020). Towards deep learning reliable gender estimation from dental panoramic radiographs. 9th Starting AI Researcher's Symposium (STAIRS), European Conference on Artificial Intelligence (ECAI), Santiago de Compostela, Spain, 2020. ISSN: 1613-0073

[*]Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) and Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.
[†]Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Spain.
[‡]Oral Sciences Research Group, Special Needs Unit, Department of Surgery and Medical-Surgical Specialities School of Medicine and Dentistry, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.

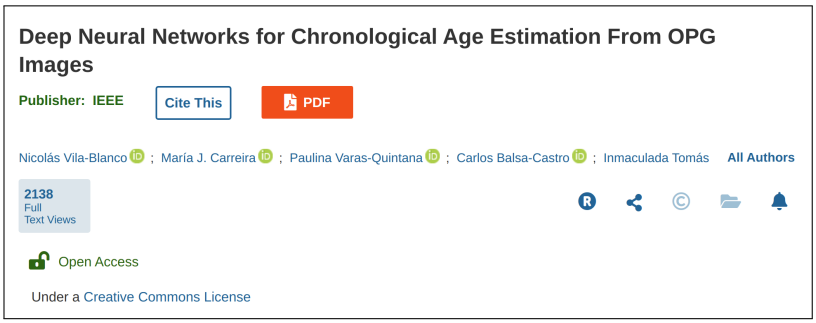


Figure 4.1: Licencing information of the publication [222], from which the contents of Chapter 4 are taken from.

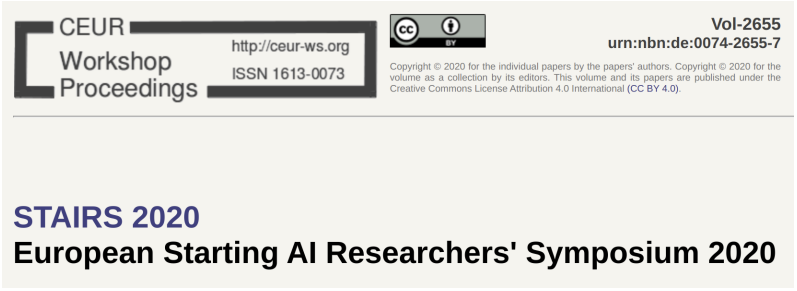


Figure 4.2: Licencing information of the publication [224], from which the contents of Chapter 4 are taken from.

4.1 Introduction

As stated in Chapter 1, chronological age estimation is a key task in a variety of scenarios, such as trials, migration management, or adoption processes. Together with sex determination, they are essential for forensic identifications. The relevant literature describes several methods for conducting chronological age and sex estimations by analysing dental maturation in OPGs, which provide a panoramic view of the entire dentition and surrounding bones in the oral cavity [27, 60, 163]. All these methods have been compared in a variety of populations [17, 20, 70, 82], 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 sex influences the pattern of dental development. Indeed, it has been proved that girls reach a particular development stage before boys [83], which is especially noticeable with permanent teeth [138]. Consequently, the methods mentioned above take sex 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 [60, 163]; in other cases, the formulae to regress the age include sex as an input variable [27, 28].

Although these age and sex prediction methods perform reasonably well and are widely used in the scientific community [19, 235], all of them rely on manual measurements or classifications, which require much time and effort. Indeed, in an example described in [111], 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 [24, 82, 148]. 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 evaluations required by some manual methods. An example is the approaches of Nolla [163] or Dermijian [60], which are not based on numerical quantification, but the assignment of an ordinal stage or category associated with dental maturity or mandible shape. Furthermore, the datasets used to initially develop and subsequently validate these methods had strict exclusion criteria, including the absence of teeth, endodontics, caries, or cysts. The outcome is techniques that do not work with subjects who have any of these conditions.

In the last few years, deep learning techniques have been used to automate tasks involving OPGs. All of these techniques can automatically extract image features for a variety of purposes, including: the segmentation of the dental region [203]; the detection and classification of individual teeth [107, 167, 216]; the detection of previous treatment, e.g., endodontics [123]; the image reconstruction where a patient was badly positioned [65]; and the diagnosis of osteoporosis [45] and jaw tumours [178].

In this chapter, a novel DL approach for chronological age prediction 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 OPGs without any preprocessing, and does not require anatomic measurements or other manual annotations to obtain accurate results.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.2 Materials and methods

A set of 2,289 OPGs included in the OPG-AS database presented in Chapter 3 was used to conduct the experiments. Table 4.1 shows the range of ages in the dataset, with most of the images corresponding to patients younger than 30. Although 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—, additional images from older subjects were included in the dataset to enable us to perform a number of experiments to confirm this belief.

We also considered the exclusion criteria applied to OPGs in earlier studies that used traditional chronological age calculation methods [215]. Of all the images in both datasets, some images portrayed the presence of one or more of the conditions mentioned in Chapter 3: orthodontics, prostheses, implants, external elements like earrings, fillings, endodontics, caries, missing teeth, dental root remains, and distorted or blurred radiographs (see examples in Figure 3.3). However, unlike the classical studies, we did not exclude these images. Instead, we labelled the images of the dataset where these conditions are observable as *imperfect* —1,752 images— and the remaining images as *perfect* —537 images—, and used this information for conducting a study regarding the influence of those conditions in the age estimation performance.

Table 4.1: Age and sex distribution of the OPG dataset, after discarding images labelled with the wrong acquisition or birth dates.

Age (years)	**Male**	**Female**	**Total**
[5, 10)	210	206	416
[10, 20)	474	489	963
[20, 30)	172	332	504
[30, 40)	64	81	145
[40, 50)	35	51	86
[50, 60)	35	49	84
[60, 70)	28	30	58
[70, 90)	12	19	31
Total	1,030	1,257	2,289

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

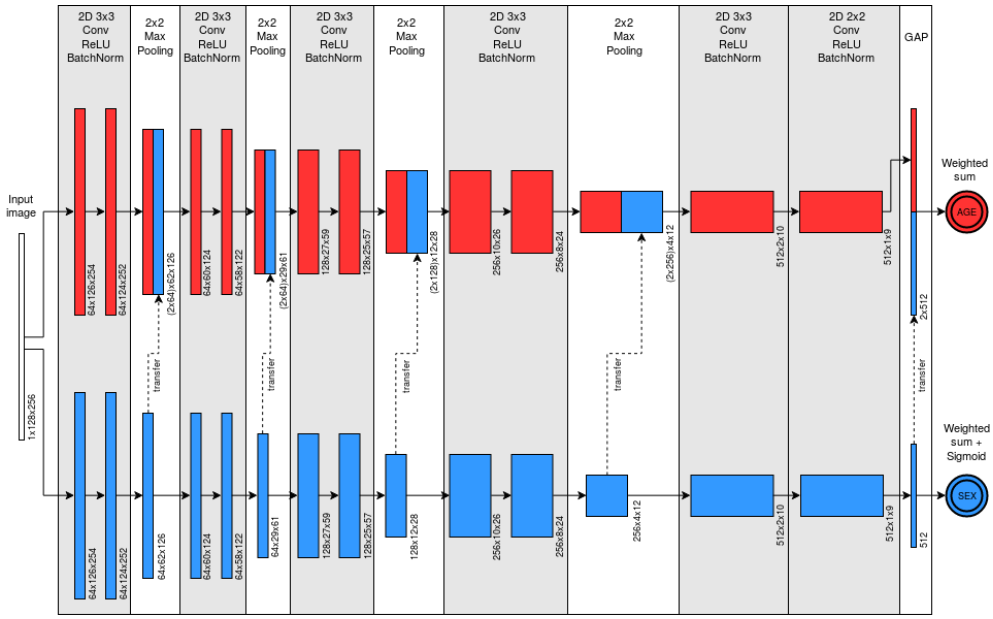


Figure 4.3: DASNet architecture. In red, the age prediction path represents the DANet approach. This consists of a set of convolutional and pooling layers with which to extract image features at different scales and, ultimately, determine a subject's age in days. In blue, the sex classification path, which complements the DANet approach and leads to the development of the DASNet 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.

The images were 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 smaller and the processing procedure is, consequently, more efficient.

In the present chapter, two different one-stage CNN architectures were developed using the Keras toolkit [42] aimed at performing chronological age estimation. The first of these, known as DANet in what follows (red items in Figure 4.3), comprises a single path that works at five different image scales. Two blocks composed of a 3x3 convolution layer and a rectified linear unit (ReLU) function [160] are applied sequentially at each scale. The application of two consecutive 3x3 convolutions is widely thought to be a less complex replacement for a single

5x5 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 apply a 2x2 max pooling layer to halve the resolution of the output by only preserving the maximum pixel in each 2x2 output block. No padding is added to the images to preserve the size of the feature maps after the convolutions, which causes a 2-pixel loss in every convolution step. As precise feature localisation is not the objective of this research, both training and inference speeds are improved without having an impact on overall performance. For the final scale, the second 3x3 convolution is replaced with a 2x2 version and the final max pooling layer by a global average pooling (GAP) layer. This produces the average value of each feature map. The values obtained are then combined in a weighted sum to predict a subject's age in days.

The second approach is known as DASNet and has a more complex architecture with the main aim of improving age estimation performance by incorporating sex-specific features. As the dental development pattern is different in boys and girls, a second path identical to DANet is added to the network to classify the images according to the sex of the subject (blue items in Figure 4.3). The prediction of this path can be seen as the probability that a given image belongs to one of the classes —in this work, we decided to encode the male class with a 0 value and the female class with a 1 value—, and so the final weighted sum is followed by a sigmoid activation to force the output to lie within the range [0,1]. The goal of this second path is 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 the entire DASNet model.

In the DASNet configuration (Figure 4.3), the sex path shares 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 is concatenated depth-wise to the output of the same pooling layer in the age path. Thereafter, the age path applies a convolutional layer to combine both age and sex information. This is also the case for the GAP layer in the sex path, the output of which is 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 sends information to the age path, this path decides 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 [103] (see Figure 4.3). 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.

USC UNIVERSIDADE

Second, as this can affect the expressive power 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.

4.3 Experiments and performance evaluation

To assess the performance of the proposed architectures in a proper way, a common training approach was defined. The learnable network parameters were initialised with the Xavier uniform initialiser [88], 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 [236] 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 (Figure 4.4). 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: 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 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.

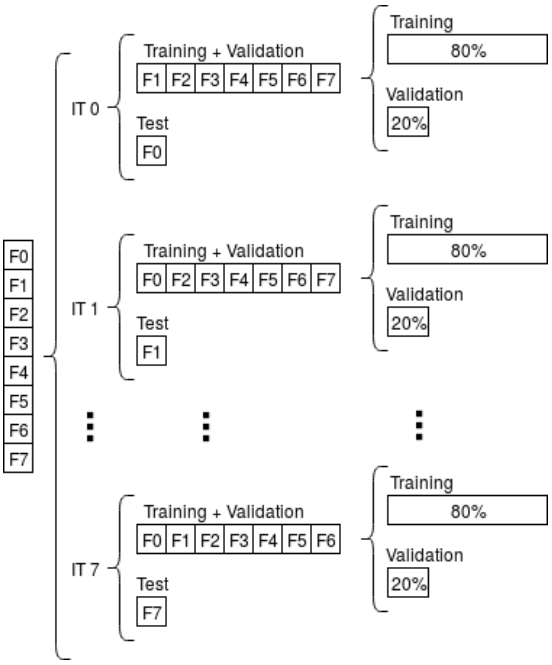


Figure 4.4: Cross validation approach defined to assess the performance of the proposed architectures. First, the dataset is split into eight folds (from $F0$ to $F7$). Then, for each CV iteration (from $IT0$ to $IT7$), a fold is omitted for testing of the model and the rest of the folds as a whole are divided into training (80%) and validation (20%) sets.

Every batch was augmented to increase the variability of the dataset and, therefore, improve the generalisation capabilities of the network. As shown in Table 4.2, 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 4.2.

Table 4.2: 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

The loss function varied depending on the output of the network. The root mean squared error (RMSE) loss function $\mathcal{L}_{DANet}$ was used in DANet, as shown in Equation (4.1), 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 $\mathcal{L}_{DASNet}$ (Equation (4.2)), the RMSE loss was used for the age regression and the binary cross entropy (BCE) 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 female 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 $\mathcal{L}_{DASNet}$.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$\mathcal{L}_{DANet}(\theta) = \sqrt{\frac{\sum_{i=1}^{n}(a_i - \hat{a}_i)^2}{n}} \tag{4.1}$$

$$\mathcal{L}_{DASNet}(\theta) = w \cdot \sqrt{\frac{\sum_{i=1}^{n}(a_i - \hat{a}_i)^2}{n}} - \frac{1}{n}\sum_{i=1}^{n}[s_i \cdot \log(\hat{s}_i) + (1 - s_i) \cdot \log(1 - \hat{s}_i)] \quad (4.2)$$

A set of regression metrics were computed to assess the performance of both DANet and DASNet. The median of 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 [9] and [150], jointly with some exploratory analysis, as further described in Appendix A (Section 4.6). First, the asymmetry observed in the real data was preserved in the predictions made by the two networks (according to the MGG symmetry test [150]). 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 DASNet was learning the sex-specific features effectively. As the output of the sex path was a continuous value 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, and it was calculated through Equation (4.3), where T_m and T_f are the images correctly classified as male and female, respectively, and F_m and F_f are the images incorrectly classified, respectively. The per-class F1 measure is calculated as the harmonic mean of the recall —the ratio of correctly classified images of that class— and precision —the proportion of images

classified as a specific class that belongs indeed to that class— metrics, and it was calculated through Equations (4.4) and (4.5) for the female and male class, respectively.

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. DANet and DASNet were trained and tested on each dataset, meaning that the age prediction performance could be assessed at different levels.

To compare the age estimation performance of 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 (error) means and medians, and their respective dispersion measures, and AE (absolute error) values were determined.

Finally, the behaviour of the network was analysed to determine which regions of the image were focused on for chronological age estimation, and this was then checked to see if it matched the opinions of the dental experts in the clinical field. The gradient-weighted class activation mapping (Grad-CAM) method [197] 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.

$$Accuracy = \frac{T_m + T_f}{T_m + T_f + F_m + F_f} \tag{4.3}$$

$$F1_f = \frac{2 \cdot R_f \cdot P_f}{R_f + P_f} \tag{4.4}$$

$$F1_m = \frac{2 \cdot R_m \cdot P_m}{R_m + P_m} \tag{4.5}$$

$$R_f = \frac{T_f}{T_f + F_m} \tag{4.6}$$

$$P_f = \frac{T_f}{T_f + F_f} \tag{4.7}$$

$$R_m = \frac{T_m}{T_m + F_f} \tag{4.8}$$

$$P_f = \frac{T_m}{T_m + F_m} \tag{4.9}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.4 Results

As shown in Tables 4.3 and 4.4, there was generally a high correlation between the ages estimated by DANet and DASNet and the real ages (R^2 = 0.87 and 0.90, respectively). DASNet provided better results than DANet in the S_{all} dataset, as the mean AE was about 2 years and 10 months, giving a 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 DASNet than DANet in all cases. In the $S_{<15}$ dataset, the mean AE was 0.78 years for DANet and 0.75 years for 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 DANet and 0.57, 0.72, and 2.58 in DASNet. As better performance results were obtained from DASNet, the following analyses focus on this model.

In Figure 4.5, 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 Figure 4.6). This started to worsen in older subjects, with this remarkable result displayed in Figure 4.6c. In this final plot, the network did not seem to have enough information to estimate the age.

The sex classification results of DASNet are also presented in Table 4.4. The top accuracy (Acc) of 85.4% was achieved in the S_{all} dataset, with $F1$ values above 0.83, while the top error was in the $S_{<15}$ dataset, where the accuracy remained below 80%. The higher female F1 ($F1_f$) with respect to the male F1 ($F1_m$) in almost all cases, with differences between 0.003 in the $S_{<20}$ dataset and 0.033 in the S_{all} dataset, shows that DASNet yields slightly better performance in female images.

As highlighted in Figure 4.7, 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]

USC UNIVERSIDADE

produced a mean AE of 1.28 ± 1.18 years, with these values rising to 5.87 ± 4.70 years in the age range [30,40].

Table 4.3: Performance of DANet (age path). E: error; AE: absolute error; med.: median; IQR: interquartile range; p99: 99th percentile; R^2: coefficient of determination. The error metrics are given in years.

		Age regression					
		E	**AE**				R^2
Dataset	**#Img**	**med.**	$\mu \pm \sigma$	**med.**	**IQR**	**p99**	
S_{all}	2,289	0.44	3.19 ± 4.32	1.66	3.01	21.09	0.87
$S_{<40}$	2,030	0.15	1.97 ± 2.26	1.19	1.97	10.80	0.85
$S_{<30}$	1,885	0.08	1.52 ± 1.54	1.04	1.57	6.85	0.88
$S_{<25}$	1,718	0.09	1.21 ± 1.17	0.87	1.20	5.53	0.89
$S_{<20}$	1,381	-0.04	0.93 ± 0.80	0.72	1.02	3.78	0.88
$S_{<15}$	1,087	-0.07	0.78 ± 0.65	0.64	0.81	2.91	0.82

Table 4.4: Performance of DASNet (age and sex paths). Age regression abbreviations: E: error; med.: median; AE: absolute error; μ: mean; σ: standard deviation; med.: median; IQR: interquartile range; p99: 99th percentile; R^2: coefficient of determination. Sex classification abbreviations: Acc.: accuracy; $F1_f$: F1 measurement for the female class; $F1_m$: F1 measurement for the male class. The error metrics are given in years.

		Age regression						**Sex classification**		
		E	**AE**				R^2	**Acc.**	$F1_f$	$F1_m$
Dataset	**#Img**	**med.**	$\mu \pm \sigma$	**med.**	**IQR**	**p99**				
S_{all}	2,289	0.12	2.84 ± 3.75	1.48	2.70	18.16	0.90	0.854	0.868	0.835
$S_{<40}$	2,030	0.04	1.80 ± 2.12	1.10	1.75	10.48	0.87	0.848	0.860	0.835
$S_{<30}$	1,885	0.06	1.43 ± 1.44	0.96	1.50	6.51	0.89	0.826	0.850	0.819
$S_{<25}$	1,718	0.07	1.17 ± 1.11	0.85	1.18	5.34	0.90	0.825	0.836	0.813
$S_{<20}$	1,381	0.02	0.89 ± 0.77	0.69	0.95	3.45	0.89	0.800	0.802	0.799
$S_{<15}$	1,087	-0.07	0.75 ± 0.57	0.64	0.72	2.58	0.84	0.750	0.741	0.754

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

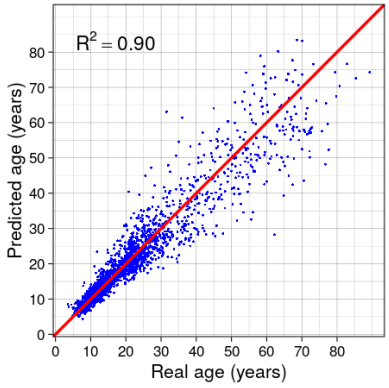
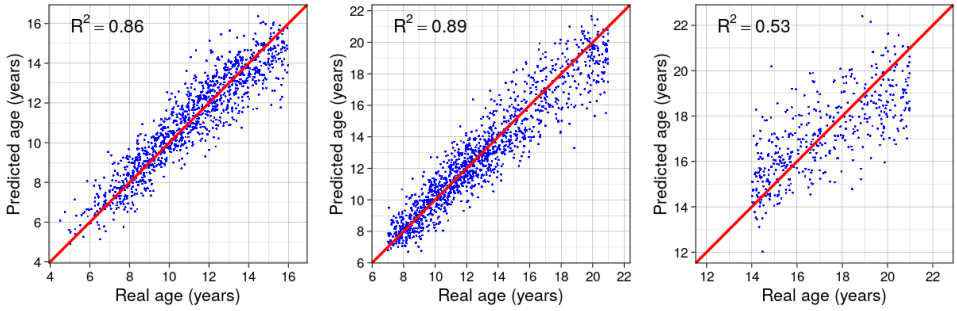


Figure 4.5: Real age *vs.* the DASNet predictions evaluated for the entire dataset, S_{all}. The closer the points are to the red line, the better the performance.



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

Figure 4.6: Real age *vs.* 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.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

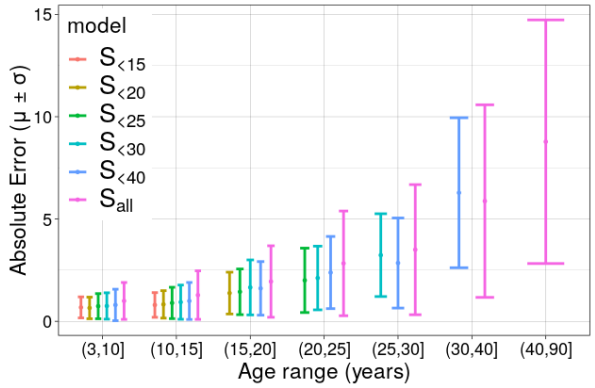


Figure 4.7: DASNet absolute error in the chronological age estimation when training with reduced datasets and evaluating performance for different age ranges.

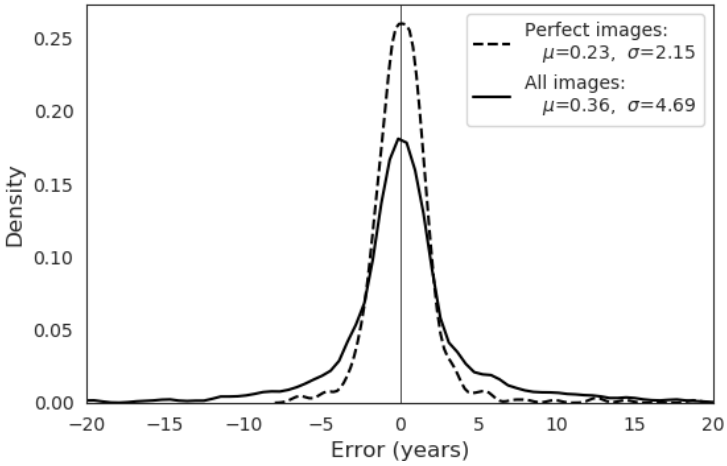


Figure 4.8: Distribution of the DASNet errors (true age - predicted age), shown as the density estimation, in all the images *vs.* the *perfect* images (those which do not present any conditioning radiological or dental characteristic, as explained in section 4.2).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Concerning the DASNet's performance on *perfect* OPGs with respect to the entire dataset (Figure 4.8), 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 ± 2.15 years, in contrast to 0.36 ± 4.69 years).

Tables 4.5 and 4.6 detail the comparisons of the manual dental age estimation methods with 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 metric was the E mean and the corresponding standard deviation. As shown in this table, and considering at least one of the metrics analysed, the DASNet approach produced better results than the manual methods in all 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 [82]; the mean AE of DASNet = 0.72 vs 1.08 and 0.93 for the Demirjian and Willems methods, respectively [29]). 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 DASNet *vs.* the manual methods = -0.01, 0.12 *vs.* -0.35, 0.88) [23, 51, 82, 180]. These observations were corroborated in the comparisons in which E medians were used [24, 29, 82]. Focusing on the dispersion measurements, and specifically the standard deviations, six of the nine comparisons revealed that 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 Figure 4.9) 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.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 4.5: Comparison of manual methods for dental age estimation (white cells) with the DASNet approach (grey cells). All the presented methods rely on manual measurements or classifications every time the age of a subject needed to be estimated.

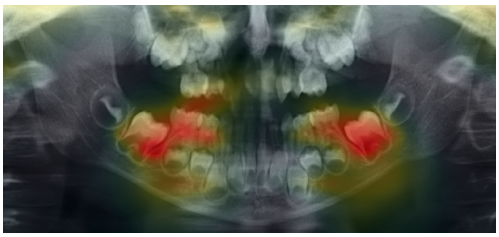
Work	Method	Age	Population	#Images	
Prieto [180]	Demirjian	14-21	Spanish	1,054	446
Melo [148]	Demirjian Nolla	7-21	Spanish	2,641	1,381
Cruz-Landeira [51]	Demirjian Chaillet	4-17	Spanish	308	1,223
Cameriere [24]	Cameriere	4-16	European Caucasian	2,652	1,160
Butti [23]	Häävikko	3.9-15.4	Italian	500	1,119
Galić [82]	Cameriere Häävikko Willems	6-13	Italian	1,089	835
Cameriere [29]	Cameriere Demirjian Willems	5-15	Italian, Spanish & Croatian	756	1,082

Table 4.6: Error metrics of the methods described in Table 4.5 with respect to the DASNet approach. E: error (real - predicted); IQR: interquartile range; AE: absolute error. The metrics are given in years.

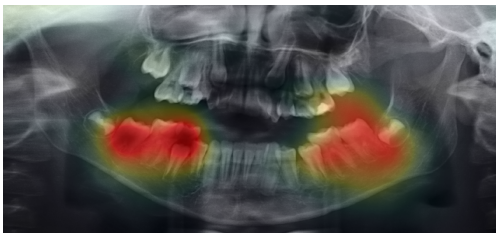
Work	E(μ)		E(σ)		E(median)		E(IQR)		AE(μ)	
[180]	-0.1 / -0.07 [(a)]	0.04	1.23 / 1.22 [(a)]	1.49	-	0.02	-	1.93	-	1.19
[148]	0.85 -0.21	0.12	0.39 / 0.56 [(b)] 0.50 / 0.23 [(b)]	1.20	- -	-0.07	- -	1.38	- -	0.91
[51]	0.76 / 0.88 [(b)] 0.37 / 0.21 [(b)]	0.10	1.01 / 1.09 [(b)] 1.04 / 1.07 [(b)]	1.00	- -	-0.07	- -	1.25	- -	0.78
[24]	-	-0.02	-	0.97	-0.11	-0.06	1.22	1.21	-	0.75
[23]	-0.35	0.06	0.72	0.94	-	0.008	-	1.21	-	0.73
[82]	-0.02 / 0.1 [(b)] -0.09 / -0.23 [(b)] 0.42 / 0.25 [(b)]	-0.01	0.71 / 0.71 [(b)] 0.79 / 0.73 [(b)] 0.77 / 0.89 [(b)]	0.80	-0.04 / 0.09 [(b)] -0.14 / -0.23 [(b)] 0.34 / 0.15 [(b)]	-0.04	- - -	1.05	0.55 / 0.53 [(b)] 0.62 / 0.59 [(b)] 0.67 / 0.69 [(b)]	0.63
[29]	- - -	0.02	- - -	0.92	0.06 -0.70 -0.07	-0.02	0.71 1.50 1.59	1.18	0.49 1.08 0.93	0.72

(a) Depending on the specific tooth used for age estimation (left third molar / right third molar)
(b) Male / female

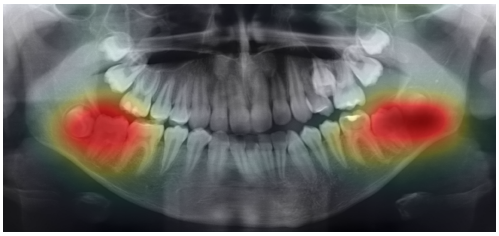
USC DE SANTIAGO DE COMPOSTELA



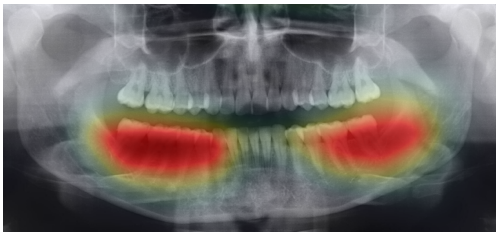
(a) 4 years and 8 months old



(b) 9 years and 8 months old



(c) 13 years and 2 months old



(d) 25 years and 2 months old

Figure 4.9: Example of Grad-CAM attention maps at different ages. The redder the region, the more attention the network pays to it.

4.5 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 [19, 235]. 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 deep learning techniques is of great interest in the clinical dental field, as it could bring about substantial improvements in terms of time and subjectivity.

This chapter presented two novel deep learning approaches to estimating the age of a subject from an OPG. 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

USC UNIVERSIDADE

estimation process. To the best of our knowledge, this work is the first to use deep learning techniques for the automatic calculation of chronological age from OPGs.

In all the examined circumstances, DASNet outperformed 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. This fact confirms that for those older than 25, which is when the permanent teeth are completely formed, it is difficult to assess age accurately [180]. 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 DASNet, the sex classification task was evaluated using the same datasets. Although the $F1$ metric was higher for the female class, the slight differences concerning the male $F1$ suggest that the sex estimation was 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 [4, 80], and improve those obtained in [10, 173].

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 sex of a subject using an automated procedure. These findings are consistent with those obtained by other authors, who used different sex estimation methods applying odontological measures [31].

Compared to the state of the art [23, 24, 29, 51, 82, 148, 180], and by taking the mean and median errors closer to zero, the DASNet approach reduces the problem of under- or over-

USC UNIVERSIDADE

estimation, which is fairly common in the classical manual methods [23, 51, 82, 148, 180]. Furthermore, the mean absolute error $AE(\mu)$ was lower than that obtained with the Demirjian and Willems methods in [29] and with the Willems approach in [82]. In contrast, the standard deviations of the errors $E(\sigma)$ of DASNet were greater than in several series that used different manual methods [23, 82, 148, 180]. 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 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 [27, 60]. In conclusion, the DASNet method can be used to automatically and accurately predict someone's chronological age, especially in young subjects with developing dentition.

4.6 Appendix A

In this appendix, further information about the statistical analysis performed to validate the agreement between the input and output data distribution and the use of the error measurements considered is described. First, the original data and predictions by DANet and DASNet were explored using graphical tools. Boxplots and histograms showed a rough summary of the data, whereas significant zero-crossing of the derivative (SiZer) maps allowed identifying features such as asymmetry or multimodality. In addition, empirical quantile-quantile plots were also produced to get information on the equality of the distributions. Formal tests of normality, equality of distributions, asymmetry, and multimodality were also performed.

The distributions of real age and predictions from DANet and DASNet were compared both qualitatively and quantitatively in the entire dataset and the reduced datasets. According to Figs. 4.10 and 4.11, real data and predictions had a similar distribution. None of the distributions were normal (p-values lower than the usual significance levels both in the Kolmogorov-Smirnov [134] and the Shapiro-Wilk [199] normality tests). Furthermore, all distributions were asymmetric, since the p-values were lower than the usual significance levels in the Miao symmetry test [150] (Table 4.7).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

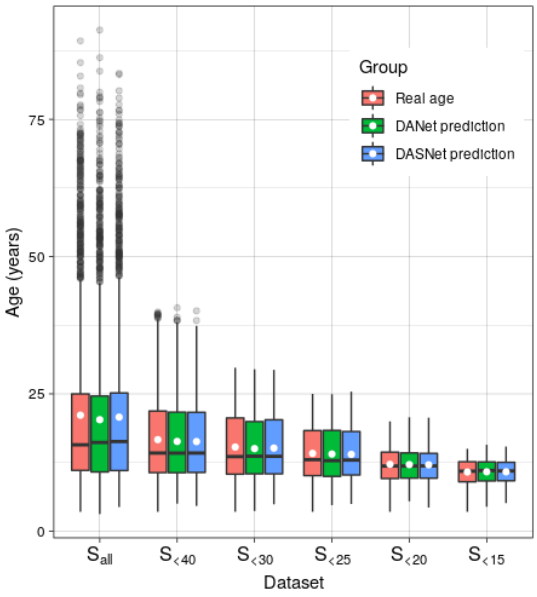


Figure 4.10: Graphical comparison of the predictions of DANet and DASNet against the real age values, for the entire dataset and the reduced datasets. The white points correspond to the means of the distribution.

Table 4.7: Results of the symmetry test proposed in [150] (given as p-values obtained with a 5% significance level).

Dataset	**Real**	**DANet**	**DASNet**
S_{all}	0.014	< 0.0001	< 0.0001
$S_{<40}$	0.016	< 0.0001	< 0.0001
$S_{<30}$	0.030	< 0.0001	< 0.0001
$S_{<25}$	0.016	< 0.0001	< 0.0001
$S_{<20}$	0.012	< 0.0001	0.002
$S_{<15}$	0.018	< 0.0001	0.008

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

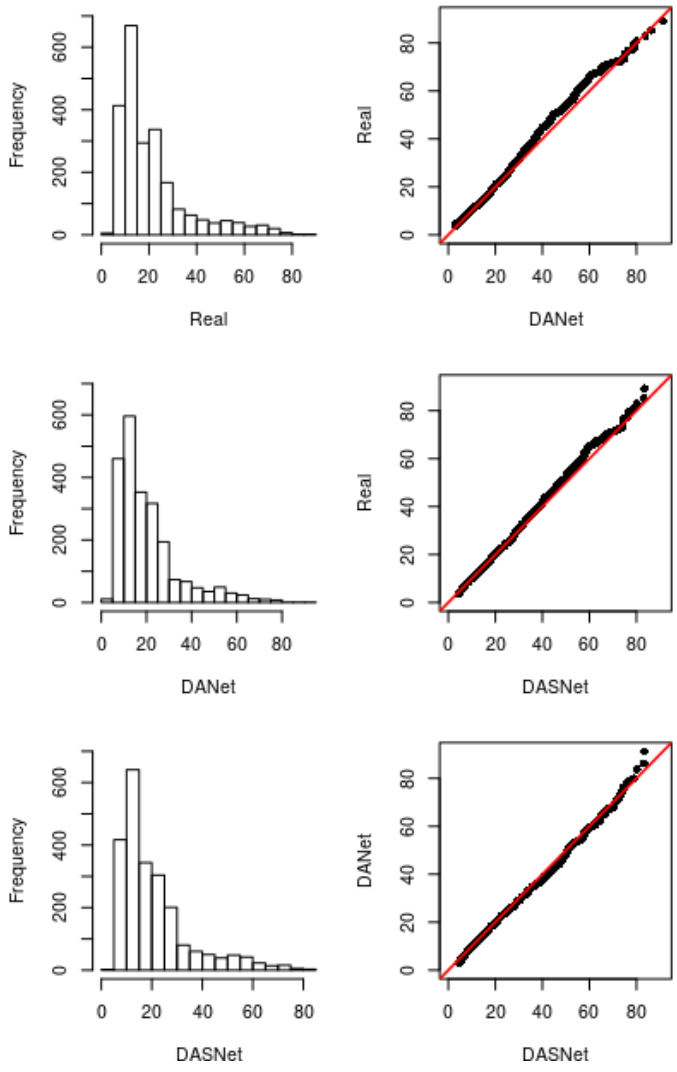


Figure 4.11: Histograms of real age and networks predictions, and quantile-quantile plots comparing each pair of distributions.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

According to the non-parametric Wilcoxon test [228], there was no evidence that DANet and DASNet predictions were different from real age (see p-values in Table 4.8). A permutation test was also applied to compare their medians, not obtaining significant differences (Table 4.9).

Table 4.8: Comparison of DANet and DASNet predictions *vs.* real age values. The results correspond to the p-values given by the Wilcoxon test [228] with a 5% significance level.

Dataset	**Real *vs.* DANet**	**Real *vs.* DASNet**
S_{all}	0.1968	0.8648
$S_{<40}$	0.4988	0.7404
$S_{<30}$	0.3753	0.6973
$S_{<25}$	0.4717	0.4726
$S_{<20}$	0.9022	0.8450
$S_{<15}$	0.7191	0.8500

Table 4.9: Comparison of the median of DANet and DASNet predictions and the median of the real age values. The results correspond to the p-values estimated from a permutation test (N=3000).

Dataset	**Real *vs.* DANet**	**Real *vs.* DASNet**
S_{all}	0.767	0.833
$S_{<40}$	0.477	0.500
$S_{<30}$	0.639	0.587
$S_{<25}$	0.251	0.365
$S_{<20}$	0.485	0.556
$S_{<15}$	0.798	0.604

To assess whether the distributions are unimodal or multimodal, the multimodality test proposed in [9] was applied. The results, presented in Table 4.10, show that in some cases, mostly in $S_{<15}$ and $S_{<20}$, the sample presented a single modal value. To explore the asymmetric and (possibly) multimodal pattern together, a SiZer map was produced for the real data and the network predictions. The SiZer map is an exploratory tool introduced in [39] to identify which increasing/decreasing patterns observed on a density estimate are underlying distribution features and which ones are just sampling artefacts. Based on a kernel density estimator, the SiZer map considers a range of smoothing values (bandwidths) and for each pair variable value - bandwidth, it checks if the estimator of the density derivative is positive, negative, or not

significantly different from zero. Areas coloured blue indicate that the function is significantly increasing (positive derivative), areas in red identify decreasing patterns (negative derivative) whereas pink areas correspond to a derivative value that is not significantly different from zero. When there is not enough data to obtain an estimator of the density derivative for a specific value with a certain bandwidth, pixels are coloured grey.

From the plots in Figure 4.12, it can be seen that there was usually a principal mode in all cases, although a spurious mode can be observed for real data and predictions in $S_{<40}$, $S_{<30}$, and $S_{<25}$, which is in line with the localisation of modes in Table 4.10. Nevertheless, from a deeper insight into the data, these spurious modes can be attributed to the unbalanced sample design, not being of primary interest for our study. The median values were also added to the plots in Figure 4.12, showing that these values were usually higher than the principal modal value (but never higher than the spurious mode) due to the asymmetry of the data distribution.

From the previous analysis, we can confirm that: 1) the distributions of the predictions of both DANet and DASNet were not significantly different from the real data, and neither of them present normality nor symmetry patterns; 2) in some cases, the distributions presented spurious modes which have to do with the sample design; 3) the median value was always close to the most significant mode, so the median error is a suitable metric to assess the performance of both networks.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 4.10: Multimodality analysis. **p**: p-value obtained with the test proposed in [9], where the null hypothesis states that the sample has one mode and the alternative hypothesis is if it has more than one modal value; **m. loc.** and **m. dens.**: mode location and density; **median**: the median of the distribution. In the case of multimodal distributions, the two most significant modes are given (those with higher density).

Dataset	**Real**			
	p	**m. loc.**	**m. dens.**	**median**
S_{all}	< 0.0001	11.64 50.49	0.05 < 0.01	15.71
$S_{<40}$	< 0.0001	11.23 18.42	0.07 0.03	14.22
$S_{<30}$	< 0.0001	11.60 18.35	0.08 0.03	13.57
$S_{<25}$	< 0.0001	11.77 18.67	0.09 0.03	13.00
$S_{<20}$	0.264	11.77	0.11	11.85
$S_{<15}$	0.282	11.78	0.14	10.91

Table 4.10: *(Continuation).*

Dataset	**DANet**				**DASNet**			
	p	**m. loc.**	**m. dens.**	**median**	**p**	**m. loc.**	**m. dens.**	**median**
S_{all}	0.197	10.93	0.05	16.14	0.184	11.20	0.05	16.31
$S_{<40}$	0.024	11.38 16.77	0.07 0.03	14.20	0.001	10.97 18.20	0.07 0.03	14.22
$S_{<30}$	< 0.0001	11.23 17.64	0.09 0.03	13.65	< 0.0001	11.16 18.00	0.08 0.03	13.62
$S_{<25}$	< 0.0001	9.93 16.44	0.09 0.03	12.83	< 0.0001	11.37 17.20	0.09 0.04	12.94
$S_{<20}$	0.759	11.83	0.12	11.85	0.174	12.02	0.12	11.88
$S_{<15}$	0.833	12.28	0.17	11.03	0.763	11.91	0.16	10.96

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

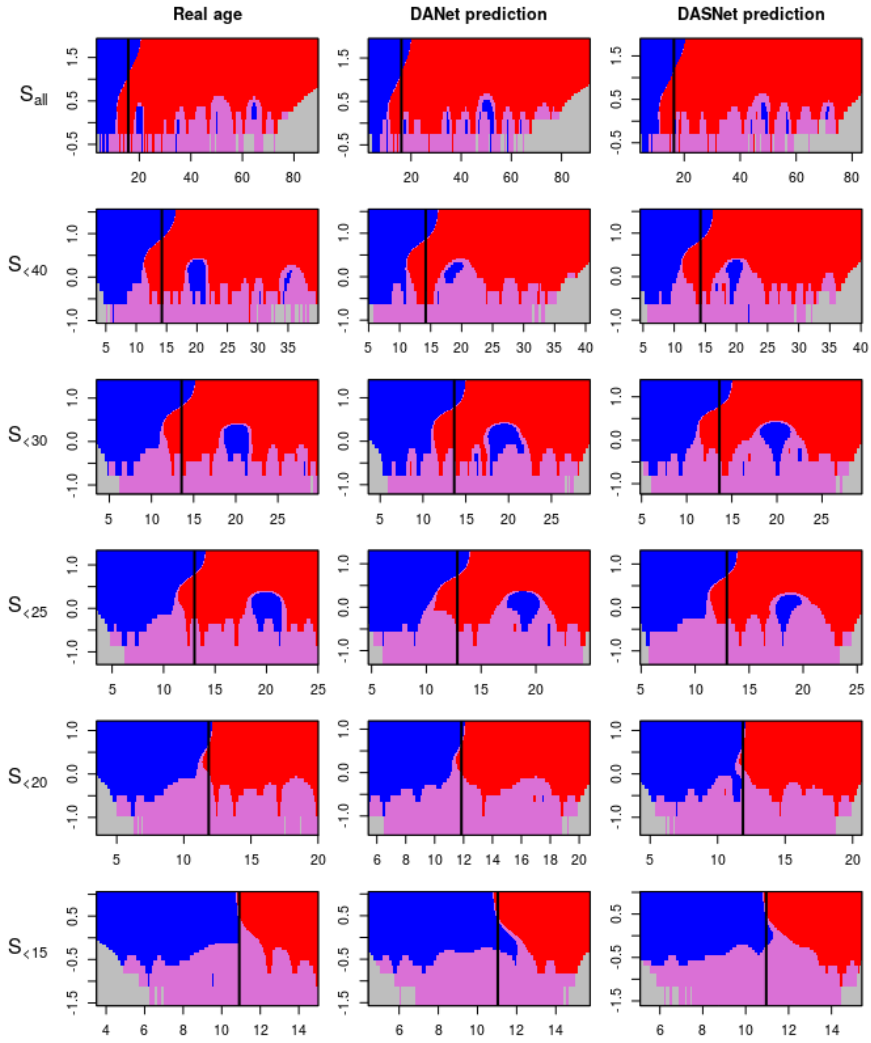


Figure 4.12: SiZer plots representing the increasing/decreasing pattern of a density estimation evaluated with different smoothing values (bandwidth). In the x-axis, the age values; in the y-axis, the bandwidth values in logarithmic scale. Each colour indicates a different pattern: blue, when the density is increasing; red, when the density is decreasing; pink, when the density is not changing significantly; and grey, when there is not enough data to obtain a density estimation for a specific value and a specific bandwidth. The vertical line corresponds to the median value.

USC DE SANTIAGO DE COMPOSTELA

4.7 Appendix B

Sexual dimorphism has been widely assessed through the analysis of some structures observable in an OPG. As it can be seen in Table 4.11, the mandible has been the most used bone due to its higher dimorphism [101], all methods proposed in this regard are based on linear distances and angles between a set of anatomic landmarks. The teeth have been studied to a lesser extent, with the study reported by Capitaneanu *et al.* [30] being particularly relevant. Moreover, other bones like the mandibular sinuses [182] or mental foramen [38] have proven to be dimorphic, although no sex determination methods have been developed in this respect. Recently, other approaches leveraged the benefits of deep learning techniques to automatically determine the sex from panoramic images [151, 222]. All these findings suggest that it is feasible to carry out sex classification in OPGs belonging to adults accurately, either through manual or automatic methods. However, the development of a sex classification method that reaches good performance in both subadult and adult subjects is still an open problem.

Following the same principle applied in this chapter to carry out chronological age estimation, a new architecture is proposed in this appendix, which is deeply based on the mentioned DASNet. Given that the main objective of DASNet was to integrate sexual features to improve chronological age estimation, a new version with reversed roles was evaluated —referred to as Dental Sex and Age Network (DSANet)—. In this case, the main CNN path corresponds to the sex classifier, and the auxiliary path is designed to estimate the chronological age and thus learn maturational features, which are propagated to the sex path to improve the sex classification.

To validate this sex prediction approach, an extended version of the OPG dataset was compiled. Specifically, 1,111 new images were collected, for a total of 3,400 images. The demographic characteristics of the dataset are presented in Table 4.12.

The DSANet architecture was trained in the same way as DASNet, as explained in Section 4.3. This included the use of Adadelta optimiser, the same loss function, a batch size of 4, data augmentation techniques for increasing the data variability, and the 8-fold CV. The performance of the network was assessed through four different metrics, already defined in Section 4.3, comprising the overall accuracy, and the per-class recall, precision, and F1. To calculate these measurements, we set the classification threshold to 0.5. DSANet was compared to the DASNet approach and the so-called VGG16 architecture [204], which has been widely used to automatically process OPGs with successful results [100, 114, 151].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 4.11: Performance of the main studies on sex prediction from OPGs. (M: Males; F: Females)

Reference	Sample	Age range	Method	Accuracy
[77]	40 (20M/20F)	20-48	Mandible measurements	95%
[4]	53 (31M/22F)	19-28	Tooth crown measurements	64-83%[(a)]
[153]	200 (100M/100F)	18-25	MCI	76%
[143]	200 (100M/100F)	20-86	Mandible measurements	84%
[12]	419 (126M/293F)	13-26	Mandibular flexure scoring[(b)]	70.9%
[205]	100 (45M/55F)	20-30	MCI	85.5%
[201]	120 (50M/70F)	16-30	MCI	64.2%
[30]	200 (100M/100F)	22-34	Tooth measurements	73 %
[5]	79 (48M/31F)	18-74	Mandible measurements	78.5%
[151]	4000 (2352M/1648F)	19-85	CNN	96.7%
[222]	2289 (1030M/1257F)	4.5-89.2	CNN	85.4%

(a) Bucolingual/mesiodistal measurements.

(b) Method from [140].

Table 4.12: Age and sex distribution of the OPG dataset, after discarding images labelled with the wrong acquisition or birthdates.

Age (years)	Male	Female	Total
[5-10)	256	254	510
[10-20)	595	606	1,201
[20-30)	232	438	670
[30-40)	118	183	301
[40-50)	109	135	244
[50-60)	93	139	232
[60-70)	77	77	154
[70-90)	35	53	88
Total	1,515	1,855	3,400

Table 4.13 shows the prediction results independently for each network and age group, following the same division as in Table 4.12 for people older than 20 and a finer division for younger people. With VGG16, the sex estimation accuracy exceeded 90% in people aged from 16 to 20 and those aged from 30 and 60, with a significant peak of almost 95% in the 18-20 age range. In the rest of the groups older than 12, the accuracy stayed above 80%, while it decreased to 74% in the youngest group. The images of females were classified better in

UEC UNIVERSIDADE

almost every age group, being especially noticeable in the group 60-70, with a recall difference of 13% with respect to the male class, while the most balanced predictions were made in the group 5-8 (recall difference of 0.12%). The most certain predictions, however, were yielded for the male class in the group 18-20, with a precision value of over 98%.

With DASNet, the classification accuracy went beyond 90% in every age group between 16 and 60 years of age, and the older age group. In younger people, the performance decreased up to 70% in the range 5-8, and the best overall performance was reached in the group 30-40, with an accuracy value of 96.68% and the highest F1 values (97.28% for the female class and 95.73% for the male class). The female images were classified better in every evaluated age group, where the most noticeable differences were seen in the older age group, with a recall difference of 8.52%. The precision values indicate that the certainty of the predictions was more balanced across both classes, although big differences were also produced as in the case of the group 5-8 (male precision was 9.57% higher than female precision).

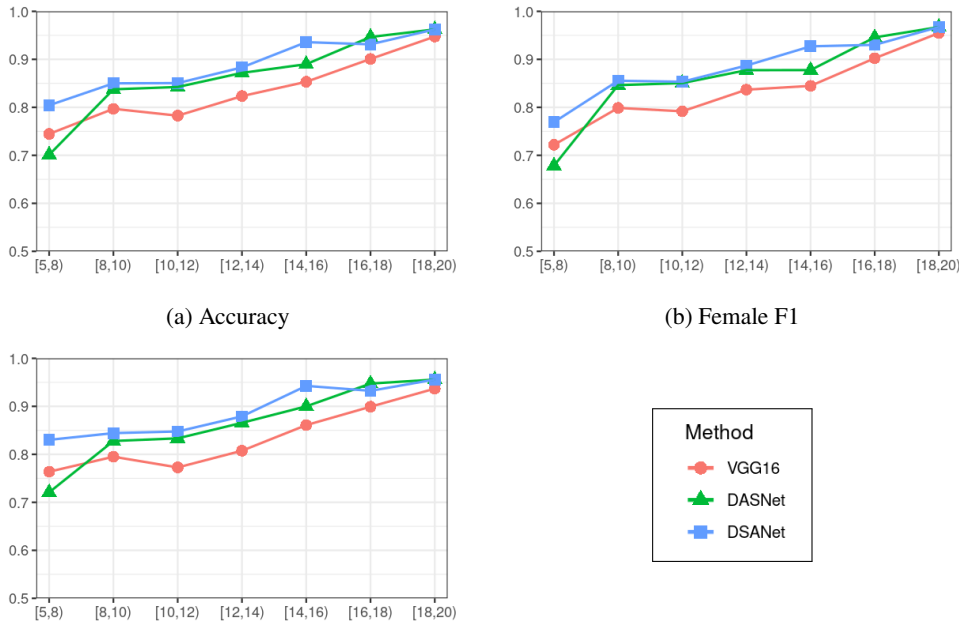
The results of DSANet showed an accuracy of over 85% in subjects older than eight and over 93% when going beyond 14, where the top accuracy of 96.24% was obtained in the group of subjects aged from 18 to 20. According to the recall metric, the classification of male images tended to be more accurate than those of females in groups younger than 12 and less accurate in the rest of the groups, being the greatest difference that obtained in the younger age group (13.1% in favour of male class). The precision metric did not follow the same trend, as changing differences between both classes were seen all across the evaluated age groups.

When comparing the three networks side by side, all metrics were highly correlated. Focusing on the age groups of people older than 20 —where it is expected the maximum accuracy—, DASNet and DSANet showed a very similar performance, clearly outperforming VGG16. The maximum accuracy difference between both of them was produced in the group 20-30, being 2.54% in favour of DSANet. In the remaining groups, that accuracy difference stayed below 2%. However, it is noticeable that the performance of DSANet was more balanced across classes, even in the age groups where DASNet produced a better accuracy value. Specifically, the male-female recall differences were reduced from 12.28% to 5.82% in the group 40-50, and from 8.52% to 5.66% in the group 70-90.

Although the accuracy of the three methods was generally lower in the people younger than 20, the performance metrics followed an improvement pattern along all that period with every tested architecture. As shown in Table 4.13 and Figure 4.13, there was a jump at about eight years of age which was especially noticeable in DASNet, where the accuracy improved

USC UNIVERSIDADE

from 70 to 84%. The better performance balance between female and male recall was obtained by DSANet in the group 12-14, being the best F1 metric obtained by DASNet and DSANet in both male and female classes of the 18-20 age group.



(a) Accuracy

(b) Female F1

(c) Male F1

Figure 4.13: Evolution of the classification metrics in subjects from seven age groups, ranging from five up to 20 years old.

Overall, DSANet clearly provided the best performance in most cases, as it showed a more homogeneous behaviour and the performance drop of the youngest and oldest age groups was not as pronounced as that produced with the other architectures. On the contrary, VGG16 tended to obtain worse results in general terms. To the best of our knowledge, this has to do with the combination of maturational and sexual features on both DASNet and DSANet architectures, which allows these architectures to learn in a more structured way. Compared to the main methods for sex determination from OPGs, the DASNet and DSANet approaches reach similar performance than the most accurate method —proposed by Milošević *et al.* [151]— in the age range between 18 and 40 years of age, and slightly worse accuracy in the

rest of adult groups. However, it is worth noting that the proposed methodology is aimed at working on a wider age range, i.e., 5-90 *vs.* 19-85.

These findings support the fact that it is quite challenging to determine the sex in people younger than 14 by looking only at the oral cavity. In the youngest age groups, the images show great variability, motivated by the presence of mixed dentition stages and the heterogeneity in mandibular growth patterns [71, 154]. Moreover, secondary sexual features are not expected to develop until puberty [187], and so sex differentiation is very weak. Thus, the networks can not go beyond 90% of accuracy.

In conclusion, the three evaluated networks provide reliable predictions of sex, with DSANet being the most accurate one in the majority of the age groups. The suitability of this approach is especially relevant in patients older than 14 years old, reporting accuracies between 90 and 96.2%. Although the performance decreases in younger people, the method is still useful in subjects older than eight when combined with other radiological methods, with accuracies over 85%, demonstrating the usefulness of automatic approaches in the prediction of sex.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 4.13: Sex classification results by age group. The highlighted values correspond to the most accurate networks according to each specific metric.

Age groups	Method	Accuracy	Female			Male		
			Recall	Precision	F1	Recall	Precision	F1
[5,8)	VGG16	0.7446	**0.7439**	0.7011	0.7219	0.7451	0.7835	0.7638
	DASNet	0.7011	0.7073	0.6517	0.6784	0.6961	0.7474	0.7208
	DSANet	**0.8043**	0.7317	**0.8108**	**0.7692**	**0.8627**	**0.8000**	**0.8302**
[8,10)	VGG16	0.7969	0.7588	0.8431	0.7988	0.8400	0.7545	0.7950
	DASNet	0.8375	**0.8412**	0.8512	0.8462	0.8333	0.8224	0.8278
	DSANet	**0.8500**	0.8353	**0.8765**	**0.8554**	**0.8667**	**0.8228**	**0.8442**
[10,12)	VGG16	0.7826	0.8000	0.7835	0.7917	0.7640	0.7816	0.7727
	DASNet	0.8424	**0.8684**	0.8333	0.8505	0.8146	**0.8529**	0.8333
	DSANet	**0.8505**	0.8421	**0.8649**	**0.8533**	**0.8596**	0.8361	**0.8476**
[12,14)	VGG16	0.8234	0.8736	0.8030	0.8368	0.7692	0.8497	0.8075
	DASNet	0.8718	**0.8846**	0.8703	0.8774	0.8580	0.8735	0.8657
	DSANet	**0.8832**	**0.8846**	**0.8895**	**0.8871**	**0.8817**	**0.8765**	**0.8791**
[14,16)	VGG16	0.8532	0.9158	0.7838	0.8447	0.8049	0.9252	0.8609
	DASNet	0.8899	0.9053	0.8515	0.8776	0.8780	0.9231	0.9000
	DSANet	**0.9358**	**0.9368**	**0.9175**	**0.9271**	**0.9350**	**0.9504**	**0.9426**
[16,18)	VGG16	0.9008	0.9375	0.8696	0.9023	0.8657	0.9355	0.8992
	DASNet	**0.9466**	**0.9531**	**0.9385**	**0.9457**	**0.9403**	**0.9545**	**0.9474**
	DSANet	0.9313	0.9375	0.9231	0.9302	0.9254	0.9394	0.9323
[18,20)	VGG16	0.9474	**0.9867**	0.9250	0.9548	0.8966	0.9811	0.9369
	DASNet	**0.9624**	**0.9867**	**0.9487**	**0.9673**	**0.9310**	**0.9818**	**0.9558**
	DSANet	**0.9624**	**0.9867**	**0.9487**	**0.9673**	**0.9310**	**0.9818**	**0.9558**
[20,30)	VGG16	0.8896	0.9018	0.9272	0.9144	0.8664	0.8238	0.8445
	DASNet	0.9179	0.9338	0.9402	0.9370	0.8879	0.8766	0.8822
	DSANet	**0.9433**	**0.9543**	**0.9587**	**0.9565**	**0.9224**	**0.9145**	**0.9185**
[30,40)	VGG16	0.9203	0.9399	0.9297	0.9348	0.8898	0.9052	0.8974
	DASNet	**0.9668**	**0.9781**	**0.9676**	**0.9728**	**0.9492**	**0.9655**	**0.9573**
	DSANet	0.9568	0.9672	0.9620	0.9646	0.9407	0.9487	0.9447
[40,50)	VGG16	0.9098	0.9333	0.9065	0.9197	0.8807	0.9143	0.8972
	DASNet	**0.9303**	**0.9852**	0.8986	**0.9399**	0.8624	**0.9792**	**0.9171**
	DSANet	0.9221	0.9481	**0.9143**	**0.9309**	**0.8899**	0.9327	0.9108
[50,60)	VGG16	0.9138	0.9353	0.9220	0.9286	0.8817	0.9011	0.8913
	DASNet	**0.9440**	**0.9640**	0.9437	**0.9537**	0.9140	**0.9444**	**0.9290**
	DSANet	0.9353	0.9424	**0.9493**	0.9458	**0.9247**	0.9149	0.9198
[60,70)	VGG16	0.8701	0.9351	0.8276	0.8780	0.8052	0.9254	0.8611
	DASNet	0.8896	**0.9221**	0.8659	0.8931	0.8571	0.9167	0.8859
	DSANet	**0.9026**	**0.9221**	**0.8875**	**0.9045**	**0.8831**	**0.9189**	**0.9007**
[70,90)	VGG16	0.8851	0.9038	0.9038	0.9038	0.8571	0.8571	0.8571
	DASNet	0.9080	**0.9423**	0.9074	0.9245	0.8571	0.9091	0.8824
	DSANet	**0.9195**	**0.9423**	**0.9245**	**0.9333**	**0.8857**	**0.9118**	**0.8986**
[5,90)	VGG16	0.8594	0.8833	0.8658	0.8745	0.8297	0.8510	0.8402
	DASNet	0.8915	**0.9146**	0.8923	0.9033	0.8627	0.8903	0.8763
	DSANet	**0.9057**	0.9135	**0.9179**	**0.9157**	**0.8983**	**0.8930**	**0.8957**

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 5

CNN AND SHAPE MODELLING FOR AGE AND SEX PREDICTION FROM THE MANDIBLE

The shape of the mandible has been analysed in a variety of fields, whether to diagnose conditions like osteoporosis or osteomyelitis, in forensics, to estimate biological information such as age, gender and race, or in orthognathic surgery. Although the methods employed produce encouraging results, most rely on dry-bone analyses or complex imaging techniques that, ultimately, hamper sample collection and, as a consequence, the development of large-scale studies. Thus, in this chapter, we proposed an objective, repeatable, and fully automatic approach to provide a quantitative description of the mandible in OPGs. Our solution relies on a CNN to localize a set of landmarks of the mandible contour automatically from the images. Furthermore, we detailed four different descriptors for the mandible shape to be used for a variety of purposes. This includes a set of linear distances and angles calculated from eight anatomical landmarks of the mandible, the centroid size, the shape variations from the mean shape, and a group of shape parameters extracted with a point distribution model (PDM). The methodology proposed, including the shape model, can be valuable in any field that requires a quantitative description of the mandible shape and a visual representation of its changes, such as clinical practice, surgery management, dental research, or legal medicine.

The contents of this chapter are extracted from the following publication, the reproduction rights of which have been granted by the CC BY licence (Figure 5.1):

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Vila-Blanco, N.*,†, Varas-Quintana, P.‡, Aneiros-Ardao, Á.‡, Tomás, I.†,‡, and Carreira, M. J.*,† (2021). Automated characterization of the mandible shape by deep learning. International Journal of Computer Assisted Radiology and Surgery, 16(12), 2215-2224. DOI:10.1007/s11548-021-02474-2

Figure 5.1: Licencing information of the publication [223], from which the contents of Chapter 5 are taken from.

5.1 Introduction

The mandible is the strongest, largest, and only movable facial bone [99]. It enables speech and mastication and hosts the lower teeth. As a consequence, mandible disorders have a significant effect on both appearance and quality of life. Furthermore, examinations of the mandible's form can be employed in the diagnosis of several conditions [15, 61, 64, 102, 156, 168].

Dentistry, orthodontics, and forensics have probably been the fields where the mandible bone has been studied the most. Regarding the latter, many works reported a strong relationship between mandibular bone features, such as morphometry and appearance, and biological variables as sex or age. Gender dimorphism has been assessed through a set of distances between anatomical landmarks [135, 179, 208] or the analysis of the mandibular shape [76]. It is worth noting that the sex estimation models reported a high accuracy for adult subjects,

*Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) and Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.
†Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Spain.
‡Oral Sciences Research Group, Special Needs Unit Department of Surgery and Medical-Surgical Specialities School of Medicine and Dentistry, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.

as the gender dimorphism is higher than in subadults [78]. The mandible evolution with age has been studied to a lesser extent, with the opposite finding, that is, the mandible changes in older people are quite limited and mainly related to tooth loss [172]. In addition to sex and age, the population-specific patterns of mandible development have also been studied [79].

Regarding the collection of mandible information, the studies have traditionally relied on dry-bone measurements. However, recent decades have seen the increasing use of imaging techniques, such as 3D optical scanner [74, 76], or computed tomography [135]. One of the most used systems is panoramic dental imaging or OPG, but this procedure has several drawbacks. Given its rotational acquisition process, the image projection leads not only to an information loss but also to a potential deformation, which is especially noticeable in the horizontal direction [121]. However, it is still non-destructive, it captures the complete mandible in a single image, which is both faster and beneficial for the storage of data and the measuring process, and it is reportedly useful to measuring the mandible [144]. Indeed, the value of OPGs has been proved in a variety of dentistry tasks, including the diagnosis of several clinical conditions [50, 164], surgery management [7], or forensic procedures [2]. Although this imaging technique has been used for decades, the mandible detection methods based on automatic image processing algorithms are still very scarce [1, 159].

The current study presents a two-step pipeline to describe the shape of the mandible automatically on OPGs, to use this description for a variety of purposes. In the first step of the proposed pipeline, a CNN is applied to automatically extract the mandible contour. In a second step, four different descriptors are employed to characterise the mandibular shape, namely a set of linear distances and angles, the centroid size, the mandible variations with respect to the mean shape, and a set of parameters given by a shape model.

5.2 Materials and methods

The workflow employed in the present study is set out in Figure 5.2. First, the contour of the mandible, given by a set of anatomical landmarks and the intermediate points —also known as semilandmarks—, is obtained through an automatic landmark detection method based on a CNN. Second, four different descriptors are applied to the mandible contour, namely a set of 11 linear distances and angles, the centroid size, the shape variations with respect to the mean shape, and the shape parameters given by a point distribution model (PDM). Both steps are explained in detail in Sections 5.2.2 and 5.2.3, respectively.

USC

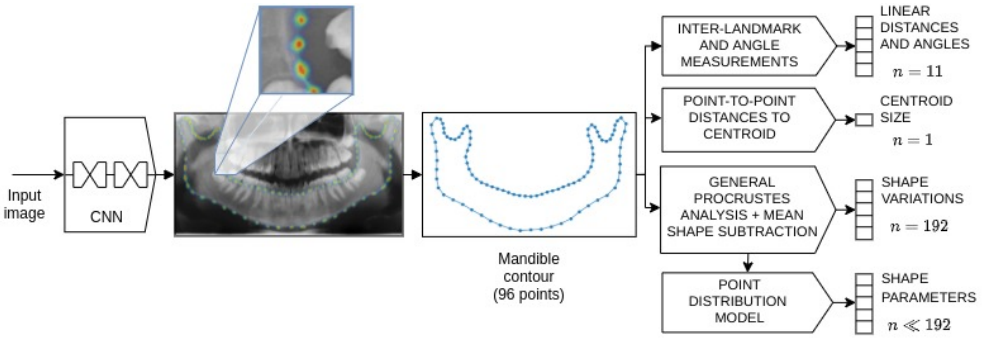


Figure 5.2: Process of describing the mandible shape from a panoramic X-ray image. In a first step, the mandible contour composed of both landmarks and semilandmarks is obtained automatically with a CNN. In a second step, four descriptors are applied, including a set of linear distances and angles; the centroid size, the variations from the mean shape, and the shape parameters produced by a PDM.

5.2.1 Data

The study presented in this chapter used a subset of the OPG-AS database described in Chapter 3, composed of 1,195 images of patients aged from five to 70. The age and gender distributions show a reasonable degree of uniformity, as it can be seen in Table 5.1.

Table 5.1: Age and sex distribution of the OPG dataset.

Age group	Male	Female	Total
[5, 10)	54	53	107
[10, 20)	93	99	192
[20, 30)	94	101	195
[30, 40)	75	109	184
[40, 50)	75	109	184
[50, 60)	65	108	173
[60, 70)	68	92	160
Total	524	671	1,195

The mandible contours were composed of eight anatomical landmarks, corresponding to the red points in Figure 5.3, namely the right and left condyles (RC and LC), the right and left coronoid processes (RCP and LCP), the right and left gonions (RG and LG), and the

USC UNIVERSIDADE

superior and inferior borders (SB and IB). On top of that, 88 semilandmarks were placed along the mandible contour to fill the gap between anatomical landmarks. To minimise the potential errors associated with the semilandmarks' placement [32], the annotators digitised them without a specific protocol regarding the position or the quantity. After that, they were automatically post-processed so there were a specific number of equally spaced semilandmarks between two consecutive anatomical landmarks (blue points in Figure 5.3). Specifically, there were eight semilandmarks between the condyles and the gonions, eight between the gonions and the inferior border, 10 between the condyles and the coronoid processes, and 18 between the coronoid processes and the superior border. The normalised mandible shape, therefore, contained 96 points in every case. This manual digitisation process was carried out through the Labelbox platform [126].

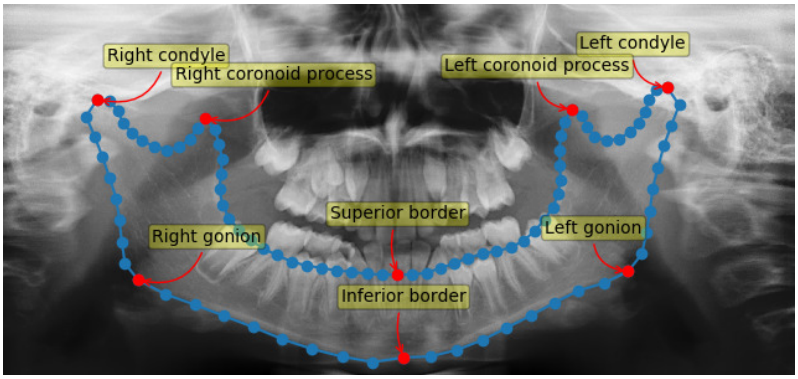


Figure 5.3: Normalised mandible digitisation.

5.2.2 Automatic digitisation of the mandible contour

To make the mandible shape description method work in a fully automatic way, an automatic method to digitise the mandible landmarks and semilandmarks without the need for an operator is proposed. This is specifically approached as a heatmap regression problem. Therefore, an fully convolutional network (FCN) is used to obtain one heatmap per contour point, i.e., 96. The target heatmaps were generated from a bivariate normal distribution, where the mean corresponds to the coordinates of the contour points, and the standard deviation was set to a fraction of the image width to ensure it works in the same way regardless of the resolution of the input image.

The point coordinates were obtained from the estimated heatmaps using the soft-argmax function, which allows for sub-pixel precision; it is also differentiable, meaning that a network can be trained end-to-end. This was applied as follows: after estimating the heatmaps, each one was normalised so that its pixel values add up to 1. Then, the coordinates of every image pixel were multiplied by the heatmap value at those coordinates. The results were summed according to Equation (5.1), where $\odot$ is the Hadamard product, P is the normalised heatmap, and w and h are the image width and height, respectively. This produced an approximation of the heatmap's peak value.

$$\langle \widetilde{x_{max}}, \widetilde{y_{max}} \rangle = \sum_{x=1}^{w} \sum_{y=1}^{h} \Big[\langle x, y \rangle \odot P_{x,y} \Big] \tag{5.1}$$

Four state-of-the-art FCNs specifically designed for landmark localisation were compared:

- The convolutional pose machines (CPMs) [226]. These involve the sequential application of a set of subnetworks. In the first stage, the network applies a set of convolution-pooling modules to output the probability map of each landmark (both anatomical landmarks and semilandmarks). In successive stages, the subnetworks operate directly over the belief maps obtained in the previous stage, enabling the inter-landmark relationships to be modeled and, therefore, the results to be refined.

- The stacked hourglass network (SHN) [162]. This is a similar approach to [226], but in this case, each subnetwork represents a downsampling-upsampling architecture and relies significantly on residual connections to overcome the vanishing gradient problem.

- The cascade pyramid network (CPN) [41]. This architecture is composed of a feature pyramid network (FPN) to detect the easiest landmarks, and a refinement network fed with the multi-resolution representations learned by the FPN to improve the detection of hard landmarks.

- The high-resolution network (HRNet) [209], which consists of several branches working in parallel at different resolutions. These branches are fused at intermediate points, providing rich multi-resolution representations.

USC UNIVERSIDADE

Although it was not a required output, we added the mandible mask as an output to increase the mandible information provided to the networks and, as a consequence, improve

their performance. The goal was for each network to produce 97 high-resolution outputs (one heatmap per contour point and one mandible mask).

Three different losses were applied to train the networks: 1. a BCE loss for the heatmap regression (Equation (5.2)); 2. a Dice loss for the mask prediction (Equation (5.3)); and 3. a RMSE loss for the point localisation (Equation (5.4)), where Y is the ground truth element, $\hat{Y}$ is the predicted element, P is the number of points of the mandible contour (in this case, 96), w and h are the image dimensions, x and y are the point coordinates, and ϵ is a small number used to avoid a division by zero. These three losses were combined in a single function as the logarithm of the product through Equation (5.5).

$$\mathcal{L}_h(Y,\hat{Y}) = -\frac{1}{P\cdot w\cdot h}\sum_{p=1}^{P}\sum_{x=1}^{w}\sum_{y=1}^{h}\left[\left[Y_{p,x,y}\cdot\log(\hat{Y}_{p,x,y})\right]+\left[(1-Y_{p,x,y})\cdot\log(1-\hat{Y}_{p,x,y})\right]\right] \tag{5.2}$$

$$\mathcal{L}_m(Y,\hat{Y}) = 1-\frac{2\cdot(Y\odot\hat{Y})}{\sum Y+\sum\hat{Y}+\epsilon} \tag{5.3}$$

$$\mathcal{L}_p(Y,\hat{Y}) = \sqrt{\frac{1}{P\cdot w\cdot h}\sum_{p=1}^{P}\sum_{x=1}^{w}\sum_{y=1}^{h}\left[Y_{x,y}-\hat{Y}_{x,y}\right]^2} \tag{5.4}$$

$$\mathcal{L} = \log\left(\mathcal{L}_m\cdot\mathcal{L}_h\cdot\mathcal{L}_p\right) \tag{5.5}$$

5.2.3 Mandible description

The quantitative description of the mandible is performed using four different descriptors. First, the mandible contours given by the anatomical landmarks and semilandmarks are employed to calculate a set of linear distances and angles. Some of these measurements are widely used in forensics and other clinical procedures, such as the ramus length [57], the bigonial and bicondylar breadth [208]; and the mandibular angle [200]. Other additional measures have been proposed to further improve the mandible description. Overall, eight linear distances and three angles were considered, as set out in Figure 5.4 and Table 5.2.

The second descriptor corresponds to the norm of the distances from each mandible contour point to the centroid and will be referred to as the centroid size [78]. To calculate the other two descriptors, the mandible contours are aligned through Generalised Procrustes Analysis (GPA) to provide optimal comparability, and centering them in the origin. The dataset of the

normalized shapes is referred to as $S_{N,2P}$, with S_i^j being the j-th coordinate of the i-th shape; N the number of shapes; and $2P$ the number of features of each shape (two coordinates per point). The mean shape ($\overline{\mathrm{S}}$) is calculated and subtracted from each aligned shape to obtain the vector of variations from the mean shape (ΔS), which is used as the third descriptor. Finally, the fourth descriptor is computed using a PDM, which involves decomposing a shape into a mean shape and a linear combination of modes of variation [48].

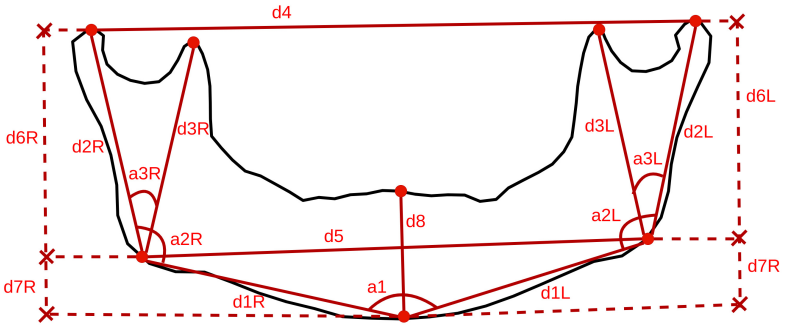


Figure 5.4: Proposed linear distances and angles.

We begin with the shape variations, $\Delta\mathrm{S}$, employing a singular value decomposition for each of them to transform $\Delta\mathrm{S}$ into $U\Sigma V^{\mathsf{T}}$, with: U being the matrix of the eigenvectors of $(\Delta\mathrm{S})(\Delta\mathrm{S})^{\mathsf{T}}$; Σ a diagonal matrix with the singular values; and V^{T} the matrix of the eigenvectors of $(\Delta\mathrm{S})^{\mathsf{T}}(\Delta\mathrm{S})$. The eigenvalues and eigenvectors are then extracted, the i-th eigenvalue giving the proportion of variance of the training shapes explained by the i-th eigenvector. As most of the shape variations can be represented with a reduced subset of modes of variation, the optimal number of modes required to explain a minimum proportion, l, of the total variance is computed from the eigenvalues.

To obtain the fourth descriptor, referred to as the shape parameters, the dataset is mapped to a k-dimensional space ($k \ll 2P$) using

$$(\Delta\mathrm{S})_k = (\Delta\mathrm{S})V_k, \tag{5.6}$$

where V_k is a matrix composed of the first k columns of V. The original dataset, S, was reconstructed via

$$\widetilde{\mathrm{S}} = \overline{\mathrm{S}} + (\Delta\mathrm{S})_k V_k^{\mathsf{T}} \tag{5.7}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 5.2: Description of the linear distances and angles. L|R: left and right sides.

Code	Measurements	Description
a1	Chin angle	Angle defined by the lines that join the gnathion and the mandibular angles
a2(L|R)	Mandibular angle	Angle formed by the lower margin of the body and the posterior margin of the ramus
a3(L|R)	Coronoid-condylar angle	Angle formed by the ramus and the imaginary line that connects the mandibular angle and the coronoid process
d1(L|R)	Diagonal length	Distance between the mandibular angle and the gnathion
d2(L|R)	Ramus length	Distance between the mandibular angle and the condyle
d3(L|R)	Coronoid-gonion	Distance between the gonion and the coronoid process
d4	Bicondylar breadth	Distance between the condylae
d5	Bigonial breadth	Distance between the gonions
d6(L|R)	Condyle-angle height	Vertical distance between the condyle and the mandibular angle
d7(L|R)	Angle-gnathion height	Vertical distance between the mandibular angle and the gnathion
d8	Chin height	Distance from interdentale to gnathion

The PDM approach has three main benefits: 1. the dimensionality of the problem is reduced, while most of the shape variation is retained; 2. the low-dimensional shapes produced by $(\Delta \mathrm{S})_k$ are orthonormal to each other; and 3. it helps us to conduct graphical assessments of variations in the mandible's shape.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.3 Experiments and performance evaluation

In this section, two different experiments are described, namely the validation of the automatic mandible digitisation system, and the assessment of the predictive capabilities of the proposed mandible descriptors in the problem of sex and age estimation.

5.3.1 Mandible digitisation

The mandible segmentation methods based on CNNs were compared in terms of the point-localisation error. The input image resolution was set to 256x512 pixels, and we used different hyperparameter tuning for each network. The stacked hourglass network (SHN) was evaluated for different depths (four, five, and six) and different initial filters in each stage (16, 32, and 64). The number of stages was set to two. The CPM was tested with one and two refinement stages, and with 64, 128, and 256 initial convolutional filters for each stage. The CPN was evaluated with two different backbones, namely ResNet50 and ResNet101. Finally, the HRNet performance was assessed using different network sizes (C parameter in [209]), specifically 24, 48, and 96.

The performance of the four networks was evaluated through a set of metrics, namely the landmark-to-landmark error corresponding to the Euclidean distance between the real and estimated anatomical landmarks; the point-to-point error (PT2PT) was obtained via the Euclidean distance between the real and estimated contour points (both landmarks and semi-landmarks), and the minimum point-to-curve error (PT2CRV) corresponded to the minimum Euclidean distance between each estimated point (both landmarks and semilandmarks) and the real mandible contour, averaged over all the contour points. All the errors were calculated in mm by using the resolution information of the X-ray acquisition device (11.11 pixels/cm).

The data were split into training (800 images), validation (200 images), and test (200 images) sets, and this division was retained in every experiment. The three subsets were equally distributed in terms of the age and gender of the subjects. The networks were trained with the Adadelta optimiser, and the batch size was set to four as a good compromise between the convergence speed and regularisation effect. Image augmentation transformations were applied to increase the dataset's variability, including the horizontal flip, translation, rotation, and contrast and brightness disturbance [222].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.3.2 Sex and age estimation

In the second experiment, the proposed mandible description method was validated in a real problem representative of mandible change and widely studied in the literature: sex and chronological age estimation. In this regard, both the shape parameters and the centroid size were used to make a visual assessment of the mandible variations according to the sex and age of a subject.

Furthermore, predictive models were developed for sex and age estimation by using each of the proposed mandible descriptors as the independent variables. To avoid potential collinearity problems, especially with the linear distances and angles, ridge regression and classification models were used for age and sex estimation, respectively. To evaluate the robustness of the proposed automatic approach, the results obtained with the CNN-digitised mandible contour were compared with those obtained with a manual digitisation process —referred to as the semiautomatic method.

The sex estimation performance was evaluated through the accuracy metric and the per-class F1 metrics, following the equations described in Section 4.3. On the other hand, the age estimation performance was assessed through the absolute error between the real and estimated ages.

Furthermore, the best sex and age estimation models obtained in the previous step were compared to other methods proposed in the literature. The metrics in this comparison were those provided by the other researchers, namely the accuracy in the case of sex classification, and the standard error (SE), the coefficient of determination (R^2), and the p-value associated with the F-test in the case of age regression.

As previously mentioned, sex and age estimations are more successful for specific age ranges. As a result, and to enable a reliable comparison with other methods, the sex estimation models were tested on subjects older than 18 and age estimation on those below that age.

5.4 Results

In this section, the results concerning the experiments described in the previous sections are presented.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.4.1 CNN comparison

The performances of the four tested architectures are set out in Table 5.3. First, to identify the best-performing setup in terms of the point-to-point error, the different configurations of each network were trained without including the mandible mask as an additional output. Although the differences were almost negligible: the SHN performed better with a depth of four and 64 initial filters; the CPM had better results with one refinement stage and 256 initial filters; the CPN functioned best with ResNet50 as the backbone network; and the optimal performance of the HRNet was achieved with a size of 48.

Table 5.3: Point detection absolute error ($\mu \pm \sigma$) of the best configurations for each CNN model, with and without the mandible mask as an additional target. RG: right gonion; LG: left gonion; SB: superior border; IB: inferior border; RC: right condyle; LC: left condyle; RCP: right coronoid process; LCP: left coronoid process; PT2PT: point-to-point errors measured at all mandible points; and PT2CRV: point-to-curve errors, also measured at all mandible points.

Model	RG[(a)]	LG[(a)]	SB[(a)]	IB[(a)]	RC[(a)]
SHN [162]	**3.1 ± 2.5**	3.5 ± 2.5	1.5 ± 1.3	1.9 ± 1.6	1.2 ± 0.8
SHN+Mask	3.2 ± 2.6	**3.2 ± 2.3**	1.4 ± 1.3	**1.6 ± 1.5**	**1.0 ± 0.7**
CPM [226]	3.5 ± 2.7	3.5 ± 2.4	1.5 ± 1.2	1.9 ± 1.6	1.4 ± 0.9
CPM+Mask	3.3 ± 2.5	3.5 ± 2.4	1.7 ± 1.5	1.8 ± 1.6	1.4 ± 0.9
CPN [41]	3.4 ± 2.8	3.5 ± 2.6	2.0 ± 1.3	1.8 ± 1.5	1.4 ± 1.0
CPN+Mask	3.3 ± 2.5	3.3 ± 2.2	1.4 ± 1.2	1.7 ± 1.5	1.5 ± 0.9
HRNet [209]	3.3 ± 2.6	3.4 ± 2.5	1.4 ± 1.3	1.7 ± 1.7	1.3 ± 0.9
HRNet+Mask	3.2 ± 2.7	3.2 ± 2.3	**1.4 ± 1.2**	1.7 ± 1.6	1.2 ± 0.9

(a) Metrics in mm, using the resolution information of the X-ray acquisition device (11.11 pixels/mm).

Without including the mandible mask as an additional network target, the gonions were the worst localised anatomical landmarks, with mean errors between 3.1 (SHN) and 3.5 (CPM) mm for the RG and between 3.4 (HRNet) and 3.5 (CPN) mm for the LG. The SB errors ranged from 1.4 (HRNet) to 2.0 (CPN) mm, and the IB errors from 1.7 (HRNet) to 1.9 (SHN and CPM) mm. The lowest errors were yielded for the condyles, ranging from 1.2 (SHN) to 1.4 (CPN) mm for the RC, and between 1.3 (SHN) and 1.5 (CPM) mm for the LC. Finally, the coronoid processes were best estimated by the SHN, with mean errors of 1.7 and 1.8 mm for the RCP and LCP, respectively, while the worst were localised by the CPM, yielding a mean

Table 5.3: *(Continuation)*

Model	LC$^{(a)}$	RCP$^{(a)}$	LCP$^{(a)}$	PT2PT$^{(a)}$	PT2CRV$^{(a)}$
SHN [162]	1.3 ± 0.9	1.7 ± 1.9	1.8 ± 2.0	1.8 ± 1.6	0.3 ± 0.3
SHN+Mask	**1.1 ± 0.9**	**1.4 ± 1.5**	**1.5 ± 1.6**	**1.7 ± 1.5**	**0.2 ± 0.2**
CPM [226]	1.5 ± 1.2	2.2 ± 2.0	2.3 ± 2.1	1.9 ± 1.5	0.3 ± 0.3
CPM+Mask	1.4 ± 0.9	2.2 ± 1.9	2.1 ± 1.9	1.9 ± 1.5	0.3 ± 0.3
CPN [41]	1.4 ± 1.0	2.0 ± 1.8	2.2 ± 1.9	1.9 ± 1.6	0.2 ± 0.3
CPN+Mask	1.6 ± 1.1	2.0 ± 1.7	2.3 ± 1.9	1.9 ± 1.5	0.3 ± 0.3
HRNet [209]	1.4 ± 1.0	1.9 ± 1.7	2.0 ± 2.0	1.8 ± 1.5	0.2 ± 0.3
HRNet+Mask	1.4 ± 1.0	1.7 ± 1.8	2.1 ± 2.0	1.7 ± 1.5	0.2 ± 0.2

error of 2.2 and 2.3 mm. The lowest point-to-point (PT2PT) and point-to-curve (PT2CRV) errors were achieved by the HRNet, at 1.8 and 0.2 mm, respectively.

These network setups were retrained from scratch with the inclusion of the mandible mask. The impact of this addition differed depending on the network: 1. in the SHN, the performance increased for every metric other than the RG error. The improvement on landmark localisation varied between 0 mm on the SB and 0.2 mm on both the LG and the RCP. The PT2PT errors also decreased, by an average of 0.1 mm, respectively; 2. the CPM improved the results noticeably in the RG, IB, LC, and LCP, but performed worse in the SB. The PT2PT error remained unchanged; 3. there was a noticeable improvement in the CPN with respect to the localisation of both gonions and the SB and IB, but the localisation accuracy fell for the other landmarks. The PT2PT and PT2CRV errors followed the same pattern as the CPM, with minor degradation of 0.1 mm for the latter; and 4. the HRNet improved for all the landmarks other than the LCP. The mandible contour fitting was also improved, with a reduction in the PT2PT errors.

Overall, the SHN with the addition of the mask outperformed the rest of the approaches tested, with the best results in the PT2PT and PT2CRV errors; it also produced the smallest errors in six out of eight anatomical landmarks. It was therefore used in the experiments described in what follows.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

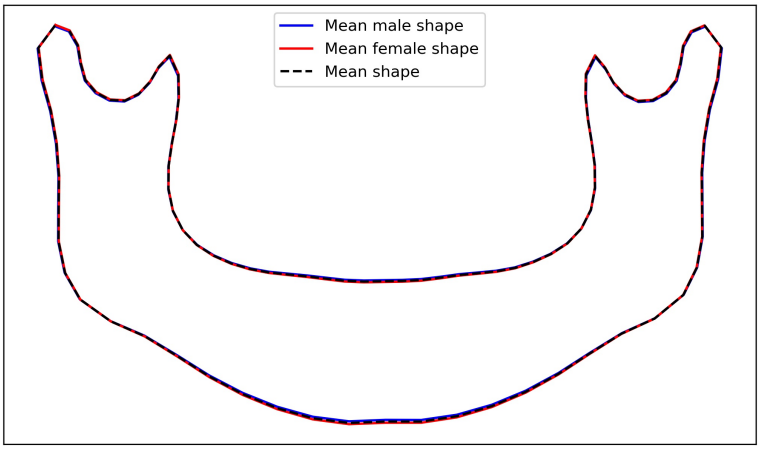
5.4.2 Mandible digitisation

As shown in Table 5.4, the greatest interobserver agreement on the issue of landmark digitisation occurred for the condyles (1.08 and 1.44 for RC and LC, respectively), and the biggest differences were related to gonion localisation (4.73 and 3.85 for RG and LG, respectively). Comparatively, the network yielded lower errors for every landmark other than the SB (1.20 *vs.* 1.43) and IB (1.58 *vs.* 1.60). The maximum difference was found for the RG, where the network reduced the degree of error by an average of 1.5mm. Concerning the linear distances and angles, the interobserver agreement in the angles was noticeably reduced by the network in the case of the chin (a1, 2.57 *vs.* 1.45) and coronoid-condylar (a3, 1.95 *vs.* 1.27) angles. The smallest interobserver error in the distance measurements was found for the chin height (d8, 0.92), while the greatest disagreement by far related to the bigonial breadth measurement (d5, 4.60). The neural network was also capable of reducing the differences between the observers and was especially noticeable for the diagonal length (d1, 3.29 *vs.* 2.24), ramus length (d2, 3.69 *vs.* 2.19), bigonial breadth (d5, 4.60 *vs.* 3.43), condyle-angle height (d6, 3.50 *vs.* 2.06), and angle-gnathion height (d7, 3.22 *vs.* 1.91), with a reduction of more than 1 mm for all of them. Overall, the overlapping of the mask of the mandible contour was slightly better with the mask estimated by the network (0.98 *vs.* 0.99).

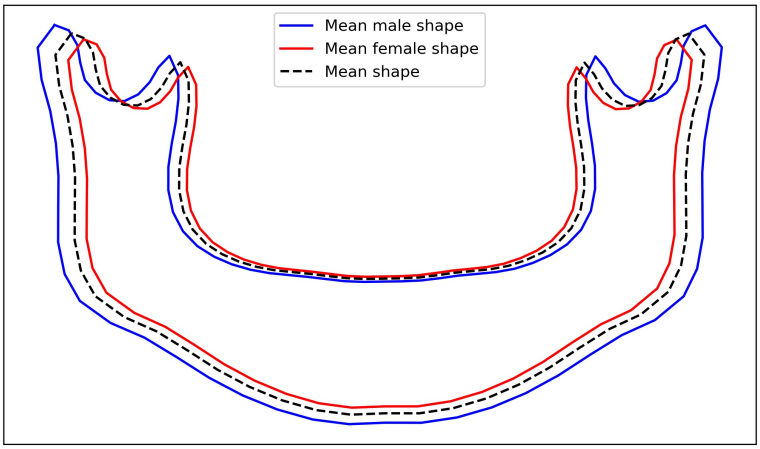
5.4.3 Sex and age estimation

For both the sex and age estimations, the first shape variation mode produced by the PDM was the most significant in relation to the classification/regression models. To visualise the main differences between the male and female mandibles, the mean male and female shapes were reconstructed using only the first mode (Equation (5.7), with $k = 1$). Furthermore, the effect of the mandible size was also assessed by scaling the mean reconstructed shapes with the mean male and female centroid sizes. Figure 5.5a demonstrates that the mean adult male mandible shape is very similar to the female mandible shape. However, when the mean centroid size is included (Figure 5.5b), the male mandible tends to be slightly bigger than that of the adult female subjects. The mean mandible shape is also reconstructed for the different age groups. As shown in Figure 5.6a, the younger age group had a more open rami, while the older age groups had a more pointy chin. When the size component is added, a clear mandible growing pattern can be seen (Figure 5.6b).

USC UNIVERSIDADE



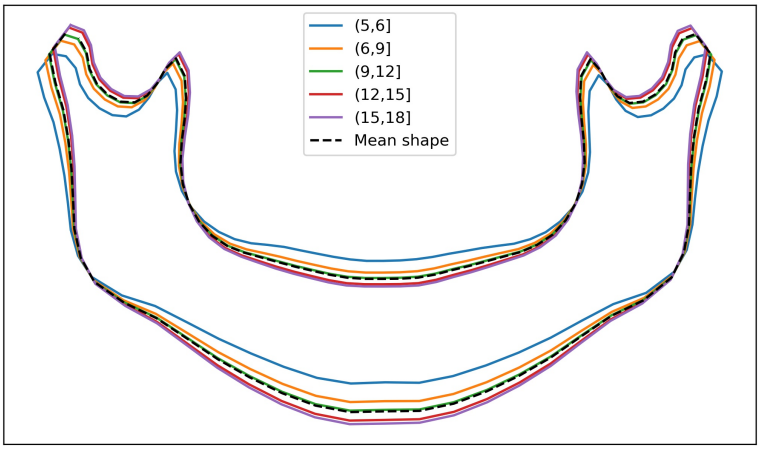
(a) First mode of variation.



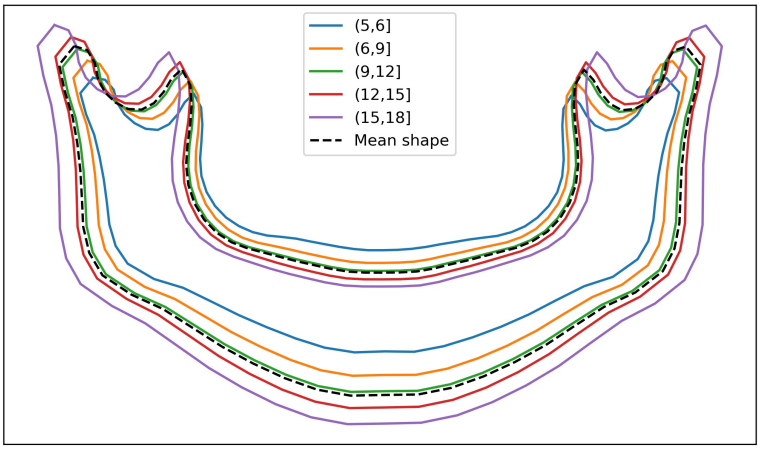
(b) First mode of variation + centroid size.

Figure 5.5: Mandible variations in subjects older than 18 regarding the sex.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) First mode of variation.



(b) First mode of variation + centroid size.

Figure 5.6: Mandible variations in subjects younger than 18 regarding the age.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 5.4: Annotation errors of the 2nd observer and the prediction errors of the best-performing network (SHN), both measured against the gold standard (1st observer).

Metric		Absolute error ($\mu \pm \sigma$)	
		Interobserver	**Network**
Point-to-point absolute error (mm)	RG	4.73 ± 2.93	3.23 ± 2.59
	LG	3.85 ± 2.83	3.21 ± 2.31
	SB	1.20 ± 1.36	1.43 ± 1.26
	IB	1.58 ± 1.35	1.60 ± 1.52
	RC	1.08 ± 0.87	0.99 ± 0.75
	LC	1.44 ± 1.27	1.13 ± 0.92
	RCP	2.09 ± 2.28	1.40 ± 1.49
	LCP	2.35 ± 2.35	1.55 ± 1.65
Point-to-curve (mm)	PT2CRV	0.20 ± 0.09	0.21 ± 0.23
Angles absolute error (degrees)	a1 (chin angle)	2.57 ± 1.68	1.45 ± 1.42
	a2 (mandibular angle) [(a)]	0.81 ± 0.71	0.81 ± 0.62
	a3 (coronoid-condylar angle) [(a)]	1.95 ± 1.49	1.27 ± 1.09
Linear distances absolute error (mm)	d1 (diagonal length) [(a)]	3.29 ± 2.24	2.24 ± 1.78
	d2 (ramus length) [(a)]	3.69 ± 2.40	2.19 ± 1.85
	d3 (coronoid-gonion)	2.17 ± 1.83	1.39 ± 1.39
	d4 (bicondylar breadth)	1.37 ± 1.30	1.28 ± 1.09
	d5 (bigonial breadth)	4.60 ± 3.29	3.43 ± 1.60
	d6 (condyle-angle height) [(a)]	3.50 ± 2.27	2.06 ± 1.79
	d7 (angle-gnathion height) [(a)]	3.22 ± 2.13	1.91 ± 1.77
	d8 (chin height)	0.92 ± 0.89	0.70 ± 0.74
Mask overlapping	DSC	0.98 ± 0.01	0.99 ± 0.00

(a) Average of right and left sides.

Table 5.5 compares the results of the semiautomatic and automatic methodologies, and is where it can be seen that the performance differences between them varied greatly depending on the mandible information used. When the linear distances and angles were employed in a fully automatic way, the accuracy increased by 2%. The classification method based on the centroid size yielded similar results both for the semiautomatic and the automatic approaches, with an accuracy value of about 0.750, while the performance for shape variations fell slightly

USC UNIVERSIDADE

with the automatic approach. The use of the shape parameters produced by the PDM led to better results in the automatic approach, with an improvement of 1.9% of accuracy and a more balanced F1 measure between males and females. Finally, the combination of the shape parameters and the centroid size produced the best results in every aspect. Specifically, the automatic approach outperformed the semiautomatic method by almost 2%, reaching an overall accuracy of 0.878. The F1 metric was also the highest, with values of 0.857 and 0.894 for males and females, respectively.

Table 5.5: Performance of the sex classification method in those aged between 18 and 70. Acc: accuracy.

Predictor	**Semiautomatic**			**Automatic**		
	Acc	$F1_{male}$	$F1_{female}$	**Acc**	$F1_{male}$	$F1_{female}$
Linear distances and angles	0.808	0.754	0.842	0.821	0.778	0.849
Centroid size	0.756	0.694	0.798	0.750	0.683	0.798
Shape variations	0.769	0.739	0.793	0.756	0.708	0.791
Shape parameters	0.731	0.691	0.761	0.750	0.748	0.769
Shape parameters + centroid size	**0.859**	**0.831**	**0.879**	**0.878**	**0.857**	**0.894**

The results produced by the automatic sex classifier were compared to the outcomes of the methods by other researchers reporting an accuracy greater than 0.8, as set out in Table 5.6. To enable a reliable comparison to be made, the findings are reported for the same age ranges used by these other authors. The proposed automatic method outperformed the other approaches in seven out of eight comparisons, with differences between −0.8% and +7.9%.

Table 5.6: Comparison of the sex classification results in the literature (semiautomatic) and those of the best-performing automatic approach presented in this chapter. Meas.: Measurements. Meas. legend: DB: dry-bone; 3DS: 3D scanner; CT: computed tomography. Method legend: DFA: discriminant function analysis; LoR: logistic regression; PDM: point distribution model; RC: ridge classification. N: sample size. Test approach legend: TT: train-test; CV: cross validation; LOO: leave-one-out. Acc: accuracy.

Work	**Age**	**Meas.**	**Method**	**N**	**Test**	**Acc.**[(a)]	**Acc of this work** [(a,b)]
Saini [192]	23-65	DB (5)	DFA	116	-	0.802	0.881 (+ 7.9%)
Giles [86]	21-75	DB (9)	DFA	265	TT	0.850	0.871 (+ 2.1%)
Steyn [208]	-	DB (5)	DFA	81	-	0.815	- [(c)]
Dayal [53]	25-69	DB (6)	DFA	60	CV	0.839	0.847 (+ 0.8%)
Pokhrel [179]	-	DB (4)	DFA	79	-	0.829	- [(c)]
Abualhija [2]	21-45	OPG (3)	LoR	50	TT	0.800	0.857 (+ 5.7%)
Franklin [76]	18-70	3DS (38)	PDM+LoR	225	CV	0.831	0.878 (+ 4.7%)
Lin [135]	21-70	3D CT (10)	DFA	240	LOO	0.879	0.871 (− 0.8%)
Ours	18-70	OPG (96)	RC	935	TT	**0.878**	

(a) Shape parameters and centroid size were used, as they yielded the best results (Table 5.7).
(b) The accuracy was calculated for the same age range as original publications.
The percentage differences were also reported.
(c) The accuracy could not be calculated for the same age range, as the original work did not report this information.

The age estimation results are presented similarly in Table 5.7. Each of the four descriptors yielded similar results when applying the semiautomatic or the fully automatic method. The main differences were obtained with the shape variations (1.57 and 1.79 for semiautomatic and automatic mode, respectively) and the shape parameters (with an improvement of 0.12 years in the error of the automatic mode). The best-performing descriptors were the shape variations and the shape parameters in the semiautomatic and fully automatic methods, respectively. When combining the shape parameters and the centroid size descriptors, the absolute error of both approaches was significantly enhanced (with improvements of 0.04 and 0.13 on average, respectively, with respect to the best performing single-descriptor model).

The age estimation methods were compared to those proposed by other authors with the same performance metrics, as set out in Table 5.8. Specifically, the performance of the proposed approach was reported for the subadult age range available in our dataset (5-17 years). Although the R^2 values were slightly worse (maximum of 0.880 *vs.* 0.804), our method outperformed these methods in terms of the SE (maximum of 2.4 *vs.* 2.0).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.5 Discussion and conclusions

This chapter presents an automatic method for detecting and describing the mandibular contour. The mandible detection was carried out with the SHN. This CNN produced, by a large margin, more confident detections than those of the experts for every anatomical landmark except SB and IB and every angle and linear measurement, as well as in the overlapping of the mandible mask.

Table 5.7: Mean and standard deviation of the absolute error (in years) in the age estimation method for subjects aged between five and 17.

Predictor	**Absolute error ($\mu \pm \sigma$)**	
	Semiautomatic	**Automatic**
Linear distances and angles	1.75 ± 1.24	1.80 ± 1.28
Centroid size	2.40 ± 1.83	2.38 ± 1.83
Shape variations	1.57 ± 1.17	1.79 ± 1.17
Shape parameters	1.82 ± 1.26	1.70 ± 1.09
Shape parameters + Centroid size	**1.53** ± **1.26**	**1.57** ± **1.21**

Table 5.8: Comparison of the best age estimation results of the automatic methodology and the semiautomatic results presented previously in the literature. Meas. legend: 3DS: 3D scanner. Method legend: LR: linear regression; PDM: point distribution model; RR: ridge regression; N: sample size; SE: standard error (in years); R^2: coefficient of determination; p: p-value of the F-test.

Work	**Age**	**Meas.**	**Method**	**N**	**SE**	**R^2**	**p**
Franklin [73]	1-17	3DS (38)	LR	79	2.4	0.834	1×10^{-31}
Franklin [74]	1-17	3DS (38)	LR	79	2.1	0.880	1.8×10^{-37}
			PDM+LR		2.4	0.827	1.8×10^{-27}
Ours	5-17	OPG (96)	RR[(a)]	260	**2.0**	0.804	0.00055

(a) Shape parameters and centroid size were used, as they yielded the best results (Table 5.7).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

To perform the quantitative description of the mandible, four different descriptors have been proposed. The combination of the shape parameters and the centroid size not only allowed us to summarise the shape and size information numerically, but to also produce comprehensive visualisations of the mandible variations between different populations, age cohorts, and sexes. In this regard, the first and main shape parameter given by the PDM represented a shape evolution in accordance with that reported in the clinical literature [40, 79]. This fact led us to confirm that the proposed approach is useful to assess the mandible shape changes both quantitatively and qualitatively.

Finally, all this shape information was used to compare mandible description for both a semiautomatic and a fully automatic method for the selected validation experiment of classifying sex and estimating chronological age. The two methods were then evaluated in five different scenarios: linear distances and angles; centroid size; shape variations; mandible shape parameters provided by the PDM; and mandible shape parameters together with centroid size. The semiautomatic method required an expert to annotate the mandible contour's landmarks, which were then used to estimate both the sex and the age. The automatic method retrieved the mandible contour extracted by the CNN.

Concerning the sex classification experiments, the top accuracy of the semiautomatic and automatic methods was achieved when combining the shape parameters and the centroid size. The F1 values of over 0.83 for both classes confirmed that the models were not biased toward a specific gender. The accuracy fell slightly when we used size-free descriptors alone, such as the shape variations and the shape parameters, or linear distances and angles. However, it is notable that the automatic method achieved a higher accuracy when relying on linear distances and angles. This is in line with the significant performance differences between the network and the observers when extracting these measurements.

Comparing the sex classification performance with that of previous studies, the proposed methodology outperformed almost every other methodology except the approach in [135], which used 3D CT images. It is also notable that three out of the 8 studies we analysed did not describe any validation scheme [179, 192, 208], while one performed a train-test split on part of the dataset [86]. It should also, therefore, be noted that the data sample used by our team is composed of 935 images, making it the largest database used in an investigation of this kind.

Regarding the age estimation results, the absolute error of the proposed automatic method was between 1.57 and 2.38 years on average. Although the proportion of the explained variance given by R^2 was slightly lower than in the other methods, the proposed method performed

USC UNIVERSIDADE

better concerning the SE. This is especially remarkable, given that our study did not include subjects younger than five; if it had been done, the results may have been even better, due to the significant development that occurs in that age range.

Although the studies using CNN-based methods that employ an entire OPG to conduct sex and age estimations performed better, they only serve the purpose for which they were developed [151, 222, 224]. On the other hand, the method we propose based on automatic mandible description performs well when estimating age and sex; it is also more versatile, as it can also be employed in other applications, such as in evaluating the mandible shape differences between populations, sexes, and age cohorts, and for disease diagnosing or surgical management.

In conclusion, the automatic method we describe in this chapter is very reliable both for the extraction of the mandible contour —with a dramatic improvement in the time it takes to do so— and the determination of the chronological age and sex from its shape. Consequently, the methodology proposed meets the requirements proposed for objective 2 of this doctoral thesis, and it can be valuable in any field that requires a quantitative description of the mandible shape and a visual representation of its changes, such as clinical practice, surgical management, dental research, or legal medicine.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CHAPTER 6

TWO-STAGE CNN FOR AGE AND SEX PREDICTION FROM TOOTH APPEARANCE

Chronological age and biological sex estimation have already led to the development of both semiautomatic and automatic prediction models. However, while the former are expensive in terms of time and human resources, the latter lack the interpretability needed to be applicable in real scenarios. In this chapter, a fully automatic method for predicting age and sex from tooth appearance is proposed. It first applies a tooth detector network, which extracts the oriented bounding boxes of each tooth, and then feeds the image features inside these boxes into a second network to produce per-tooth age and sex probability distributions. Finally, our method aggregates those estimated distributions by following an uncertainty-aware policy. Besides, a system to generate per-tooth heatmaps is proposed to represent the regions of the image where the network is paying more attention. The results showed that this approach yields a lower mean absolute error than any other previous work, being 0.97 years. Moreover, the accuracy of the sex classification was 91.82%, enforcing the suitability of the teeth to carry out this task. The developed approach also enables the analysis of age and sex estimations on each tooth, both quantitative and qualitative, allowing the experts to find the most relevant teeth for each task or population cohort or to detect potential development problems. In conclusion, this method provides excellent performance in both age and sex prediction with a high degree of explainability, making it suitable for a wide range of application scenarios.

The contents of this chapter are extracted from the following publication:

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Vila-Blanco, N.[*,†], Varas-Quintana, P.[‡], Aneiros-Ardao, Á.[‡], Tomás, I.[†,‡], and Carreira, M. J.[*,†] (2021). XAS: automatic yet eXplainable Age and Sex determination by combining imprecise per-tooth predictions. Article submitted to IEEE Transactions on Medical Imaging and currently under review.

6.1 Introduction

As mentioned in previous chapters, chronological age and biological sex are, together with race, the most important human features to be considered in anthropological and forensic studies [72]. Moreover, the development status of the bones has been utilised successfully for carrying out both age and sex estimations, being the only available method when examining archaeological remains or severely injured corpses. It is worth noting that there is not a single method based on bone development that outperforms others systematically, as the performance of each one depends on numerous factors. For instance, there are specific age prediction methods developed for subadults and others that work better in adults [44]. In the same way, sex estimation works better in adults because dimorphic skeletal indicators are not developed until after puberty [44].

The most widely used body part for carrying out age prediction is the dentition. Tooth development is highly correlated with chronological age and its physical properties make it extremely resistant and less affected by external factors, allowing a wide range of application scenarios [229]. Although the potential of the teeth to estimate the biological sex is more limited than other skeletal parts, they are still used in case they are the only available body remains or as a complement for other sex determination methods [109]. Specifically, the canines were reported as the most dimorphic teeth [110], though other teeth also showed significant differences [195].

X-ray-based imaging techniques, particularly dental panoramic radiographs (OPGs), represented a step forward because they allowed clinicians to assess dentition development *in situ*, with less invasive and faster procedures. On the one hand, this enabled them to perform

[*]Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) and Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.
[†]Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Spain.
[‡]Oral Sciences Research Group, Special Needs Unit, Department of Surgery and Medical-Surgical Specialities School of Medicine and Dentistry, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.

chronological age and sex estimations through many comprehensive methodologies based on simple dental measurements [27, 60, 163]. On the other hand, the advancements in image processing, especially the breakthrough development of deep learning techniques, led to the development of faster methods for age and sex determination, which reduce the inherent subjectivity of manual methods [55, 151, 222, 224]. However, the most interpretable solutions still need human intervention [58, 149], while the fully automatic approaches are much less interpretable [55, 100].

In this chapter, a method for the prediction of age and sex from the appearance of the teeth is proposed. It relies on deep learning techniques to detect the bounding box of each tooth, taking into account its orientation. Then, it reuses the learned tooth features to estimate the age and sex probability distributions of each tooth. Finally, it aggregates these distributions through an uncertainty-aware policy.

6.2 Related work

Automatic tooth extraction in dental images has been approached from many perspectives. The first works were aimed at human identification on bitewing dental images and carried out tooth detection through the recognition of gaps between the lower and upper jaws and between consecutive teeth, usually by integral projections [108]. However, the most successful approaches in recent years have used methods based on deep learning techniques —more specifically CNNs— on OPGs. In this regard, the available research has approached the problem from three different points of view, as seen in Table 6.1: localisation —estimation of tooth bounding boxes—, segmentation —extraction of tooth mask—, and classification —prediction of tooth type or numbering.

Regarding tooth localisation, Oktay *et al.* [167] 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 [142, 202, 216] or centre-based tooth detectors [46]. Regarding tooth segmentation, this was initially approached by Silva *et al.* [203] and Koch *et al.* [117], targeting the dental mask as a whole through Mask R-CNN and U-Net architectures, respectively. This approach was refined by Jader *et al.* [107] and Lee *et al.* [129] to perform instance segmentation, that is, to get the individual mask for each tooth.

UGR

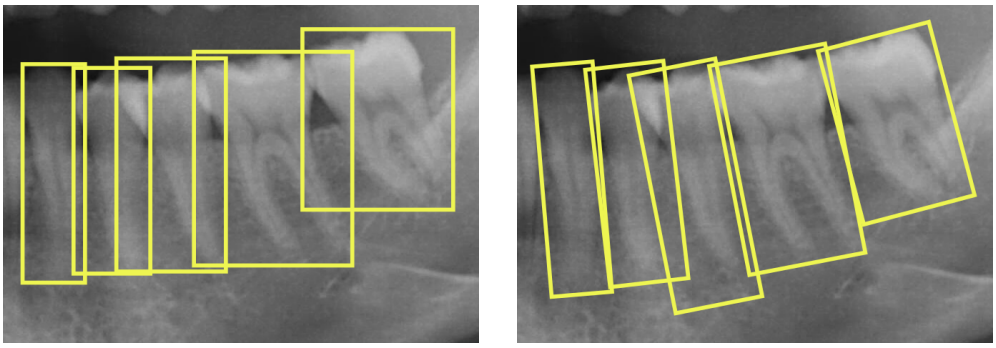
Table 6.1: Main deep learning methods for tooth detection/segmentation in OPGs. Loc.: localisation; Seg.: segmentation; Class.: classification; T: tooth type; I: individual tooth; Dent.: dentition; C: children; A: adult.

Work	Loc.	Seg.	Class.		Dent.		Method	#Images
			T	I	C	A		
Oktay [167]	✓	✗	✓	✗	✗	✓	Modified AlexNet	100
Silva [203]	✗	✓	✗	✗	✓	✓	Mask R-CNN	1,500
Jader [107]	✓	✓	✗	✗	✓	✓	Mask R-CNN	1,500
Tuzoff [216]	✓	✗	✓	✓	✗	✓	Faster R-CNN + VGG-16	1,574
Koch [117]	✗	✓	✗	✗	✓	✓	U-Net	1,500
Leite [132]	✓	✓	✗	✗	✗	✓	Deeplab-v3 + FCN-ResNet101	153
Silva [202]	✓	✓	✓	✓	✗	✓	PANet	543
Kim [115]	✓	✗	✓	✓	✗	✓	Faster R-CNN + Heuristics	303
Muramatsu [157]	✓	✗	✓	✗	✗	✓	DetectNet + ResNet50	100
Lee [129]	✓	✓	✗	✗	✗	✓	Mask R-CNN	50
Muresan [158]	✓	✓	✗	✗	✗	✓	ERFNet + Histogram classification	1,000
Mahdi [142]	✓	✗	✓	✓	✗	✓	Faster R-CNN + Heuristics	1,000
Chung [46]	✓	✗	✓	✓	✗	✓	Center detector + Bbox regressor	818
Ours	✓[(a)]	✗	✓	✓	✓	✓	**Rotated R-CNN**	**1,746**

(a) The result is an oriented bounding box.

The datasets used to validate these methodologies vary in size, being the biggest the one collected in [216] with 1,574 OPGs. They also varied regarding the age of the population sample. Although all works used OPGs belonging to adult subjects with fully developed dentition, only a small subgroup of investigations included images from children with deciduous or mixed dentition [107, 117, 203], none of which performed tooth numbering. It is remarkable that the work by Silva *et al.* [202] even discarded the images belonging to children, although they were available.

The available studies have carried out tooth localisation through standard bounding boxes, that is, rectangles that are parallel to the x-y axes. This aspect results in an easier box representation —only four coordinates are required— and easier implementation of detection steps —such as the region of interest (ROI) Pooling operation in Faster R-CNN architectures. However, the lack of orientation alignment between the tooth and its box, which is especially noticeable when the teeth are significantly rotated, can cause a high overlap between adjacent boxes, as shown in Fig. 6.1. Moreover, standard bounding boxes do not provide information regarding the tooth aspect ratio —width-height ratio— or its orientation, which is valuable in fields like paediatric dentistry.



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

Figure 6.1: Comparison of standard (a) and oriented (b) bounding boxes. In the former, there is a high overlapping between neighbouring boxes, which is especially noticeable in the posterior teeth.

In the last few years, some methodologies have been also proposed to estimate age and sex from OPGs in a fully automatic way (Table 6.2), most of them based on deep learning techniques. Concerning age estimation methods, the available studies focused on different objectives: staging of specific teeth [13, 58, 149], numeric age regression [52, 55, 222], age group classification [116], and legal age classification [90]. The methodologies developed in this regard follow two different principles. In some cases, the authors proposed to restrict the age estimation to image crops or segmentations belonging to a specific tooth or group of teeth [13, 52, 58, 115, 149]. In these scenarios, the previous step for detecting these teeth has been carried out manually [58, 115, 149], semiautomatically [52], or automatically [13]. Other works proposed one-step pipelines to estimate the chronological age directly from the raw image, with no intermediate steps [55, 90, 222]. Regardless of the method applied, the authors agreed that age estimation from dental panoramic radiographs is feasible only during the tooth development period.

The automatic sex estimation methods from OPGs are scarcer, and all of them follow the same approach: a one-stage CNN to perform predictions directly from the images [100, 114, 151, 224]. The sex classification accuracy varied among the studies and evaluated age cohorts, although it remained above 90% for subjects older than 16. The performance was reportedly lower when younger subjects were included, especially for those younger than 8 [224]. The almost non-existence of dimorphic skeletal indicators before puberty and the significant tooth

USC UNIVERSIDADE

development activity at that age make it difficult to find sexually dimorphic features in the images belonging to children and, thus, the sex classification.

Table 6.2: Main automatic methods for age/sex estimation from OPGs. 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 [58]	AlexNet[b]	38	-	DSC
	Čular [52]	Active Appearance Model + Radial Basis Network	48	10-25	NA
	De Back [55]	Bayesian CNN	-	5-25	NA
	Merdietio [149]	DenseNet201[b]	38	-	DSC
	Banar [13]	YOLO-like CNN + U-Net + DenseNet201	38	7-24	DSC
	Vila-Blanco [222]	Two-path CNN	-	5-89	NA
	Kim [116]	ResNet152[b]	18, 28, 38, 48	-	AGC
	Guo [90]	SE-ResNet101	-	5-24	LAD
Sex	Ilić [100]	VGG-16	-	-	BSC
	Milošević [151]	VGG-16	-	19-85	BSC
	Ke [114]	Multi-Feature Fusion Model	-	16-70	BSC
	Vila-Blanco [224]	Two-path CNN	-	5-90	BSC

(a) FDI notation.

(b) This method requires an initial manual crop of the regions of interest.

Among the proposed automatic age and sex prediction methods, there are noticeable differences regarding the interpretability and the potential usefulness of the estimations. Age prediction approaches aimed at third molar staging are easily understandable since the proposed stages are based on well-known scoring systems relying on certain tooth characteristics [58]. However, they require the presence of this tooth and are only useful for estimating the age in the period of developing the third molar. On the other hand, the methods focused on predicting the numeric age or the binary sex directly from the raw images are very useful as they yield very accurate estimations. Also, they do not require the presence of specific teeth, so they are applicable in a wider range of age cohorts. Despite the solutions provided in some studies reflecting the attention behaviour of these methods [151, 222], the interpretability of these models is still far from that of the traditional manual and semiautomatic methods, which in the end can hinder their potential applicability.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.3 Materials and methods

The experiments detailed in this chapter were conducted in a subset of the OPG-AS database described in Chapter 3, which consisted of 1,746 images belonging to subjects aged from five to 60. Regarding the database demographics, the number of female and male images was almost the same for every age group, as seen in Table 6.3. To provide a good balance between developing and fully developed dentitions, a similar amount of subjects younger and older than 20 was included. The image annotation procedure was carried out by two trained and previously calibrated dental experts through the Labelbox platform [126], and consisted of the tracing of the oriented bounding box of each tooth, preserving its natural orientation.

Table 6.3: Dataset demographics.

Age (years)	**Male**	**Female**	**Total**
[5, 10)	140	150	290
[10, 15)	150	150	300
[15, 20)	146	149	295
[20, 25)	108	104	212
[25, 30)	45	45	90
[30, 35)	45	45	90
[35, 40)	49	47	96
[40, 45)	49	48	97
[45, 50)	49	49	98
[50, 55)	47	47	94
[55, 60)	44	40	84
Total	872	874	1,746

Due to the properties of the OPG geometry, the maxillary teeth show a greater overlap than the mandibular counterparts, especially in the premolar region, and the maxillary sinuses can hamper the visibility of the apices [98]. Moreover, the anterior teeth both in the maxilla and the mandible are usually affected by ghost shadows produced by the cervical spine [174]. This, together with the suitability of the posterior teeth for estimating the chronological age —as reflected in the heatmaps provided in Chapter 4 (Figure 4.9)— and the canines for determining

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

the sex [110], led us to select mandibular canines, premolars, and molars for conducting the experiments.

Fig. 6.2 shows the methodology proposed in this work for automatic age and sex estimation, which has two main parts: teeth detection and age and sex prediction. In the first step, the selected subset of teeth is detected in the image, obtaining the position of each tooth via oriented bounding boxes, and its number according to the *Fédération Dentaire Internationale* (FDI) notation [106]. In the second step, the image information retrieved by the CNN at 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 left part of Fig. 6.2 is based on the Faster R-CNN architecture [186] and so it is composed of three main modules: the feature extraction network —also called backbone—, the region proposal network (RPN), and the ROI head. Firstly, 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 containing a tooth, being the most probable tooth locations referred to as 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 classify it into one of the tooth classes or the background class.

This work includes some Faster R-CNN improvements proposed in posterior works. Specifically, the FPN was used as the feature extractor [136], and ROIAlign was used as the ROI pooler operation [95]. Also, some modules of the whole pipeline were modified to work with oriented bounding boxes. This includes the RPN, whose anchors were parameterised not only by their area and aspect ratio but also by their rotation, the ROIAlign operation, which needed to include the angle variable to cope with the oriented boxes, and the ROI head, which had to compute the refined positions of oriented boxes.

Once the teeth are localised and classified, the ROIAlign operation is applied again to evaluate the features extracted by the backbone at the final tooth positions. The resulting per-tooth features are then processed to obtain the chronological age and sex, as shown in the right part of Fig. 6.2. We will refer to the age and sex estimation modules as age head and sex head, respectively.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

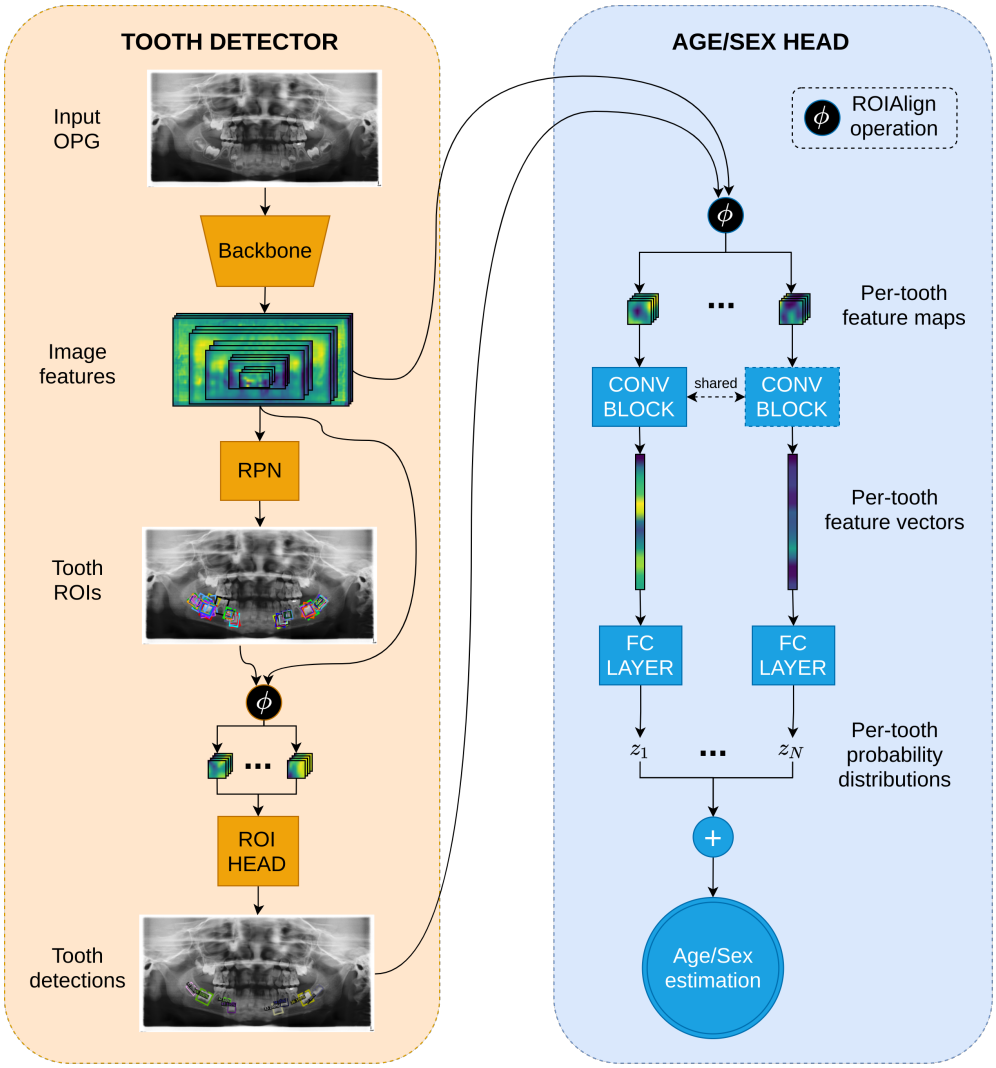


Figure 6.2: Proposed approach for age and sex estimation. First, the teeth are detected in the OPG with an R-CNN-based architecture. This includes the feature extraction process, through an FPN backbone, the proposal of teeth locations (ROIs), by an RPN, and the refinement and classification of each tooth ROI, with an ROI head. Second, the per-tooth probability distributions of age and sex are estimated. This process consists of the evaluation of the features extracted by the backbone at the estimated tooth locations, and the application of 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.

The architecture of the age head is composed firstly of a convolutional block, which is shared by all tooth classes, to extract per-tooth high-level features, which are expected to be associated with tooth development —and thus to the chronological age. The convolutional block consists of a first convolutional layer plus a max pooling operation, followed by a second convolutional layer plus an average pooling operation. In every case, the applied convolutions rely on a 3x3 kernel. The output of this block is then flattened into a 1-dimensional per-tooth-feature vector. As the dental development stages follow a different timing for each specific tooth, performing a tooth-wise estimation without taking into account the specific tooth type can be misleading. Thus, per-tooth fully connected layers are applied after the convolutional block. In this way, the development features can be summarised differently for each tooth type.

It is well known that some teeth are more relevant for performing chronological age estimation than others, and so the age estimates of those teeth should be taken into account to a greater extent. In this regard, the per-tooth fully connected layers produce not only a point estimation of age but a parameterised Gaussian distribution $\mathcal{N}_t(\mu_t, \sigma_t^2), t \in T$, being T the set of detected teeth. The Gaussian parameters are very interpretable in the case of age estimation, as they provide a measurement of centrality (mean) and a measurement of uncertainty (variance).

To constrain the variance parameter to the positive domain, Bishop [18] suggested the use of an exponential activation function. However, this function grows very fast and can lead to unstable behaviour, especially in datasets with high variance. Thus, we followed the approach proposed in [21] and used a modified version of the exponential linear unit (ELU) activation function (Equation (6.1)). As it can be seen, the ELU' function keeps the ELU behaviour —exponential growth for small values and linear growth for higher values—, and adds $1 + \epsilon$ to get strictly positive output.

$$ELU'(x) = ELU(x) + 1 + \epsilon$$
$$ELU(x) = \begin{cases} \alpha(e^x - 1) & \text{for} \quad x \leq 0 \\ x & \text{for} \quad x > 0 \end{cases} \tag{6.1}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The aggregation of the per-tooth age estimations is carried out by following the principle that the most uncertain estimation contributes to the final estimation to a lesser extent than the less uncertain estimation. Thus, the overall estimated age is calculated through

$$\begin{aligned} \hat{y} &= \sum_{t \in T} \mu_t \cdot w_t \\ w_t &= \frac{1/[(\sigma_t^2)^n]}{\sum_{t \in T} 1/[(\sigma_t^2)^n]} \end{aligned} \tag{6.2}$$

being T the set of detected teeth, μ_t and σ_t^2 the mean and variance of the distribution for tooth t, w_t the weight associated with the estimation of tooth t, and n a penalty term which gives even less importance to those per-tooth estimations with high variance. The higher the n parameter is, the closer the aggregated estimation will be to the most confident individual estimation.

Following the same idea applied in Chapter 4, the Grad-CAM method was used to enhance the interpretability of the chronological age estimations [197]. The main idea is to generate an attention heatmap indicating which regions of each tooth contribute to a greater extent to the age estimation by combining the mean and variance components of the estimated per-tooth age distributions. To do so, the heatmaps are calculated individually through Algorithm 1, which takes as input the estimated per-tooth age distributions, the per-tooth weight —calculated through Equation 6.2, with $n = 1$—, and the per-tooth activations of the last convolutional layer —as it is the last layer which provides spatial information—.

First, the gradient of μ_t with respect to the activations is obtained through backpropagation, which gives a measure of how each activation pixel contributes to increasing or decreasing the μ_t component. Second, the per-tooth activation tensor is weighted channelwise by the previously calculated gradient, averaged over the channel dimension. Third, the absolute value of the activations is taken —as the sign of their contributions is meaningless for this objective—, and the result is normalised to the (0,1) interval. Fourth, the per-tooth heatmap is calculated by taking the product between the per-tooth activation and the per-tooth weight. Finally, the alpha channel α_t of the heatmap is obtained with the aim of making the less relevant regions more transparent.

After per-tooth attention heatmaps are calculated, they need to be superimposed on the original OPG for a better understanding. Firstly, a colour map is applied to each heatmap to convert it from greyscale to RGB, where blue, yellow, and red regions correspond to irrelevant,

significant, and highly relevant contributions, respectively. Then, the alpha channel is added, conforming an RGB image, which is resized and translated to fit in the detected tooth bounding box.

Algorithm 1: Generation of the per-tooth attention heatmaps.

Input: Set of detected teeth: T
Input: Estimated per-tooth age distributions: $\mathcal{N}_t(\mu_t, \sigma_t^2), t \in T$
Input: Per-tooth estimation weight: w_t (Equation 6.2 with $n = 1$)
Input: Per-tooth activations of the last convolutional layer: $a_t^{(l)} \in \mathbb{R}^{C \times W \times H}$
Output: Per-tooth heatmaps: $H_t \in \mathbb{R}^{W \times H}$
Output: Per-tooth heatmap alphas: $\alpha_t \in \mathbb{R}^{W \times H}$

for $t \in T$ **do**
  $\nabla_t \leftarrow \frac{\partial \mu_t}{\partial a^{(l)}}$
  for $c = 1, \ldots, C$ **do**
    $a_t^{(l)}[c, :, :] \leftarrow a_t^{(l)}[c, :, :] \cdot \texttt{mean}(\nabla_t[c, :, :])$
  $a_t^{(l)} \leftarrow \frac{|a_t^{(l)}|}{\texttt{max}(|a_t^{(l)}|)}$
  $H_t \leftarrow H_t \cdot w_t$
  $\alpha_t \leftarrow \texttt{softmax}(H_t)$

The architecture of the sex head is almost identical to that of the age head. In this case, the convolutional block is aimed to catch the most sexually dimorphic features of the teeth. Again, the dimorphism has proven 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 the parameter p_t of a Bernoulli distribution such that p_t is the probability that the 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 give 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 through Equation (6.2).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.4 Experiments and performance evaluation

Although the tooth detector and the age and sex heads work together to perform age and sex estimations directly from a given image, the training process was split into two steps. First, the tooth detector was trained without including the age and sex heads, and the detection performance 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 through the Detectron2 platform [233].

6.4.1 Tooth detection

For the experiments related to the tooth detector, the initial dataset of 1,746 images was split into training, validation, and test subsets, with 60%, 20%, and 20% of the cases, respectively. The initial age and sex distribution of the dataset, provided in Table 6.3, was preserved across the three subsets. To increase the size and variability of the data, image augmentation techniques were applied, specifically horizontal flip, translation in both axes, rotation, and brightness and contrast disturbing.

Regarding 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 by 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 iterations. The learning rate was set initially to 0.01 and multiplied by a factor of 0.1 at iterations 4,000 and 7,000. At the end of the training step, the model state was reverted to the validation point where the loss function was minimum, and the performance metrics were obtained in the test set.

The main evaluation metric for the tooth detector was the mean average precision (mAP) [68]. This consists, firstly, of establishing a minimum overlap —intersection over union (IoU)— between the object ground truth and the prediction from which the detection is considered correct. Then, the Average Precision is used to summarise the precision/recall curve for each tooth class. Finally, the metric is averaged over all classes. Specifically, three different versions of the mAP metric were 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 ten different IoU values, from 0.5 to 0.95.

USC UNIVERSITAT DE VALÈNCIA

6.4.2 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, and as it was mentioned, there are specific age ranges where it is not feasible to estimate the chronological age or sex. Therefore, the experimentation regarding age estimation was focused on subjects younger than 25 (1,097 cases), while that related to sex estimation was focused on subjects older than 16 (1,096 cases).

The loss function used to train the age head combines the negative log-likelihood, which is 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 which compute the mean and variance components. The overall loss for the age head $\mathcal{L}_a$ is calculated through the following equation:

$$\mathcal{L}_a = \frac{\sum\limits_{t \in T} -\ln\left[\frac{1}{\sqrt{2\pi}\sigma_t} \quad e^{\frac{-(y-\mu_t)^2)}{2\sigma_t^2}}\right]}{|T|} + \lambda\Big(\sum W_l^2\Big) \tag{6.3}$$

where y is the real age, λ is the weight of the L2-loss, and W_l is the tensor of 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.

Age estimation performance was evaluated through regression metrics on the residuals —real age minus estimated age—, absolute errors —absolute value of the residuals—, and the explained variance. For the residuals, the median value is reported, as it has proven to be more robust for this problem [222]. For the absolute errors, mean, standard deviation, median, and interquartile range were reported. Moreover, the 95$^{\text{th}}$ percentile of the absolute errors was used as a measure of performance in the worst scenarios. Finally, the coefficient of determination R^2 was given.

To train the sex head, the formula for the BCE loss was used:

$$\mathcal{L}_s = \frac{\Sigma_{t \in T}\left[y \ln\Big(p_t(y)\Big) + \Big(1-y\Big)\ln\Big(1-p_t(y)\Big)\right]}{|T|} \tag{6.4}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

where y is the real sex of the subject belonging to the image —1 if it is female and 0 if it is male.

Sex estimation capabilities were assessed via traditional classification metrics, namely, the overall accuracy, and the precision, recall, and F1 metrics for each class —female and male.

The impact of the uncertainty-aware distribution aggregation was analysed both for age and sex heads. Specifically, three different aggregation policies were tested. The first one simply gives equal weights to each per-tooth estimation, i.e., $n = 0$ in Equation (6.2). In the second version, the per-tooth estimations are linearly weighted by their uncertainty, i.e., $n = 1$. In the third version, the highly uncertain estimations are heavily penalised by increasing the parameter n in Equation (6.2). The impact of the n value in the performance was also analysed.

6.5 Results

6.5.1 Tooth detection performance

The detection performance was reported in Table 6.4 along with other tooth detection networks available in the literature. The proposed approach outperformed the others in terms of the mAP@0.75 metrics (94.4 *vs.* 89.0 and 90.0), while it remained in the second position regarding the mAP@[0.5:0.95] and mAP@0.5 metrics.

Table 6.4: Comparison of systems for tooth detection in OPGs[(a)].

Model	mAP@[0.5:0.95]	mAP@0.5	mAP@0.75
Silva [202]	74.0	**99.7**	89.0
Chung [46]	**81.0**	91.0	90.0
Ours	77.6	96.4	**94.4**

(a) Only those methods reporting AP-based measurements were included.

The mAP@[0.5:0.95] metric was represented in Fig. 6.3. It can be seen that the highest detection accuracy was reached in teeth 46 and 36, corresponding to the first molars, with values of 81.9 and 82.6, respectively. The next teeth in terms of detection accuracy were 47 and 37, also known as the second molars, with values of 79.8 and 82.6. The detection performance decreased 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.

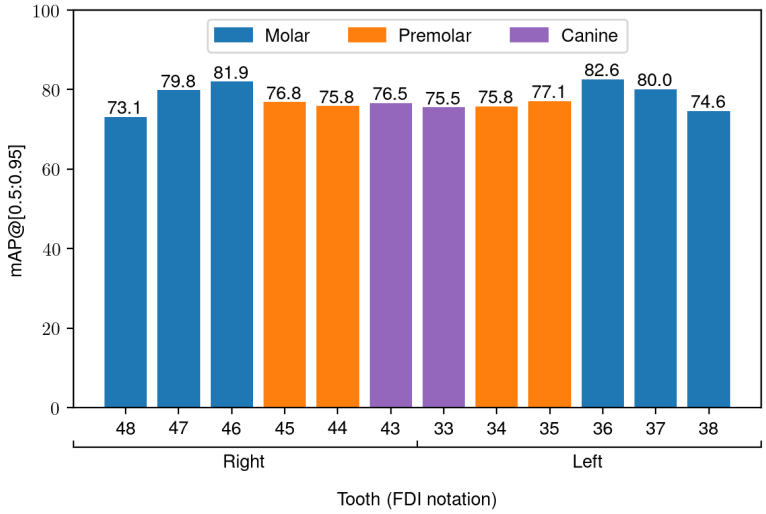
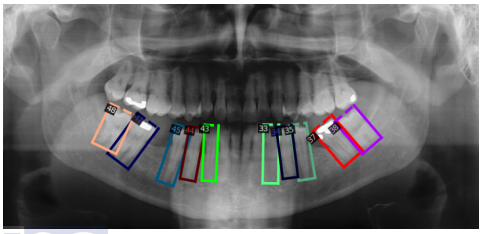
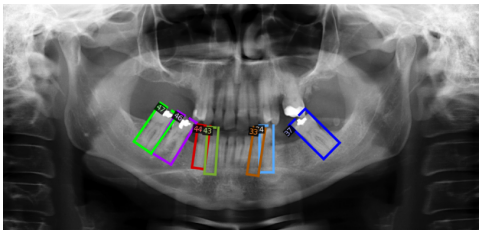


Figure 6.3: Detection performance for each tooth type. The first and second molars (46, 45, 45, and 36) were clearly the best detected teeth.

Qualitative results presented in Figure 6.4 provides an easy way to evaluate the behaviour of the proposed tooth detection system in both favourable and unfavourable scenarios. For example, the system correctly detected missing teeth, as shown in Figures 6.4a, 6.4b, and 6.4d. It also showed a robust behaviour in cases of dental implants (Figure 6.4e), and very accurate performance when dealing with images of children with mixed dentitions, as it can be seen in Figures 6.4f, 6.4g, and 6.4h. Moreover, it was able to detect teeth that are in a very early stage of development, as in the case of the left third molar in Figure 6.4h, and those that are more rotated than expected (Figure 6.4c).

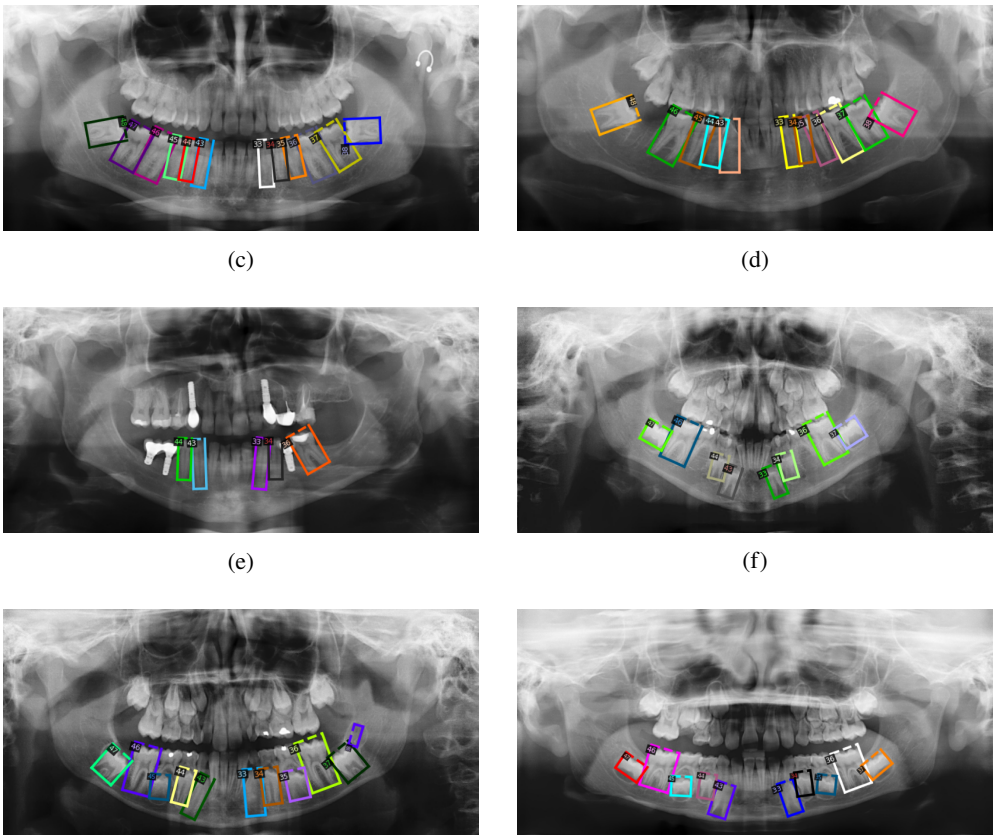


(a)



(b)

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(c) (d)

(e) (f)

(g) (h)

Figure 6.4: Qualitative results of the proposed tooth detection system.

6.5.2 Age estimation performance

The performance of age estimation regarding the different aggregation policies can be seen in Table 6.5 and Fig. 6.5 (blue color). In terms of age estimation results when the aggregation consists in averaging the means of the per-tooth distributions ($n = 0$ in Equation (6.2)), there was a noticeable systematic overestimation (the median E is -0.20 years), while the mean AE was 1.11 years. The 5% worst-estimations yielded an AE greater or equal than 3.29 years. The inclusion of a parameter that linearly penalises the uncertainty of the per-tooth estimations ($n = 1$) helped to slightly reduce the overestimation, being the median E -0.17 years. The

metrics based on the AE were also improved, being remarkable that the mean AE was below 1 year and the median AE decreased from 0.82 to 0.63 years. Also, the explained variance improved from 92% to 94%.

Table 6.5: 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**

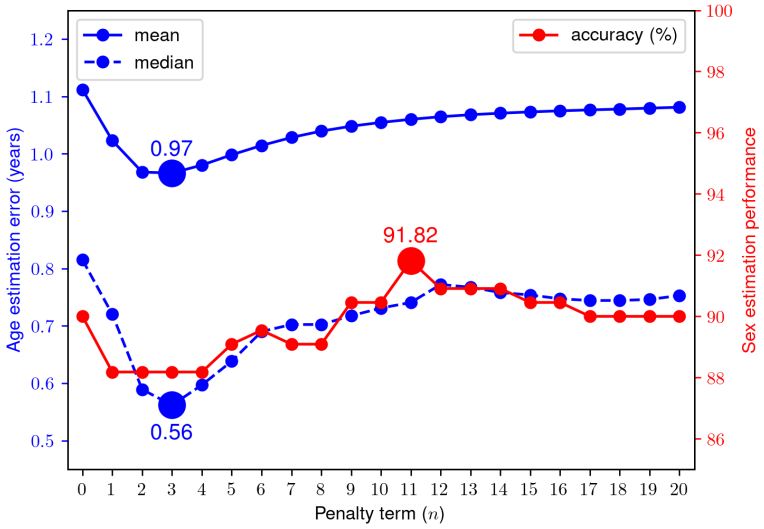


Figure 6.5: Impact of the penalty term n in the age and sex estimation performance (blue and red lines, respectively) for values greater or equal than 1. The best age estimations were obtained with $n = 3$, while the most accurate sex classification was reached with $n = 11$.

As shown in Fig. 6.5, when the n parameter is further increased —which leads to exponential uncertainty penalisations—, the age estimation errors decreased until they reach a minimum point at $n = 3$, and started to increase from there on. When the n value was set to that optimal point, the overestimation was reduced to almost zero, being the median E -0.06

years. Every other single metric was also improved, being particularly noticeable that the *AE* interquartile range decreased to 1.11 years and the 95th decreased to less than 3 years (2.88).

In Fig. 6.6, an example of the estimated per-tooth age distributions for a patient aged 18 years and 3 months was presented. In this case, the most anterior teeth —33 and 43 (canines) and 34, 44, and 45 (premolars)— yielded age distributions centred in a position far away from the real age. In addition, these distributions showed a high variance. As a counterpart, the most posterior teeth —37, 47, 38, and 48 (second and third molars)— produced distributions whose means were closer to the target age and whose variances were noticeably lower. The first aggregation policy ($n = 0$), which does not take into account the variance of the distributions, resulted in a poor estimation 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, as it relies on the widest distributions to a lesser degree than those tight distributions. Finally, the third aggregation policy ($n = 3$, according to the optimal value reported in Fig. 6.5), which penalises the uncertainty exponentially, got the final estimation even closer to the real age of the subject, as the difference regarding the weights given to the most and the less reliable estimations was significantly increased.

The performance of the proposed age estimation method was compared to other methods aimed at estimating automatically the numeric age in Table 6.6. As it can be seen, the method described in this chapter yielded better performance than any other, with improvements on the mean *AE* ranging 0.2 (*vs.* Vila-Blanco *et al.* [222]) to 1.31 years (*vs.* Čular *et al.* [52]). Furthermore, the distribution of the *AE* was tighter than those obtained in other studies according to its standard deviation. Finally, the overestimation of the proposed method is near zero, being comparable in absolute terms to the underestimation obtained in previous work [222].

Table 6.6: Comparison of systems for automatic age estimation.

Model	**Age range**	**E (med.)**	**AE ($\mu \pm \sigma$)**
Čular [52]	10-25	-	2.28 ± 2.17
De Back [55]	5-25	-	1.75 ± -
Vila-Blanco [222]	5-25	0.07	1.17 ± 1.11
Ours	5-25	**-0.06**	**0.97** ± **0.94**

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

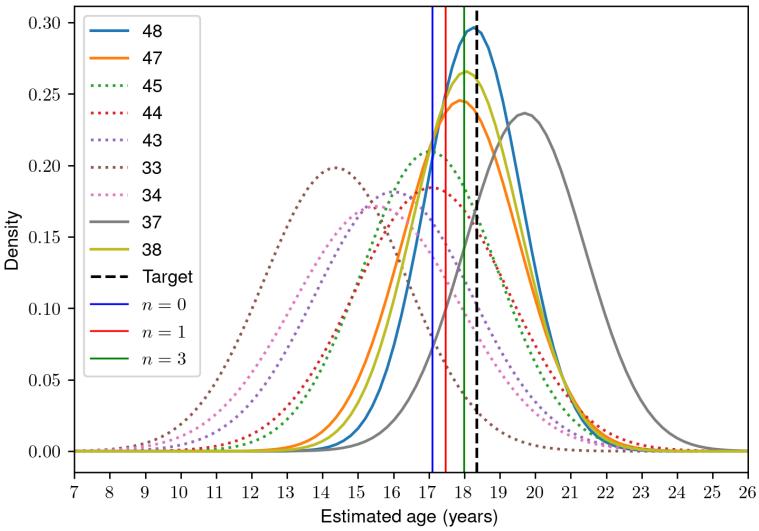
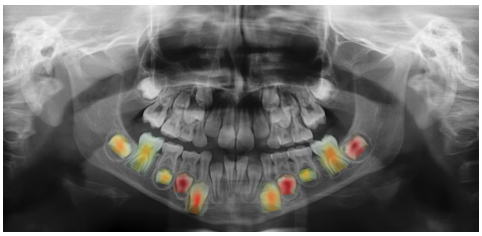


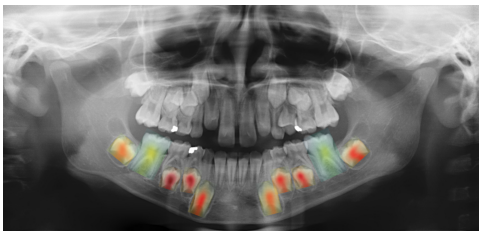
Figure 6.6: Example of the estimated per-tooth age distributions for a subject aged 18 years and 3 months, and the final prediction according to the different aggregation policies ($n \in \{0, 1, 3\}$). The solid and dotted per-tooth distributions correspond to the most and least confident teeth, respectively.

In Figure 6.7, the behaviour of the proposed system for highlighting the most relevant per-tooth regions for age estimation is represented through a set of examples. As mentioned before, the redder the regions are, the more relevant their contribution is. In the first three examples, corresponding to subjects aged 7, 8, and 11 years, respectively, the most relevant regions were located as a general rule over the first and third molars (36, 38, 46, and 48), and the second premolars (35 and 45). From then on, the development of most teeth is almost finished and thus their relevance starts to decrease. The most posterior teeth, and specifically the third molars (38 and 48) due to their later development ending, are clearly the most relevant teeth in adulthood. It is also noticeable that the most relevant regions within each tooth are mostly located at the pulp chamber, the radicular canals and the apical area.

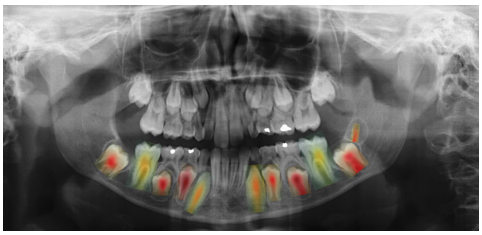
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



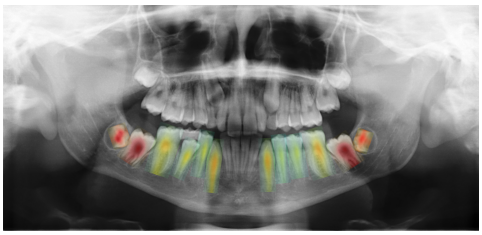
(a) 7 years and 7 months.



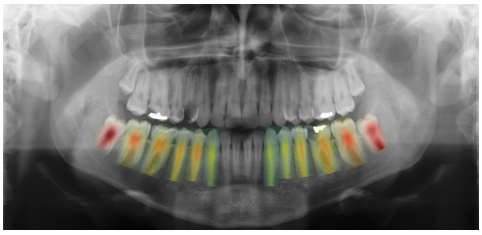
(b) 8 years and 4 months.



(c) 11 years and 8 months.



(d) 14 years and 4 months.



(e) 18 years and 6 months.

(f) 22 years and 8 months.

Figure 6.7: Attention heatmaps indicating the importance of tooth regions in the chronological age estimations.

6.5.3 Sex estimation performance

The sex estimation performance was reported in Table 6.7 and the red graph in Fig. 6.5. When the aggregation of the per-tooth distributions was done using a simple average of the estimated probabilities ($n = 0$), the accuracy was 90%. The recall for females was 94.54% and 85.45% for males, while the precision was 86.67% and 94%, respectively. In both classes, the F1 measure was around 90%. When a linear uncertainty penalisation was added ($n = 1$), the overall accuracy dropped to 88.18%. The female recall and male precision

USC UNIVERSITAT DE VALÈNCIA

decreased noticeably, while the female precision and the male recall increased to a lesser extent, resulting in a more balanced classification performance across both classes. The F1 measurements were noticeably lower, with a difference of about 2% with respect to the aggregation with no uncertainty ($n = 0$).

Table 6.7: 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**

The sex classification accuracy followed a different pattern than the age estimation performance when the value of n was increased. As shown in Fig. 6.5 (red line), every value under 9 yielded an accuracy below that obtained with the simplest aggregation policy. From $n = 9$ to $n = 14$ the results improved the baseline ($n = 0$), and then the performance started to drop. The highest accuracy was reached with $n = 11$, being 91.82%. With this optimal setup, the recall for the female class and the precision for the male class stayed very similar to those obtained without uncertainty-aware aggregation. However, the precision for the female class, the recall for the male class, and the F1 measure improved more than 2%.

An example of the aggregation of per-tooth sex distributions for a female subject is presented in Fig. 6.8. In this case, the best per-tooth estimations were produced for the canines (33 and 43), first premolars (34 and 44), and the right first molar (46), while the worst estimations were yielded for the second premolars (35 and 45) and the right third molar (48). The three aggregation policies described in Section 6.3 were evaluated, being the penalisation parameter n in the exponential penalisation policy set to 11 according to the optimal value reported in the red graph in Fig. 6.5. In every case, the returned probability classified correctly the sex of the subject using the threshold of 0.5. However, when the uncertain estimations —the ones that give 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.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

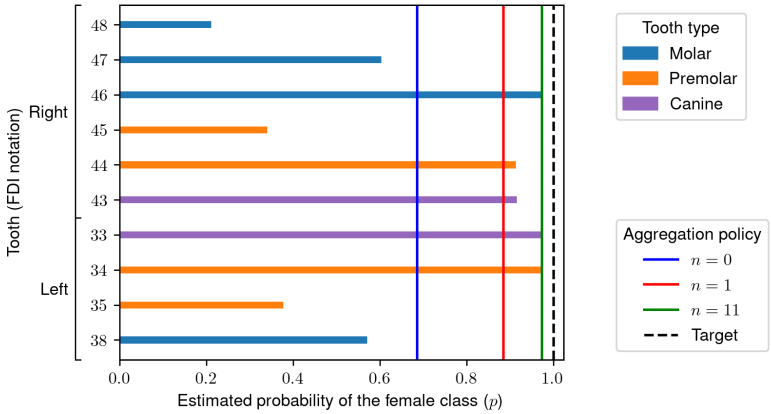


Figure 6.8: Example of the estimated per-tooth sex distributions for a female subject, and the final prediction according to the different aggregation policies ($n \in \{0, 1, 11\}$).

The results of the proposed sex estimation method were compared to the other automatic sex estimation approaches in Table 6.8. In general terms, the proposed method was outperformed by the other approaches regarding the obtained classification accuracy, being the differences between 2.5% (*vs.* Ilić *et al.* [100]) and 5.1% (*vs.* Milošević *et al.* [151]).

Table 6.8: Comparison of systems for automatic sex determination.

Model	**Age range**	**Accuracy (%)**
Ilić [100]	-	94.3
Milošević [151]	19-85	**96.9**
Ke [114]	16-70	94.6
Vila-Blanco [224]	5-90	90.6
Ours	16-60	91.8

6.6 Discussion and conclusions

The estimation of chronological age and biological sex is a crucial task in a number of clinical or forensic procedures, and it has been widely tackled by the research community [109, 229]. Nevertheless, the majority of studies have introduced manual or semiautomatic approaches,

which are very time-consuming and subjected to a high degree of subjectivity. In the last years, fully automatic methods have been proposed to process a dental X-ray image and establish the chronological age and sex of the subject [55, 151, 222, 224]. These studies reported high performance, with fast and confident estimations, suggesting that the subjectivity problems of manual processes can be greatly reduced. However, and despite the efforts to enhance the interpretability of these methods [222], they still lack explanation mechanisms to assess the reasons behind those estimations.

In this chapter, a two-step process has been proposed to perform chronological age and biological sex estimation from an OPG. In the first step, a CNN is used to detect a set of mandibular teeth, outputting the oriented bounding boxes where each tooth is located. The features learned by this network are evaluated at the located tooth boxes and fed into a second CNN, which applies a set of convolutional layers and a fully connected layer to produce a per-tooth age/sex estimation. 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. The per-tooth age/sex 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 inversely proportional to the distribution's variance. Also, a method to produce explanatory heatmaps on the contribution of each tooth region to the age estimation has been proposed to enhance the interpretability of the results.

Regarding the tooth detection subsystem, its performance was comparable to the other methods, outperforming them in the mAP@[0.5:0.95] metric and staying in the second position in the other two proposed metrics. This is quite remarkable since the proposed approach was evaluated in mixed dentitions (children and adults) and the other methods were tested only on fully developed adult dentitions. It is expected that the performance of the other methods would decrease to some extent if child subjects were included in their experimentation, as they contain very small objects —when tooth development is beginning— and great variability in terms of tooth aspect ratio —the teeth are wider when they start to develop, and they get higher as they grow.

In addition, and to the best of our knowledge, the proposed approach is the first attempt to detect the teeth through oriented bounding boxes. This kind of box fits the tooth better than the horizontally aligned boxes, as it avoids overlapping with the neighbour teeth. Thus, it enables to extract both the orientation of the teeth —which is crucial to detect some conditions such as impacted wisdom teeth— and some other measurements like their aspect ratio.

USC UNIVERSIDADE

Concerning the age and sex estimation subsystem, the proposed approach was evaluated according to three 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, and the weighted sum of the distribution means through an exponential penalisation of the uncertainty.

The age estimation experiments were performed in subjects younger than 25, as it is reportedly inaccurate in older subjects due to the fully developed teeth [180]. The results confirmed that the uncertainty-aware aggregation policies increased the predictive capabilities of the model and reduced the error 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 policy and 0.06 years (0.72 months) by the third policy. Similarly, the AE-based metrics were also improved, remarking the 0.97 years for the mean AE and the capability of the model to catch 94% of the data variability.

These results outperformed those obtained by other automatic age estimation systems, both in terms of under- or overestimation and absolute errors. This is very remarkable as the proposed approach only uses image information retrieved inside the tooth bounding boxes, as opposed to the works by De Back *et al.* [55] and Vila-Blanco *et al.* [222], which used the whole OPG to produce age estimations. This suggests that the teeth are the bones in the oral cavity which best correlate to the chronological age, at least in OPGs.

The proposed system for generating attention heatmaps supposed an important step forward with respect to those proposed in Chapter 4, as they are focused specifically on the teeth and provide a higher resolution. This allows fine-grained studies not only on the relevance of each tooth but also on the tooth parts that contribute to a greater extent. The examples presented demonstrated that the proposed method clearly agrees with the clinical findings reported in the literature. The developing teeth were obviously the more relevant ones, whereas the posterior teeth were proven to yield the most confident estimations in the older age cohorts, in line with the findings reported in Chapter 4. Similarly, the most influential tooth parts represented in the attention maps show a significant clinical coherence. The pulp chamber stands out as the most important tooth region —especially in fully developed teeth—, agreeing with the traditional methodologies relying on this area [25]. The apical area has also proven to be relevant in most developing teeth, supporting the previous works of Cameriere *et al.* in this regard [27]. Finally, the tooth enamel presents a negligible importance, as its formation —especially at the occlusal surface— ends very early [60].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The sex estimation results were assessed in subjects older than 16, as puberty is the point at which dimorphic skeletal indicators start to become noticeable [44]. The performance according to the same three 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 detection performance was reduced. With the optimal n value ($n = 11$), the classification performance was higher than that achieved with both the first and second aggregation policies in every single metric, with an overall accuracy of almost 92%.

The top achieved accuracy was remarkable, which implies that the teeth have a great potential to detect the biological sex. However, the results are still behind other automatic methods for sex estimation from OPGs [100, 114, 151, 224], suggesting that there are other structures in the OPG that are highly dimorphic. This is in line with other studies reporting a higher level of dimorphism in the mandible than in the teeth [79].

The visual representations of the predicted per-tooth distributions have proven to be consistent with the findings reported in the clinical literature. As an example, the most confident age estimations for the example presented in Fig. 6.6 —a subject aged 18 years and 3 months— were the third molars (38 and 48), which are expected to be the only teeth in development at that age [44]. In Fig. 6.8 the canines (33 and 43) yielded the most confident sex estimations, followed by the first premolars (34 and 44), which is in line with the study by Zorba *et al.* [237].

Overall, the proposed age and sex estimation models have some advantages over the previous automatic approaches. First, they do not require the presence of any specific tooth to work, as opposed to other methods [13, 52, 58, 116, 149], ensuring its applicability in a wider range of scenarios. Although the model could produce age and sex estimations with a single tooth, the performance is expected to increase as multiple teeth are available. Moreover, the age estimation method outputs a numeric value, which gives a more precise output than the N-stage classification system [13, 58, 115, 149].

The explainability of the age and sex estimations is a key aspect of the proposed approach, as they can be easily traced back to build understandable explanations. In addition to the above-mentioned attention heatmaps for the age estimation problems, the fact that an individual probability distribution is produced for each detected tooth gives the experts a powerful tool to conduct further studies regarding the teeth that are contributing to the age and sex estimations to a greater extent, those that are more suitable for a specific age cohort, the behaviour of

USC UNIVERSIDADE

the proposed approach when the most relevant teeth are missing, or the minimum set of teeth yielding the optimal performance. Besides, the age estimation system could also be used to detect development problems for specific teeth, such as advanced or delayed growth.

In conclusion, the proposed age and sex estimation method fulfils objective 3 of this doctoral thesis, as it is useful to estimate the chronological age and the biological sex accurately from OPGs in a variety of scenarios, enabling an interpretable per-tooth estimation analysis.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 7

CONCLUSIONS

In this PhD Thesis, three different automatic approaches for predicting chronological age and sex from dental panoramic radiographs have been proposed. All these methods are heavily based on deep learning techniques, and more specifically on convolutional neural networks (CNN), which have proven to be extremely useful to automate medical imaging processing. The performance evaluation was carried out on different subsets of the OPG-AS database provided by the School of Medicine and Dentistry, Universidade de Santiago de Compostela (Spain), composed of dental panoramic radiographs belonging to Caucasian subjects. In no case an image exclusion protocol was defined, creating a test environment that was as realistic as possible.

In the first contribution, two one-stage CNN architectures for chronological age estimation were proposed. The first one —Dental Age Network (DANet)— consists of a single-path CNN oriented to predict the age by extracting development-related features in the image. The second architecture —Dental Age and Sex Network (DASNet)— adds a secondary path aimed at extracting the most sexually-dimorphic features and propagating them at intermediate points of the main path to reduce the age estimation error. Both architectures yielded state-of-the-art performance, especially in the younger age groups, with mean absolute errors around eight months. However, DASNet provided more confident estimations, which confirms the clinical hypothesis that the development of skeletal structures follows different timing patterns in males and females, and thus this information should be taken into account to develop more accurate age estimation models. Visual representations of the OPG regions on which DASNet relies most heavily for age estimation were also obtained to improve the interpretability of the

USC UNIVERSIDADE

method. Moreover, the DASNet architecture was modified for the problem of sex estimation by reversing the roles of both paths, that is, making the sex path the main path and the age path the secondary path. The results in this regard confirmed that the sex classification is more accurate in adults but also good enough in children older than eight.

In the second contribution, a fully automatic pipeline for the extraction of the mandible contour in an OPG, the characterisation of its shape, and the use of these shape-related features for the prediction of both age and sex were described. The first step was carried out through the application of a two-stage CNN oriented to regress the heatmaps of each contour point; in the second step, four different descriptors are proposed to summarise both the shape and the size of the mandible; in the third step, the obtained mandible features are used to feed a set of highly interpretable regression and classification models for age and sex prediction, respectively. The proposed CNN architecture for mandible contour detection was compared with the landmarks digitised by two trained dentists, resulting in a lower error and better confidence for the former. Furthermore, the optimal configurations of mandible descriptors yielded a better accuracy compared to traditional methods for mandible ageing and sexing. The proposed methodology also allows providing visual representations of mandible shape changes, which were in accordance with those reported in the clinical literature.

In the third contribution, an automatic two-component method for predicting chronological age and sex based on tooth appearance was also proposed. The first component addresses tooth detection through a CNN architecture aimed at obtaining oriented tooth bounding boxes instead of the traditional horizontal boxes. It reaches state-of-the-art performance, with the benefit that it allows minimising the overlap between neighbouring teeth and collecting accurate information regarding tooth dimensions and orientation. The second component reuses the image features extracted by the tooth detector at each tooth position to estimate a per-tooth probability distribution for age and sex, respectively. A method for aggregating these per-tooth estimations by penalising the most uncertain ones was proven to improve the previously proposed methods for automatic age estimation from teeth, yielding for the first time a mean absolute error of less than one year for subjects under 25. Overall, the aggregation-based methodology allows reasoning some important questions, such as knowing which tooth is more important for each specific task, or which one shows a greater development variability. In addition, a method to generate per-tooth attention heatmaps for representing the tooth regions that contribute the most to the age estimation was proposed. The examples shown demonstrate that these representations are in line with the clinical literature regarding not only

USC UNIVERSIDADE

the most relevant teeth for each age cohort, but also the most significant regions within each tooth.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 8

FUTURE WORK

The methods developed in this PhD Thesis are useful tools for a variety of experts, like general and paediatric dentists, forensic clinicians, or anthropologists, among others. On the one hand, they offer high-performance solutions for chronological age and sex prediction. On the other hand, they provide mechanisms for improving the explainability of the results, which is crucial for every potential application. However, there is still room for improvement both in technical and clinical terms to enhance the applicability in real scenarios. In the following, some of these potential enhancements have been identified.

An initial research direction to improve the age and sex prediction performance is the combination of development- and sexual-related image features —which led to a noticeable performance improvement in Chapter 4— to the aggregation-based system described in Chapter 6. The main idea is to keep the architecture of the tooth detector as a mechanism to extract low- and mid-level features, and perform the high-level feature integration between the age and sex heads. Two different approaches, and the combination of the two, will be evaluated. In the first one, the integration will be performed in the convolutional blocks, in a similar way as described in Chapter 4. Thus, if the maximum age estimation performance is required, the per-tooth feature maps from the sex head will be propagated to the age path and concatenated at the beginning of each convolutional block. Reversely, the per-tooth feature maps from the age head will be integrated into the sex path if the objective is the maximisation of the sex prediction accuracy. In the second approach, the integration will be carried out once the per-tooth age and sex estimated distributions are obtained. The idea is to add fully connected layers after each tooth prediction, so each layer will get three inputs: the mean and variance of

the estimated per-tooth age distribution, and the p parameter of the per-tooth sex distribution. The output will correspond to the refined age and sex distribution parameters. After that, the aggregation policy described in Chapter 6 will be applied.

We also believe that the two-stage probabilistic approach described in Chapter 6 could be adapted to use additional information from the mandible shape and other relevant skeletal structures observable in a dental panoramic radiograph. This would entail the development of more object detectors and/or landmark localisation systems —on top of the mandible and tooth detectors proposed in Chapters 5 and 6, respectively. Moreover, it would be interesting to hierarchise the proposed aggregation system. The idea is to define different levels among the detected skeletal structures and use this hierarchy to enrich the prediction explanations. For example, the first level will include items such as the dentition, the mandible, the maxillary sinuses, the nasal fossa, etc. The dentition could be subdivided into the different teeth, whereas the mandible contour could be divided into segments, such as the ramus, inferior border, condylae, etc. This sort of tree structure does not necessarily have to be full, that is, each skeletal structure will reach different depth levels, as many as required. Two main benefits would be obtained with this system. On the one hand, the prediction performance will be expectedly higher, as more information will be used. On the other hand, the hierarchical reasoning would allow knowing not only the contribution of each detected structure to the final prediction, but also the contribution of their individual parts, which is relevant from a clinical point of view.

Another interesting continuation of this PhD Thesis has to do with the applicability of the proposed methodologies to other populations. As the performance of our age and sex prediction models —trained on Caucasian samples— will be expectedly lower when assessing it in other samples —i.e., African, Asian, ...— due to the different skeletal development patterns, there will be a need to adapt it to reach the highest possible performance in each case. However, and to the best of our knowledge, the different development milestones or sexually dimorphic features are the same for every population, being the main differences related to the timing when those features can be observed. In terms of image processing, we can affirm that most of the image features are common to all populations, and only the high-level connections between those features vary from one population to another. Thus, the process of adapting our method to other populations will consist mainly of the retraining of the last layers of the proposed network on each specific database. As the number of parameters

USC UNIVERSIDADE

of these layers is small compared to the whole system, a relatively small database should be enough for performing those retrainings.

A step would also be the curation and publication of the database provided to us to validate the proposed methods. The legal and privacy requirements associated with the collection of medical records lead to a scarcity of databases in this regard, which hinders the development and evaluation of automatic methods. The publication of this database would have two main advantages. On the one hand, it would entail the biggest population-specific public OPG database, and the only one labelled with the age and sex of every subject, which would allow assessing the performance of other age and sex prediction methodologies. On the other hand, the fact that the database includes images with a high variety of conditions makes it ideal to use it for developing automatic diagnostic methods.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Bibliography

[1] A.H. Abdi, S. Kasaei, and M. Mehdizadeh. Automatic segmentation of mandible in panoramic X-ray. *Journal of Medical Imaging*, 2(4):044003, 2015.

[2] D. Abualhija, G. Revie, and S. Manica. Mandibular ramus as a sex predictor in adult Jordanian subjects. *Forensic Imaging*, 21:200366, 2020.

[3] A. B. Acharya and S. Mainali. Limitations of the mandibular canine index in sex assessment. *Journal of Forensic and Legal Medicine*, 16(2):67–69, 2009.

[4] A.B. Acharya and S. Mainali. Sex discrimination potential of buccolingual and mesiodistal tooth dimensions. *Journal of Forensic Sciences*, 53(4):790–792, 2008.

[5] A. Alias, A.N. Ibrahim, S.N.A. Bakar, M.S. Shafie, S. Das, N. Abdullah, H.M. Noor, I.Y. Liao, and F.M. Nor. Anthropometric analysis of mandible: an important step for sex determination. *La Clinica Terapeutica*, 169(5):e217–e223, 2018.

[6] S.J. AlQahtani, M.P. Hector, and H.M. Liversidge. Brief communication: the London atlas of human tooth development and eruption. *American Journal of Physical Anthropology*, 142(3):481–490, 2010.

[7] A. Alqerban, M. Hedesiu, M. Baciut, O. Nackaerts, R. Jacobs, S. Fieuws, SedentexCT Consortium, and G. Willems. Pre-surgical treatment planning of maxillary canine impactions using panoramic vs cone beam CT imaging. *Dentomaxillofacial Radiology*, 42(9):20130157, 2013.

[8] V. Ambarkova, I. Galić, M. Vodanović, D. Biočina-Lukenda, and H. Brkić. Dental age estimation using Demirjian and Willems methods: cross sectional study on children

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

from the Former Yugoslav Republic of Macedonia. *Forensic Science International*, 234:187–e1, 2014.

[9] J. Ameijeiras-Alonso, R.M. Crujeiras, and A. Rodríguez-Casal. Mode testing, critical bandwidth and excess mass. *Test*, 28(3):900–919, 2019.

[10] D.L. Anderson and G.W. Thompson. Interrelationships and sex differences of dental and skeletal measurements. *Journal of Dental Research*, 52(3):431–438, 1973.

[11] P.J. Atkinson and C. Woodhead. Changes in human mandibular structure with age. *Archives of Oral Biology*, 13(12):1453–IN16, 1968.

[12] D.H. Badran, D.A. Othman, H.W. Thnaibat, and W.M. Amin. Predictive accuracy of mandibular ramus flexure as a morphologic indicator of sex dimorphism in Jordanians. *International Journal of Morphology*, 33(4), 2015.

[13] N. Banar, J. Bertels, F. Laurent, R. Merdietio Boedi, J. De Tobel, P. Thevissen, and D. Vandermeulen. Towards fully automated third molar development staging in panoramic radiographs. *International Journal of Legal Medicine*, pages 1–11, 2020.

[14] I. Bankman. *Handbook of medical image processing and analysis*. Elsevier, 2008.

[15] D.A. Baur, M.A. Altay, A. Flores-Hidalgo, Y. Ort, and F.A. Quereshy. Chronic osteomyelitis of the mandible: diagnosis and management—an institution's experience over 7 years. *Journal of Oral and Maxillofacial Surgery*, 73(4):655–665, 2015.

[16] R. Behling. *Modern diagnostic X-ray sources: technology, manufacturing, reliability*. CRC Press, 2021.

[17] M.E. Berkvens, S.I. Fairgrieve, and S. Keenan. A comparison of techniques in age estimation using the third molar. *Canadian Society of Forensic Science Journal*, 50(2):74–83, 2017.

[18] C.M. Bishop. Mixture density networks. 1994.

[19] M.A.V. Bittencourt, G.O. Cericato, A. Franco, R. Silva-Girão, A.P.B. 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 Radiology*, 47(4):20170362, 2018.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[20] M.V. Bolanos, M.C. Manrique, M.J. Bolanos, and M.T. Briones. Approaches to chronological age assessment based on dental calcification. *Forensic Science International*, 110(2):97–106, 2000.

[21] A. Brando-Guillaumes. Mixture density networks for distribution and uncertainty estimation. Master's thesis, Universitat Politècnica de Catalunya, 2017.

[22] P.H. Buschang, A. Santos-Pinto, and A. Demirjian. Incremental growth charts for condylar growth between 6 and 16 years of age. *The European Journal of Orthodontics*, 21(2):167–173, 1999.

[23] A.C. Butti, A. Clivio, M. Ferraroni, E. Spada, A. Testa, and A. Salvato. Häävikko's method to assess dental age in Italian children. *The European Journal of Orthodontics*, 31(2):150–155, 2008.

[24] 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. *International Journal of Legal Medicine*, 121(6):449–453, 2007.

[25] R. Cameriere, L. Ferrante, M. G. Belcastro, B. Bonfiglioli, E. Rastelli, and M. Cingolani. Age estimation by pulp/tooth ratio in canines by mesial and vestibular peri-apical X-rays. *Journal of Forensic Sciences*, 52(5):1151–1155, 2007.

[26] R. Cameriere, L. Ferrante, and M. Cingolani. Variations in pulp/tooth area ratio as an indicator of age: a preliminary study. *Journal of Forensic Science*, 49(2):1–3, 2004.

[27] R. Cameriere, L. Ferrante, and M. Cingolani. Age estimation in children by measurement of open apices in teeth. *International Journal of Legal Medicine*, 120(1):49–52, 2006.

[28] 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. *International Journal of Legal Medicine*, 122(6):493–497, 2008.

[29] R. Cameriere, L. Ferrante, H.M. Liversidge, J.L. Prieto, and H. Brkic. Accuracy of age estimation in children using radiograph of developing teeth. *Forensic Science International*, 176(2-3):173–177, 2008.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[30] C. Capitaneanu, G. Willems, R. Jacobs, S. Fieuws, and P. Thevissen. Sex estimation based on tooth measurements using panoramic radiographs. *International Journal of Legal Medicine*, 131(3):813–821, 2017.

[31] C. Capitaneanu, G. Willems, and P. Thevissen. A systematic review of odontological sex estimation methods. *The Journal of Forensic Odonto-Stomatology*, 35(2):1, 2017.

[32] A. Cardini. Less tautology, more biology? a comment on "high-density" morphometrics. *Zoomorphology*, 139(4):513–529, 2020.

[33] I. Castiglioni, L. Rundo, M. Codari, G. Di Leo, C. Salvatore, M. Interlenghi, F. Gallivanone, A. Cozzi, N.C. D'Amico, and F. Sardanelli. AI applications to medical images: from machine learning to deep learning. *Physica Medica*, 83:9–24, 2021.

[34] P. Cattell. *Dentition as a measure of maturity*. Number 9-12. Harvard University Press, 1928.

[35] M. Celikoglu, K. Cantekin, and I. Ceylan. Dental age assessment: the applicability of Demirjian method in eastern Turkish children. *Journal of Forensic Sciences*, 56(s1), 2011.

[36] N. Chaillet, M. Nyström, and A. Demirjian. Comparison of dental maturity in children of different ethnic origins: international maturity curves for clinicians. *Journal of Forensic Science*, 50(5):JFS2005020–11, 2005.

[37] N. Chaillet, M. Nyström, M. Kataja, and A. Demirjian. Dental maturity curves in Finnish children: Demirjian's method revisited and polynomial functions for age estimation. *Journal of Forensic Science*, 49(6):JFS2004211–8, 2004.

[38] A. Chandra, A. Singh, M. Badni, R. Jaiswal, and A. Agnihotri. Determination of sex by radiographic analysis of mental foramen in North Indian population. *Journal of Forensic Dental Sciences*, 5(1):52, 2013.

[39] P. Chaudhuri and J.S. Marron. SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, 94(447):807–823, 1999.

[40] S.Y.Y. Chen, P.E. Lestrel, W.J.S. Kerr, and J.H. McColl. Describing shape changes in the human mandible using elliptical Fourier functions. *European Journal of Orthodontics*, 22(3):205–216, 2000.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[41] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun. Cascaded pyramid network for multi-person pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7103–7112, 2018.

[42] F. Chollet et al. Keras. `https://keras.io`, 2015.

[43] M.E. Chovalopoulou, E.D. Valakos, and S.K. Manolis. Sex determination by three-dimensional geometric morphometrics of the palate and cranial base. *Anthropologischer Anzeiger*, 70(4):407–25, 2013.

[44] A.M. Christensen, N.V. Passalacqua, and E.J. Bartelink. Chapter 10 - Age estimation. In A.M. Christensen, N.V. Passalacqua, and E.J. Bartelink, editors, *Forensic anthropology (second edition)*, pages 307–349. Academic Press, second edition edition, 2019.

[45] P. Chu, C. Bo, X. Liang, J. Yang, V. Megalooikonomou, F. Yang, B. Huang, X. Li, and H. Ling. Using octuplet siamese network for osteoporosis analysis on dental panoramic radiographs. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2579–2582. IEEE, 2018.

[46] M. Chung, J. Lee, S. Park, M. Lee, C.E. Lee, J. Lee, and Y.G. Shin. Individual tooth detection and identification from dental panoramic x-ray images via point-wise localization and distance regularization. *Artificial Intelligence in Medicine*, 111:101996, 2021.

[47] D. Cipriani. *Children's rights and the minimum age of criminal responsibility: a global perspective*. Routledge, 2016.

[48] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[49] M. Coquerelle, F.L. Bookstein, J. Braga, D.J. Halazonetis, G.W. Weber, and P. Mitteroecker. Sexual dimorphism of the human mandible and its association with dental development. *American Journal of Physical Anthropology*, 145(2):192–202, 2011.

[50] E.F. Corbet, D.K.L. Ho, and S.M.L. Lai. Radiographs in periodontal disease diagnosis and management. *Australian Dental Journal*, 54:S27–S43, 2009.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[51] 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. *International Journal of Legal Medicine*, 124(2):105–112, 2010.

[52] L. Čular, M. Tomaić, M. Subašić, T. Šarić, V. Sajković, and 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*, pages 25–30. IEEE, 2017.

[53] M.R. Dayal, M.A. Spocter, and M.A. Bidmos. An assessment of sex using the skull of black South Africans by discriminant function analysis. *Homo*, 59(3):209–221, 2008.

[54] D. De Angelis, D. Gibelli, D. Gaudio, F.C. Noce, N. Guercini, G. Varvara, E. Sguazza, C. Sforza, and C. Cattaneo. Sexual dimorphism of canine volume: a pilot study. *Legal Medicine*, 17(3):163–166, 2015.

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

[56] H.H. de Boer, S. Blau, T. Delabarde, and L. Hackman. The role of forensic anthropology in disaster victim identification (DVI): recent developments and future prospects. *Forensic Sciences Research*, 4(4):303–315, 2019.

[57] F.T. de Oliveira, M.Q.S. Soares, V.A. Sarmento, C.M.F. Rubira, J.R.P. Lauris, and I R.F. Rubira-Bullen. Mandibular ramus length as an indicator of chronological age and sex. *International Journal of Legal Medicine*, 129(1):195–201, 2014.

[58] J. De Tobel, P. Radesh, D. Vandermeulen, and P.W. Thevissen. An automated technique to stage lower third molar development on panoramic radiographs for age estimation: a pilot study. *The Journal of Forensic Odonto-Stomatology*, 35(2):42, 2017.

[59] A. Demirjian and H. Goldstein. New systems for dental maturity based on seven and four teeth. *Annals of Human Biology*, 3(5):411–421, 1976.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[60] A. Demirjian, H. Goldstein, and J.M. Tanner. A new system of dental age assessment. *Human Biology*, pages 211–227, 1973.

[61] H. Devlin, K. Karayianni, A. Mitsea, R. Jacobs, C. Lindh, P. van der Stelt, E. Marjanovic, J. Adams, S. Pavitt, and K. Horner. Diagnosing osteoporosis by using dental panoramic radiographs: the OSTEODENT project. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 104(6):821–828, 2007.

[62] M. Dhillon, S.M. Raju, S. Verma, D. Tomar, R.S. Mohan, M. Lakhanpal, and B. Krishnamoorthy. Positioning errors and quality assessment in panoramic radiography. *Imaging Science in Dentistry*, 42(4):207–212, 2012.

[63] P. Diz, J. Limeres, A.F.P. Salgado, I. Tomás, L.F. Delgado, E. Vázquez, and J.F. Feijoo. Correlation between dental maturation and chronological age in patients with cerebral palsy, mental retardation, and Down syndrome. *Research in Developmental Disabilities*, 32(2):808–817, 2011.

[64] S. Dostalova, K. Sonka, Z. Smahel, V. Weiss, J. Marek, and D. Horinek. Craniofacial abnormalities and their relevance for sleep apnoea syndrome aetiopathogenesis in acromegaly. *European Journal of Endocrinology*, 144(5):491–497, 2001.

[65] 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 *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2615–2618. IEEE, 2018.

[66] T. Ekert, J. Krois, L. Meinhold, K. Elhennawy, R. Emara, T. Golla, and F. Schwendicke. Deep learning for the radiographic detection of apical lesions. *Journal of Endodontics*, 45(7):917–922, 2019.

[67] S. Ellingham and J. Adserias-Garriga. Complexities and considerations of human age estimation. In *Age estimation*, pages 1–15. Elsevier, 2019.

[68] M. Everingham, L. Van-Gool, C.K. I. Williams, J. Winn, and A. Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2):303–338, 2010.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[69] J.P. Fancher and P. Hampl. Missing and unidentified persons. In *Forensic Odontology*, pages 103–120. Elsevier, 2018.

[70] C.S. Farah, D.R. Booth, and S.C. Knott. Dental maturity of children in Perth, Western Australia, and its application in forensic age estimation. *Journal of Clinical Forensic Medicine*, 6(1):14–18, 1999.

[71] M.F.N. Feres, T.S. Muniz, S.H. de Andrade, M.M. Lemos, and S. Shizue N. Pignatari. Craniofacial skeletal pattern: is it really correlated with the degree of adenoid obstruction? *Dental Press Journal of Orthodontics*, 20:68–75, 2015.

[72] D. Franklin. Forensic age estimation in human skeletal remains: current concepts and future directions. *Legal Medicine*, 12(1):1–7, 2010.

[73] D. Franklin and A. Cardini. Mandibular morphology as an indicator of human subadult age: interlandmark approaches. *Journal of Forensic Sciences*, 52(5):1015–1019, 2007.

[74] D. Franklin, A. Cardini, P. O'Higgins, C.E. Oxnard, and I. Dadour. Mandibular morphology as an indicator of human subadult age: geometric morphometric approaches. *Forensic Science, Medicine and Pathology*, 4(2):91–99, 2007.

[75] D. Franklin, A. Flavel, J. Noble, L. Swift, and S. Karkhanis. Forensic age estimation in living individuals: methodological considerations in the context of medico-legal practice. *Research and Reports in Forensic Medical Science*, 5:53–66, 2015.

[76] D. Franklin, P. O'Higgins, and C.E. Oxnard. Sexual dimorphism in the mandible of indigenous South Africans: a geometric morphometric approach. *South African Journal of Science*, 104(3-4):101–106, 2008.

[77] D. Franklin, P. O'Higgins, C.E. Oxnard, and I. Dadour. Determination of sex in South African blacks by discriminant function analysis of mandibular linear dimensions. *Forensic Science, Medicine, and Pathology*, 2(4):263–268, 2006.

[78] D. Franklin, C.E. Oxnard, P. O'Higgins, and I. Dadour. Sexual dimorphism in the subadult mandible: quantification using geometric morphometrics. *Journal of Forensic Sciences*, 52(1):6–10, 2007.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[79] D. Franklin, P. O'Higgins, C.E. Oxnard, and I. Dadour. Sexual dimorphism and population variation in the adult mandible. *Forensic Science, Medicine and Pathology*, 3(1):15–22, 2007.

[80] D. Franklin, P. O'Higgins, C.E. Oxnard, and I. Dadour. Discriminant function sexing of the mandible of indigenous South Africans. *Forensic Science International*, 179(1):84–e1, 2008.

[81] M. Fukuda, K. Inamoto, N. Shibata, Y. Ariji, Y. Yanashita, S. Kutsuna, K. Nakata, A. Katsumata, H. Fujita, and E. Ariji. Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography. *Oral Radiology*, 36(4):337–343, 2020.

[82] I. Galić, M. Vodanović, R. Cameriere, E. Nakaš, E. Galić, E. Selimović, and H. Brkić. Accuracy of Cameriere, Haavikko, and Willems radiographic methods on age estimation on Bosnian–Herzegovian children age groups 6–13. *International Journal of Legal Medicine*, 125(2):315–321, 2011.

[83] S.M. Garn, A.B. Lewis, and S.L. Polacheck. Variability of tooth formation. *Journal of Dental Research*, 38(1):135–148, 1959.

[84] V.M. Georgieva, A.D. Mihaylova, and P.P. Petrov. An application of dental X-ray image enhancement. In *2017 13th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS)*, pages 447–450. IEEE, 2017.

[85] R. Ghafari, A. Ghodousi, and E. Poordavar. Comparison of the accuracy of the London atlas and Smith method in dental age estimation in 5–15.99-year-old Iranians using the panoramic view. *International Journal of Legal Medicine*, 133(1):189–195, 2019.

[86] E. Giles. Sex determination by discriminant function analysis of the mandible. *American Journal of Physical Anthropology*, 22(2):129–135, 1964.

[87] I. Gleiser and E.E. Hunt. The permanent mandibular first molar: its calcification, eruption and decay. *American Journal of Physical Anthropology*, 13(2):253–283, 1955.

[88] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[89] J.A. Gómez-Valdés, A. Menéndez Garmendia, L. García-Barzola, G. Sánchez-Mejorada, C. Karam, J. P. Baraybar, and A. Klales. Recalibration of the Klales et al. (2012) method of sexing the human innominate for Mexican populations. *American Journal of Physical Anthropology*, 162(3):600–604, 2017.

[90] Y.C. Guo, M. Han, Y. Chi, H. Long, D. Zhang, J. Yang, Y. Yang, T. Chen, and S. Du. Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images. *International Journal of Legal Medicine*, pages 1–9, 2021.

[91] G. Gustafson. Age determinations on teeth. *The Journal of the American Dental Association*, 41(1):45–54, 1950.

[92] K. Haavikko. The formation and the alveolar and clinical eruption of the permanent teeth. An orthopantomographic study. *Suomen Hammaslaakariseuran toimituksia. Finska tandlakarsallskapets forhandlingar*, 66(3):103–170, 1970.

[93] M. Halperin-Sternfeld, E.E. Machtei, C. Balkow, and J. Horwitz. Patient movement during extraoral radiographic scanning. *Oral Radiology*, 32(1):40–47, 2016.

[94] E.F. Harris and D.P. Dinh. Intercusp relationships of the permanent maxillary first and second molars in American whites. *American Journal of Physical Anthropology*, 130(4):514–528, 2006.

[95] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2961–2969, 2017.

[96] P.M. Helfman and J.L. Bada. Aspartic acid racemization in tooth enamel from living humans. *Proceedings of the National Academy of Sciences*, 72(8):2891–2894, 1975.

[97] C.F. Hildebolt, M.W. Vannier, M.J. Gravier, M.K. Shrout, R.H. Knapp, and R.K. Walkup. Digital dental image processing of alveolar bone: Macintosh ii personal computer software. *Dentomaxillofacial Radiology*, 21(3):162–169, 1992.

[98] P. Hölttä, M. Nyström, M. Evälahti, and S. Alaluusua. Root–crown ratios of permanent teeth in a healthy Finnish population assessed from panoramic radiographs. *European Journal of Orthodontics*, 26(5):491–497, 2004.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[99] K. S. Hu, K. S. Koh, S. H. Han, K. J. Shin, and H. J. Kim. Sex determination using nonmetric characteristics of the mandible in Koreans. *Journal of Forensic Sciences*, 51(6):1376–1382, 2006.

[100] I. Ilić, M. Vodanović, and M. Subašić. Gender estimation from panoramic dental X-ray images using deep convolutional networks. In *IEEE EUROCON 18th International Conference on Smart Technologies*, pages 1–5. IEEE, 2019.

[101] A.P. Indira, A. Markande, and M.P. David. Mandibular ramus: an indicator for sex determination-a digital radiographic study. *Journal of Forensic Dental Sciences*, 4(2):58, 2012.

[102] S. Ingawale and T. Goswami. Temporomandibular joint: disorders, treatments, and biomechanics. *Annals of Biomedical Engineering*, 37(5):976–996, 2009.

[103] S. Ioffe and C. Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[104] R. Iqbal, S. Zhang, and C. Mi. Reliability of mandibular canine and mandibular canine index in sex determination: a study using Uyghur population. *Journal of Forensic and Legal Medicine*, 33:9–13, 2015.

[105] J. Irurita-Olivares and I. Alemán-Aguilera. Validation of the sex estimation method elaborated by Schutkowski in the Granada osteological collection of identified infant and young children: analysis of the controversy between the different ways of analyzing and interpreting the results. *International Journal of Legal Medicine*, 130(6):1623–1632, 2016.

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

[107] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira. Deep instance segmentation of teeth in panoramic X-ray images. In *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images*, pages 400–407. IEEE, 2018.

[108] A.K. Jain and H. Chen. Matching of dental X-ray images for human identification. *Pattern Recognition*, 37(7):1519–1532, 2004.

UDC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[109] A.P. Joseph, R.K. Harish, P.K.R. Mohammed, and R.B. Vinod-Kumar. How reliable is sex differentiation from teeth measurements. *Journal of Oral and Maxillofacial Pathology*, 4(1):289–92, 2013.

[110] R. Kapila, K.S. Nagesh, A.R. Iyengar, and S. Mehkri. Sexual dimorphism in human mandibular canines: a radiomorphometric study in South Indian population. *Journal of Dental Research, Dental Clinics, Dental Prospects*, 5(2):51, 2011.

[111] P. Kapoor and V. Jain. Comprehensive chart for dental age estimation (DAEcc8) based on Demirjian 8-teeth method: simplified for operator ease. *Journal of Forensic and Legal Medicine*, 59:45–49, 2018.

[112] C.A. Karlo, P. Stolzmann, S. Habernig, L. Müller, R. Saurenmann, and C.J. Kellenberger. Size, shape and age-related changes of the mandibular condyle during childhood. *European Radiology*, 20(10):2512–2517, 2010.

[113] L. Kats, M. Vered, S. Blumer, and E. Kats. Neural network detection and segmentation of mental foramen in panoramic imaging. *Journal of Clinical Pediatric Dentistry*, 44(3):168–173, 2020.

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

[115] C. Kim, D. Kim, H. Jeong, S.J. Yoon, and S. Youm. Automatic tooth detection and numbering using a combination of a CNN and heuristic algorithm. *Applied Sciences*, 10(16):5624, 2020.

[116] S. Kim, Y.H. Lee, Y.K. Noh, F.C. Park, and Q.S. Auh. Age-group determination of living individuals using first molar images based on artificial intelligence. *Scientific Reports*, 11, 2021.

[117] T.L. Koch, M. Perslev, C. Igel, and S.S. Brandt. Accurate segmentation of dental panoramic radiographs with U-Nets. In *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, pages 15–19. IEEE, 2019.

[118] B.S. Kraus and R.E. Jordan. *The human dentition before birth*. Lea & Febiger, 1965.

USC

[119] K. Krishan, P.M. Chatterjee, T. Kanchan, S. Kaur, N. Baryah, and R.K. Singh. A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Science International*, 261:165–e1, 2016.

[120] A. Kumagai, G. Willems, A. Franco, and P. Thevissen. Age estimation combining radiographic information of two dental and four skeletal predictors in children and subadults. *International Journal of Legal Medicine*, 132(6):1769–1777, 2018.

[121] S.S. Kumar, V. Thailavathy, D. Srinivasan, D. Loganathan, and J. Yamini. Comparison of orthopantomogram and lateral cephalogram for mandibular measurements. *Journal of Pharmacy & Bioallied Sciences*, 9(Suppl 1):S92, 2017.

[122] S. Kundu. AI in medicine must be explainable. *Nature Medicine*, 27(8):1328–1328, 2021.

[123] Y.F. Kuo, S.Y. Lin, C.H. Wu, S.L. Chen, T.L. Lin, N.H. Lin, C.H. Mai, and J.F. Villaverde. A convolutional neural network approach for dental panoramic radiographs classification. *Journal of Medical Imaging and Health Informatics*, 7(8):1693–1704, 2017.

[124] S.I. Kvaal, K.M. Kolltveit, I.O. Thomsen, and T. Solheim. Age estimation of adults from dental radiographs. *Forensic Science International*, 74(3):175–185, 1995.

[125] O. Kwon, T.H. Yong, S.R. Kang, J.E. Kim, K.H. Huh, M.S. Heo, S.S. Lee, S.C. Choi, and W.J. Yi. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. *Dentomaxillofacial Radiology*, 49(8):20200185, 2020.

[126] Labelbox. Labelbox. `https://labelbox.com/`, 2021.

[127] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[128] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[129] J.H. Lee, S.S. Han, Y.H. Kim, C. Lee, and I. Kim. Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radio-

UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

graphs. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 129(6):635–642, 2020.

[130] K.S. Lee, S.K. Jung, J.J. Ryu, S.W. Shin, and J. Choi. Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. *Journal of Clinical Medicine*, 9(2):392, 2020.

[131] T.M. Lehmann, E. Troeltsch, and K. Spitzer. Image processing and enhancement provided by commercial dental software programs. *Dentomaxillofacial Radiology*, 31(4):264–272, 2002.

[132] A.F. Leite, A. Van Gerven, H. Willems, T. Beznik, P. Lahoud, H. Gaêta-Araujo, M. Vranckx, and R. Jacobs. Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs. *Clinical Oral Investigations*, 25(4):2257–2267, 2021.

[133] A.B. Lewis and S.M. Garn. The relationship between tooth formation and other maturational factors. *The Angle Orthodontist*, 30(2):70–77, 1960.

[134] H.W. Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318):399–402, 1967.

[135] C. Lin, B. Jiao, S. Liu, F. Guan, N.E. Chung, S.H. Han, and U.Y. Lee. Sex determination from the mandibular ramus flexure of Koreans by discriminant function analysis using three-dimensional mandible models. *Forensic Science International*, 236:191–e1, 2014.

[136] T.Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2117–2125, 2017.

[137] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, and C.I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[138] H. Liversidge. Variation in modern human dental development. *Cambridge Studies in Biological and Evolutionary Anthropology*, pages 73–113, 2003.

UDC Universidade de Santiago de Compostela

[139] L.J. Lopes, T.O. Gamba, J.V.J. Bertinato, and D.Q. Freitas. Comparison of panoramic radiography and CBCT to identify maxillary posterior roots invading the maxillary sinus. *Dentomaxillofacial Radiology*, 45(6):20160043, 2016.

[140] S.R. Loth and M. Henneberg. Mandibular ramus flexure: a new morphologic indicator of sexual dimorphism in the human skeleton. *American Journal of Physical Anthropology*, 99(3):473–485, 1996.

[141] M. Maber, H.M. Liversidge, and M.P. Hector. Accuracy of age estimation of radiographic methods using developing teeth. *Forensic Science International*, 159:S68–S73, 2006.

[142] F.P. Mahdi, K. Motoki, and S. Kobashi. Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs. *Scientific Reports*, 10(1):1–12, 2020.

[143] M. Marinescu, V. Panaitescu, and M. Rosu. Sex determination in Romanian mandible using discriminant function analysis: comparative results of a time-efficient method. *Romanian Journal of Legal Medicine*, 21(4):305–8, 2013.

[144] G. Markic, L. Müller, R. Patcas, M. Roos, N. Lochbühler, T. Peltomäki, C.A. Karlo, O. Ullrich, and C.J. Kellenberger. Assessing the length of the mandibular ramus and the condylar process: a comparison of OPG, CBCT, CT, MRI, and lateral cephalometric measurements. *European Journal of Orthodontics*, 37(1):13–21, 2015.

[145] S. Martin-de las Heras, P. Garcia-Fortea, A. Ortega, S. Zodocovich, and A. Valenzuela. Third molar development according to chronological age in populations from Spanish and Magrebian origin. *Forensic Science International*, 174(1):47–53, 2008.

[146] M. Mayil, G. Keser, and F.N. Pekiner. Clinical image quality assessment in panoramic radiography. *Marmara Üniversitesi Saglik Bilimleri Enstitüsü Dergisi*, 4(3):126, 2014.

[147] D. McCloe, I. Marion, M.A. da Fonseca, M. Colvard, and S. AlQahtani. Age estimation of Hispanic children using the London Atlas. *Forensic Science International*, 288:332–e1, 2018.

[148] 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 Science International*, 270:276–e1, 2017.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

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

[150] 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*, pages 199–214. World Scientific, 2006.

[151] D. Milošević, M. Vodanović, I. Galić, and M. Subašić. Estimating biological gender from panoramic dental X-ray images. In *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 105–110. IEEE, 2019.

[152] C.F.A. Moorrees, E.A. Fanning, and E.E. Hunt Jr. Age variation of formation stages for ten permanent teeth. *Journal of Dental Research*, 42(6):1490–1502, 1963.

[153] I. Mughal, A.S. Saqib, and F. Manzur. Mandibular canine index (MCI). *The Professional Medical Journal*, 17(03):459–463, 2010.

[154] I.C.L. Muñoz and P.B. Orta. Comparison of cephalometric patterns in mouth breathing and nose breathing children. *International Journal of Pediatric Otorhinolaryngology*, 78(7):1167–1172, 2014.

[155] P. Murail, J. Bruzek, F. Houët, and E. Cunha. DSP: a tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, (17 (3-4)):167–176, 2005.

[156] C. Muramatsu, K. Horiba, T. Hayashi, T. Fukui, T. Hara, A. Katsumata, and H. Fujita. Quantitative assessment of mandibular cortical erosion on dental panoramic radiographs for screening osteoporosis. *International Journal for Computer Assisted Radiology and Surgery*, 11(11):2021–2032, 2016.

[157] C. Muramatsu, T. Morishita, R. Takahashi, T. Hayashi, W. Nishiyama, Y. Ariji, X. Zhou, T. Hara, A. Katsumata, and H. Fujita. Tooth detection and classification on panoramic radiographs for automatic dental chart filing: improved classification by multi-sized input data. *Oral Radiology*, 37(1):13–19, 2021.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[158] M.P. Muresan, A.R. Barbura, and 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)*, pages 457–463. IEEE, 2020.

[159] A. Naik, S. Tikhe, S. Bhide, K.P. Kaliyamurthie, and T Saravanan. Automatic segmentation of lower jaw and mandibular bone in digital dental panoramic radiographs. *Indian Journal of Science and Technology*, 9(21), 2016.

[160] V. Nair and G.E. Hinton. Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814, 2010.

[161] S.P. Nawrocki. The nature and sources of error in the estimation of age at death from the skeleton. In *Age estimation of the human skeleton*, pages 79–101. Charles C Tomas Publisher Illinois, 2010.

[162] A. Newell, K. Yang, and J. Deng. Stacked hourglass networks for human pose estimation. In *European Conference on Computer Vision – ECCV 2016*, pages 483–499. Springer, 2016.

[163] C.M. Nolla. The development of the human dentition. *ASDC Journal of Dentistry for Children*, 27(27):254–266, 1960.

[164] I. Nurtanio, E.R. Astuti, I. Ketut Eddy Purnama, M. Hariadi, and M.H. Purnomo. Classifying cyst and tumor lesion using support vector machine based on dental panoramic images texture features. *IAENG International Journal of Computer Science*, 40(1):29–37, 2013.

[165] E. Nuzzolese, P. Randolph-Quinney, J. Randolph-Quinney, and G. Di Vella. Geometric morphometric analysis of sexual dimorphism in the mandible from panoramic X-ray images. *The Journal of Forensic Odonto-Stomatology*, 37(2):35, 2019.

[166] R. Obuchowicz, K. Nurzynska, B. Obuchowicz, A. Urbanik, and A. Piórkowski. Caries detection enhancement using texture feature maps of intraoral radiographs. *Oral Radiology*, 36(3):275–287, 2020.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[167] A.B. Oktay. Tooth detection with convolutional neural networks. In *2017 Medical Technologies National Congress (TIPTEKNO)*, pages 1–4. IEEE, 2017.

[168] K. Ortakoglu, T. Akcam, M. Sencimen, O. Karakoc, H. A. Ozyigit, and O. Bengi. Osteochondroma of the mandible causing severe facial asymmetry: a case report. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 103(5):e21–e28, 2007.

[169] A.G. Ortiz, C. Costa, R.H.A. Silva, M.G.H. Biazevic, and E. Michel-Crosato. Sex estimation: anatomical references on panoramic radiographs using machine learning. *Forensic Imaging*, 20:200356, 2020.

[170] M. Paknahad, S. Shahidi, and Z. Zarei. Sexual dimorphism of maxillary sinus dimensions using cone-beam computed tomography. *Journal of Forensic Sciences*, 62(2):395–398, 2017.

[171] M. Paknahad, M. Vossoughi, and F. A. Zeydabadi. A radio-odontometric analysis of sexual dimorphism in deciduous dentition. *Journal of Forensic and Legal Medicine*, 44:54–57, 2016.

[172] N.M. Parr, N.V. Passalacqua, and K. Skorpinski. Investigations into age-related changes in the human mandible. *Journal of Forensic Sciences*, 62(6):1586–1591, 2017.

[173] T.R. Peckmann, S. Meek, N. Dilkie, and M. Mussett. Sex estimation using diagonal diameter measurements of molar teeth in African American populations. *Journal of Forensic and Legal Medicine*, 36:70–80, 2015.

[174] S. Perschbacher. Interpretation of panoramic radiographs. *Australian Dental Journal*, 57:40–45, 2012.

[175] L.B. Petersen, K.R. Olsen, J. Christensen, and A. Wenzel. Image and surgery-related costs comparing cone beam CT and panoramic imaging before removal of impacted mandibular third molars. *Dentomaxillofacial Radiology*, 43(6):20140001, 2014.

[176] V.M. Phillips and T.J. van Wyk Kotze. Testing standard methods of dental age estimation by Moorrees, Fanning and Hunt and Demirjian, Goldstein and Tanner on three South African children samples. *Journal of Forensic Odonto-Stomatology*, 27(2):20–28, 2009.

[177] M.A. Pilloud and G.R. Scott. Dentition in the estimation of sex. In *Sex estimation of the human skeleton*, pages 149–169. Elsevier, 2020.

UDC Universidade de Santiago de Compostela

[178] W. Poedjiastoeti and S. Suebnukarn. Application of convolutional neural network in the diagnosis of jaw tumors. *Healthcare informatics research*, 24(3):236–241, 2018.

[179] R. Pokhrel and R. Bhatnagar. Sexing of mandible using ramus and condyle in Indian population: a discriminant function analysis. *European Journal of Anatomy*, 17(1):39–42, 2013.

[180] J.L. Prieto, E. Barbería, R. Ortega, and C. Magaña. Evaluation of chronological age based on third molar development in the Spanish population. *International Journal of Legal Medicine*, 119(6):349–354, 2005.

[181] C. Priyadarshini, M.P. Puranik, and S.R. Uma. *Dental age estimation methods. A review.* LAP Lambert Academic Publ, 2015.

[182] C.L. de Queiroz, A.S.S.D. Terada, T.U. Dezem, L. Gomes de Araújo, R. Galo, C. Oliveira-Santos, and R.H. Alves da Silva. Sex determination of adult human maxillary sinuses on panoramic radiographs. *Acta Stomatologica Croatica*, 50(3):215–221, 2016.

[183] B. Rai, J. Kaur, M. Cingolani, L. Ferrante, and R. Cameriere. Age estimation in children by measurement of open apices in teeth: an Indian formula. *International Journal of Legal Medicine*, 124(3):237–241, 2010.

[184] N.G. Rao, N.N. Rao, M.L. Pai, and M.S. Kotian. Mandibular canine index — a clue for establishing sex identity. *Forensic Science International*, 42(3):249–254, 1989.

[185] V.M. Reddy, S. Saxena, and P. Bansal. Mandibular canine index as a sex determinant: a study on the population of western Uttar Pradesh. *Journal of Oral and Maxillofacial Pathology*, 12(2):56, 2008.

[186] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: towards real-time object detection with region proposal networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in neural information processing systems*, volume 28. Curran Associates, Inc., 2015.

[187] J.E. Richards and R.S. Hawley. Chapter 8 - Sex determination: how genes determine a developmental choice. In J.E. Richards and R.S. Hawley, editors, *The human genome (third edition)*, pages 273–298. Academic Press, San Diego, third edition edition, 2011.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[188] S. Ritz-Timme, C. Cattaneo, M.J. Collins, E.R. Waite, H.W. Schütz, H.J. Kaatsch, and H.I.M. Borrman. Age estimation: the state of the art in relation to the specific demands of forensic practise. *International Journal of Legal Medicine*, 113(3):129–136, 2000.

[189] M. Roberts, J. Yuan, J. Graham, R. Jacobs, and H. Devlin. Changes in mandibular cortical width measurements with age in men and women. *Osteoporosis International*, 22(6):1915–1925, 2011.

[190] R.H.N. Rondon, Y.C.L. Pereira, and G.C. do Nascimento. Common positioning errors in panoramic radiography: a review. *Imaging Science in Dentistry*, 44(1):1–6, 2014.

[191] A. Sabarudin and Y.J. Tiau. Image quality assessment in panoramic dental radiography: a comparative study between conventional and digital systems. *Quantitative Imaging in Medicine and Surgery*, 3(1):43, 2013.

[192] V. Saini, R. Srivastava, R.K. Rai, S.N. Shamal, T.B. Singh, and S.K. Tripathi. Mandibular ramus: an indicator for sex in fragmentary mandible. *Journal of Forensic Sciences*, 56:S13–S16, 2011.

[193] A. Sakuma, S. Ohtani, H. Saitoh, and H. Iwase. Comparative analysis of aspartic acid racemization methods using whole-tooth and dentin samples. *Forensic Science International*, 223(1-3):198–201, 2012.

[194] A. Schmeling, C. Grundmann, A. Fuhrmann, H.J. Kaatsch, B. Knell, F. Ramsthaler, W. Reisinger, T. Riepert, S. Ritz-Timme, F.W. Rösing, K. others, Rötzscher, and G. Geserick. Criteria for age estimation in living individuals. *International Journal of Legal Medicine*, 122(6):457–460, 2008.

[195] G.T. Schwartz and M.C. Dean. Sexual dimorphism in modern human permanent teeth. *American Journal of Physical Anthropology*, 128(2):312–317, 2005.

[196] P. Selliah, F. Martino, M. Cummaudo, L. Indra, L. Biehler-Gomez, C. P. Campobasso, and C. Cattaneo. Sex estimation of skeletons in middle and late adulthood: reliability of pelvic morphological traits and long bone metrics on an Italian skeletal collection. *International Journal of Legal Medicine*, 134(5):1683–1690, 2020.

[197] 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 *Proceedings of the IEEE International Conference on Computer Vision*, pages 618–626, 2017.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[198] P.H. Shah, R. Venkatesh, C.B. More, V Vaishnavee, et al. Age-and sex-related mandibular dimensional changes: a radiomorphometric analysis on panoramic radiographs. *Indian Journal of Dental Research*, 31(1):113, 2020.

[199] S.S. Shapiro and M.B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[200] M. Sharma, R.K. Gorea, A. Gorea, and A. Abuderman. A morphometric study of the human mandible in the Indian population for sex determination. *Egyptian Journal of Forensic Sciences*, 6(2):165–169, 2016.

[201] A.M. Silva, M.L. Pereira, S. Gouveia, J.N. Tavares, A. Azevedo, and I.M. Caldas. A new approach to sex estimation using the mandibular canine index. *Medicine, Science and Law*, 56(1):7–12, 2016.

[202] B. Silva, L. Pinheiro, L. Oliveira, and 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)*, pages 164–171. IEEE, 2020.

[203] G. Silva, L. Oliveira, and M. Pithon. Automatic segmenting teeth in X-ray images: trends, a novel data set, benchmarking and future perspectives. *Expert Systems with Applications*, 107:15–31, 2018.

[204] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[205] S. Singh, A. Gupta, B. Padmavathi, S. Kumar, S. Roy, and A. Kumar. Mandibular canine index: a reliable predictor for gender identification using study cast in Indian population. *Indian Journal of Dental Research*.

[206] D.M. Son, Y.A. Yoon, H.J. Kwon, C.H. An, and S.H. Lee. Automatic detection of mandibular fractures in panoramic radiographs using deep learning. *Diagnostics*, 11(6):933, 2021.

[207] M.K. Spradley and R.L. Jantz. Sex estimation in forensic anthropology: skull versus postcranial elements. *Journal of Forensic Sciences*, 56(2):289–296, 2011.

[208] M. Steyn and M.Y. Işcan. Sexual dimorphism in the cranial and mandibles of South African whites. *Forensic Science International*, 98(1-2):9–16, 1998.

UDC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[209] K. Sun, B. Xiao, D. Liu, and J. Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 5693–5703, 2019.

[210] D. Sweet. Forensic dental identification. *Forensic Science International*, 201(1-3):3–4, 2010.

[211] N.S. A. Taleb and M.E. Beshlawy. Mandibular ramus and gonial angle measurements as predictors of sex and age in an Egyptian population sample: A digital panoramic study. *Journal of Forensic Research*, 6(308):2, 2015.

[212] D. Tardivo, J. Sastre, M. Ruquet, L. Thollon, P. Adalian, G. Leonetti, and B. Foti. Three-dimensional modeling of the various volumes of canines to determine age and sex: a preliminary study. *Journal of Forensic Sciences*, 56(3):766–770, 2011.

[213] A. Teivens and H. Mörnstad. A modification of the Demirjian method for age estimation in children. *The Journal of Forensic Odonto-Stomatology*, 19(2):26–30, 2001.

[214] B. Thanathornwong and S. Suebnukarn. Automatic detection of periodontal compromised teeth in digital panoramic radiographs using faster regional convolutional neural networks. *Imaging Science in Dentistry*, 50(2):169, 2020.

[215] L.F. Tomás, L.S.M. Mónico, I. Tomás, P. Varela-Patiño, and B. Martin-Biedma. The accuracy of estimating chronological age from Demirjian and Nolla methods in a Portuguese and Spanish sample. *BMC Oral Health*, 14(1):160, 2014.

[216] D.V. Tuzoff, L.N. Tuzova, M.M. Bornstein, A.S. Krasnov, M.A. Kharchenko, S.I. Nikolenko, M.M. Sveshnikov, and G.B. Bednenko. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofacial Radiology*, 48(4):20180051, 2019.

[217] D.H. Ubelaker and C.M. DeGaglia. Population variation in skeletal sexual dimorphism. *Forensic Science International*, 278:407–e1, 2017.

[218] D.H. Ubelaker and C.M. DeGaglia. Factors of population variation in sex estimation methodology. In *Sex estimation of the human skeleton*, pages 281–293. Elsevier, 2020.

[219] D.H. Ubelaker and C.G. Volk. A test of the Phenice method for the estimation of sex. *Journal of Forensic Science*, 47(1):19–24, 2002.

UDC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[220] J. Viciano, R. D'Anastasio, and L. Capasso. Odontometric sex estimation on three populations of the Iron Age from Abruzzo region (central–southern Italy). *Archives of Oral Biology*, 60(1):100–115, 2015.

[221] J. Viciano, S. López-Lázaro, and I. Alemán. Sex estimation based on deciduous and permanent dentition in a contemporary Spanish population. *American Journal of Physical Anthropology*, 152(1):31–43, 2013.

[222] N. Vila-Blanco, M.J. Carreira, P. Varas-Quintana, C. Balsa-Castro, and I. Tomás. Deep neural networks for chronological age estimation from OPG images. *IEEE Transactions on Medical Imaging*, 39(7):2374–2384, 2020.

[223] N. Vila-Blanco, P. Varas-Quintana, Á. Aneiros-Ardao, I. Tomás, and M.J. Carreira. Automated description of the mandible shape by deep learning. *International Journal of Computer Assisted Radiology and Surgery*, pages 1–10, 2021.

[224] N. Vila-Blanco, R.R. Vilas, M.J. Carreira, and I. Tomás. Towards deep learning reliable gender estimation from dental panoramic radiographs. In *Proceedings 9th European Starting AI Researchers' Symposium co-located with 24th European Conference on Artificial Intelligence (ECAI 2020)*, volume 2655. CEUR Workshop Proceedings, 2020.

[225] P.L. Walker. Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, 136(1):39–50, 2008.

[226] S.E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4724–4732, 2016.

[227] S.C. White and M.J. Pharoah. *Oral radiology: principles and interpretation*. Elsevier Health Sciences, 2014.

[228] F. Wilcoxon. Probability tables for individual comparisons by ranking methods. *Biometrics*, 3(3):119–122, 1947.

[229] G. Willems. A review of the most commonly used dental age estimation techniques. *The Journal of Forensic Odonto-Stomatology*, 19(1):9–17, 2001.

UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

[230] G. Willems, A. Van Olmen, B. Spiessens, and C. Carels. Dental age estimation in Belgian children: Demirjian's technique revisited. *Journal of Forensic Science*, 46(4):893–895, 2001.

[231] A. Wirtz, S.G. Mirashi, and S. Wesarg. Automatic teeth segmentation in panoramic X-ray images using a coupled shape model in combination with a neural network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention — MICCAI 2018*, pages 712–719. Springer, 2018.

[232] World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310(20):2191–2194, 11 2013.

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

[234] X. Ye, F. Jiang, X. Sheng, H. Huang, and X. Shen. Dental age assessment in 7–14-year-old Chinese children: comparison of Demirjian and Willems methods. *Forensic Science International*, 244:36–41, 2014.

[235] 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 Science International*, 280:245–e1, 2017.

[236] M.D. Zeiler. ADADELTA: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

[237] E. Zorba, K. Moraitis, and S.K. Manolis. Sexual dimorphism in permanent teeth of modern Greeks. *Forensic Science International*, 210(1-3):74–81, 2011.

[238] E. Zorba, V. Vanna, and K. Moraitis. Sexual dimorphism of root length on a Greek population sample. *Homo*, 65(2):143–154, 2014.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

List of Figures

Fig. 1.1 Different views of dental radiographs. 12

Fig. 1.2 Anatomic structures observable in a dental panoramic radiograph. 13

Fig. 3.1 Approval of the ethical committee of the Universidade de Santiago de Compostela (USC) to use the OPG-AS database. 29

Fig. 3.2 OPG examples by age. 31

Fig. 3.3 OPG examples by condition. 32

Fig. 4.1 Licencing information of the publication [222], from which the contents of Chapter 4 are taken from. 35

Fig. 4.2 Licencing information of the publication [224], from which the contents of Chapter 4 are taken from. 35

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Fig. 4.3 DASNet architecture. In red, the age prediction path represents the DANet 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 sex classification path, which complements the DANet approach and leads to the development of the DASNet 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. 38

Fig. 4.4 Cross validation approach defined to assess the performance of the proposed architectures. First, the dataset is split into eight folds (from $F0$ to $F7$). Then, for each CV 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. 41

Fig. 4.5 Real age *vs.* the DASNet predictions evaluated for the entire dataset, S_{all}. The closer the points are to the red line, the better the performance. 47

Fig. 4.6 Real age *vs.* 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. 47

Fig. 4.7 DASNet absolute error in the chronological age estimation when training with reduced datasets and evaluating performance for different age ranges. . 48

Fig. 4.8 Distribution of the DASNet errors (true age - predicted age), shown as the density estimation, in all the images *vs.* the *perfect* images (those which do not present any conditioning radiological or dental characteristic, as explained in section 4.2). 48

Fig. 4.9 Example of Grad-CAM attention maps at different ages. The redder the region, the more attention the network pays to it. 51

Fig. 4.10 Graphical comparison of the predictions of DANet and DASNet against the real age values, for the entire dataset and the reduced datasets. The white points correspond to the means of the distribution. 54

Fig. 4.11 Histograms of real age and networks predictions, and quantile-quantile plots comparing each pair of distributions. 55

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Fig. 4.12 SiZer plots representing the increasing/decreasing pattern of a density estimation evaluated with different smoothing values (bandwidth). In the x-axis, the age values; in the y-axis, the bandwidth values in logarithmic scale. Each colour indicates a different pattern: blue, when the density is increasing; red, when the density is decreasing; pink, when the density is not changing significantly; and grey, when there is not enough data to obtain a density estimation for a specific value and a specific bandwidth. The vertical line corresponds to the median value. 59

Fig. 4.13 Evolution of the classification metrics in subjects from seven age groups, ranging from five up to 20 years old. 63

Fig. 5.1 Licencing information of the publication [223], from which the contents of Chapter 5 are taken from. 68

Fig. 5.2 Process of describing the mandible shape from a panoramic X-ray image. In a first step, the mandible contour composed of both landmarks and semi-landmarks is obtained automatically with a CNN. In a second step, four descriptors are applied, including a set of linear distances and angles; the centroid size, the variations from the mean shape, and the shape parameters produced by a PDM. 70

Fig. 5.3 Normalised mandible digitisation. 71

Fig. 5.4 Proposed linear distances and angles. 74

Fig. 5.5 Mandible variations in subjects older than 18 regarding the sex. 81

Fig. 5.6 Mandible variations in subjects younger than 18 regarding the age. 82

Fig. 6.1 Comparison of standard (a) and oriented (b) bounding boxes. In the former, there is a high overlapping between neighbouring boxes, which is especially noticeable in the posterior teeth. 93

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Fig. 6.2 Proposed approach for age and sex estimation. First, the teeth are detected in the OPG with an R-CNN-based architecture. This includes the feature extraction process, through an FPN backbone, the proposal of teeth locations (ROIs), by an RPN, and the refinement and classification of each tooth ROI, with an ROI head. Second, the per-tooth probability distributions of age and sex are estimated. This process consists of the evaluation of the features extracted by the backbone at the estimated tooth locations, and the application of 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. 97
Fig. 6.3 Detection performance for each tooth type. The first and second molars (46, 45, 45, and 36) were clearly the best detected teeth. 104
Fig. 6.4 Qualitative results of the proposed tooth detection system. 105
Fig. 6.5 Impact of the penalty term n in the age and sex estimation performance (blue and red lines, respectively) for values greater or equal than 1. The best age estimations were obtained with $n = 3$, while the most accurate sex classification was reached with $n = 11$. 106
Fig. 6.6 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 most and least confident teeth, respectively. 108
Fig. 6.7 Attention heatmaps indicating the importance of tooth regions in the chronological age estimations. 109
Fig. 6.8 Example of the estimated per-tooth sex distributions for a female subject, and the final prediction according to the different aggregation policies ($n \in \{0, 1, 11\}$). 111

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

List of Tables

Tab. 3.1 Demographic characteristics of the OPG-AS database at the end of this PhD Thesis. 30

Tab. 4.1 Age and sex distribution of the OPG dataset, after discarding images labelled with the wrong acquisition or birth dates. 37

Tab. 4.2 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. 42

Tab. 4.3 Performance of DANet (age path). E: error; AE: absolute error; med.: median; IQR: interquartile range; p99: 99th percentile; R^2: coefficient of determination. The error metrics are given in years. 46

Tab. 4.4 Performance of DASNet (age and sex paths). Age regression abbreviations: E: error; med.: median; AE: absolute error; μ: mean; σ: standard deviation; med.: median; IQR: interquartile range; p99: 99th percentile; R^2: coefficient of determination. Sex classification abbreviations: Acc.: accuracy; $F1_f$: F1 measurement for the female class; $F1_m$: F1 measurement for the male class. The error metrics are given in years. 46

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tab. 4.5 Comparison of manual methods for dental age estimation (white cells) with the DASNet approach (grey cells). All the presented methods rely on manual measurements or classifications every time the age of a subject needed to be estimated. 50

Tab. 4.6 Error metrics of the methods described in Table 4.5 with respect to the DASNet approach. E: error (real - predicted); IQR: interquartile range; AE: absolute error. The metrics are given in years. 50

Tab. 4.7 Results of the symmetry test proposed in [150] (given as p-values obtained with a 5% significance level). 54

Tab. 4.8 Comparison of DANet and DASNet predictions *vs.* real age values. The results correspond to the p-values given by the Wilcoxon test [228] with a 5% significance level. 56

Tab. 4.9 Comparison of the median of DANet and DASNet predictions and the median of the real age values. The results correspond to the p-values estimated from a permutation test (N=3000). 56

Tab. 4.10 Multimodality analysis. **p**: p-value obtained with the test proposed in [9], where the null hypothesis states that the sample has one mode and the alternative hypothesis is if it has more than one modal value; **m. loc.** and **m. dens.**: mode location and density; **median**: the median of the distribution. In the case of multimodal distributions, the two most significant modes are given (those with higher density). 58

Tab. 4.11 Performance of the main studies on sex prediction from OPGs. (M: Males; F: Females) 61

Tab. 4.12 Age and sex distribution of the OPG dataset, after discarding images labelled with the wrong acquisition or birthdates. 61

Tab. 4.13 Sex classification results by age group. The highlighted values correspond to the most accurate networks according to each specific metric. 65

Tab. 5.1 Age and sex distribution of the OPG dataset. 70

Tab. 5.2 Description of the linear distances and angles. L|R: left and right sides. . . . 75

Tab. 5.3 Point detection absolute error ($\mu \pm \sigma$) of the best configurations for each CNN model, with and without the mandible mask as an additional target. RG: right gonion; LG: left gonion; SB: superior border; IB: inferior border; RC: right condyle; LC: left condyle; RCP: right coronoid process; LCP: left coronoid process; PT2PT: point-to-point errors measured at all mandible points; and PT2CRV: point-to-curve errors, also measured at all mandible points. 78
Tab. 5.3 *(Continuation)* 79
Tab. 5.4 Annotation errors of the 2nd observer and the prediction errors of the best-performing network (SHN), both measured against the gold standard (1st observer). 83
Tab. 5.5 Performance of the sex classification method in those aged between 18 and 70. Acc: accuracy. 84
Tab. 5.6 Comparison of the sex classification results in the literature (semiautomatic) and those of the best-performing automatic approach presented in this chapter. Meas.: Measurements. Meas. legend: DB: dry-bone; 3DS: 3D scanner; CT: computed tomography. Method legend: DFA: discriminant function analysis; LoR: logistic regression; PDM: point distribution model; RC: ridge classification. N: sample size. Test approach legend: TT: train-test; CV: cross validation; LOO: leave-one-out. Acc: accuracy. 85
Tab. 5.7 Mean and standard deviation of the absolute error (in years) in the age estimation method for subjects aged between five and 17. 86
Tab. 5.8 Comparison of the best age estimation results of the automatic methodology and the semiautomatic results presented previously in the literature. Meas. legend: 3DS: 3D scanner. Method legend: LR: linear regression; PDM: point distribution model; RR: ridge regression; N: sample size; SE: standard error (in years); R^2: coefficient of determination; p: p-value of the F-test. . 86

Tab. 6.1 Main deep learning methods for tooth detection/segmentation in OPGs. Loc.: localisation; Seg.: segmentation; Class.: classification; T: tooth type; I: individual tooth; Dent.: dentition; C: children; A: adult. 92
Tab. 6.2 Main automatic methods for age/sex estimation from OPGs. DSC: development stage classification; NA: numeric age; AGC: age group classification; LAD: legal age determination; BSC: binary sex classification. 94
Tab. 6.3 Dataset demographics. 95

USC UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

Tab. 6.4 Comparison of systems for tooth detection in OPGs[(a)]. 103
Tab. 6.5 Impact of the aggregation policy of age head on the estimation performance. 106
Tab. 6.6 Comparison of systems for automatic age estimation. 107
Tab. 6.7 Impact of the aggregation policy of sex head on the estimation performance. 110
Tab. 6.8 Comparison of systems for automatic sex determination. 111

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Publications

The contributions of this PhD Thesis have been described in the following publications:

Journals

> Vila-Blanco, N.[*,†], Carreira, M. J.[*,†], Varas-Quintana, P.[‡], Balsa-Castro, C.[†,‡], and Tomás, I.[†,‡] (2020). Deep neural networks for chronological age estimation from OPG images. IEEE transactions on medical imaging, 39(7), 2374-2384. 103615, 2020. DOI:10.1109/TMI.2020.2968765

- **Quality indicators:**

  JCR (2020):
  - Impact factor: 10.048
  - Quartile: Q1 in categories *Radiology, Nuclear Medicine & Medical Imaging* (4/133) and *Computer Science, Interdisciplinary Applications* (5/111)

  Scopus (2020):
  - Citescore: 13.8
  - Quartile: Q1 in categories *Computer Science Applications* (21/693) and *Computer Science: Software* (17/389)

[*]Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) and Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.
[†]Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Spain.
[‡]Oral Sciences Research Group, Special Needs Unit, Department of Surgery and Medical-Surgical Specialities School of Medicine and Dentistry, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.

- **PhD candidate contribution:** research conceptualisation, design and implementation of the experiments, evaluation and verification of the methodology, and manuscript writing.

- **Reproduction rights:** this publication was partially reproduced in Chapter 4. The reproduction rights have been granted by the CC BY licence (Figure 8.1)

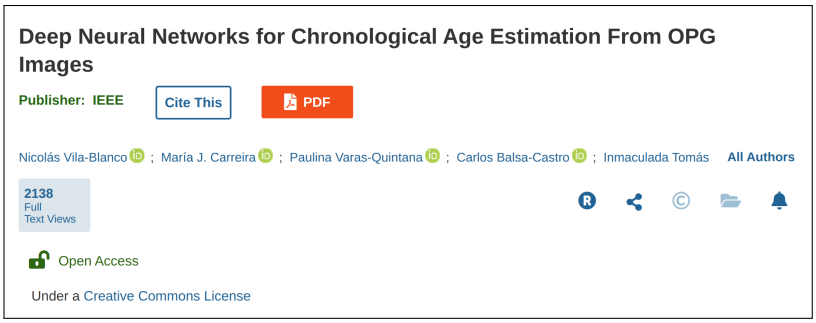


Figure 8.1: Licencing information of the publication [222], from which the contents of Chapter 4 are taken from.

> Vila-Blanco, N.[*,†], Varas-Quintana, P.[‡], Aneiros-Ardao, Á.[‡], Tomás, I.[†,‡], and Carreira, M. J.[*,†] (2021). Automated characterization of the mandible shape by deep learning. International Journal of Computer Assisted Radiology and Surgery, 16(12), 2215-2224. DOI:10.1007/s11548-021-02474-2

- **Quality indicators:**

  JCR (2020):

  - Impact factor: 2.924
  - Quartile: Q2 in category *Radiology, Nuclear Medicine & Medical Imaging* (66/133)

  Scopus (2020):

*Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) and Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.
†Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Spain.
‡Oral Sciences Research Group, Special Needs Unit, Department of Surgery and Medical-Surgical Specialities School of Medicine and Dentistry, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.

  - Citescore: 5.3
  - Quartile: Q1 in categories *Radiology, Nuclear Medicine & Imaging* (51/288), *Health Informatics* (21/95), and *Computer Science Applications* (162/693)

- **PhD candidate contribution:** research conceptualisation, design and implementation of the experiments, evaluation and verification of the methodology, and manuscript writing.

- **Reproduction rights:** this publication was partially reproduced in Chapter 5. The reproduction rights have been granted by the CC BY licence (Figure 8.2)

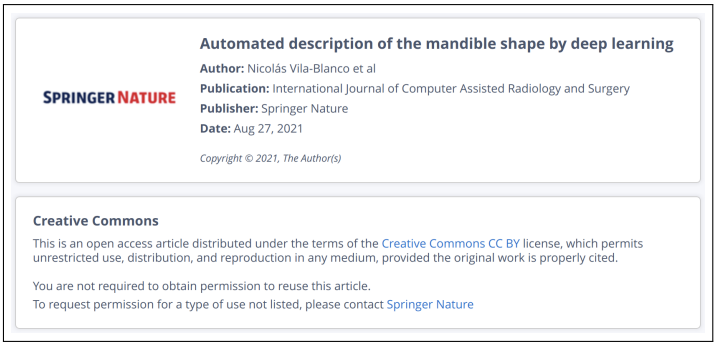


Figure 8.2: Licencing information of the publication [223], from which the contents of Chapter 5 are taken from.

> Vila-Blanco, N.[*,†], Varas-Quintana, P.[‡], Aneiros-Ardao, Á.[‡], Tomás, I.[†,‡], and Carreira, M. J.[*,†] (2021). XAS: automatic yet eXplainable Age and Sex determination by combining imprecise per-tooth predictions. Article submitted to IEEE Transactions on Medical Imaging and currently under review.

- **Quality indicators:**

  JCR (2020):

[*]Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) and Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.
[†]Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Spain.
[‡]Oral Sciences Research Group, Special Needs Unit, Department of Surgery and Medical-Surgical Specialities School of Medicine and Dentistry, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.

  - Impact factor: 10.048
  - Quartile: Q1 in categories *Radiology, Nuclear Medicine & Medical Imaging* (4/134) and *Computer Science, Interdisciplinary Applications* (5/112)

  Scopus (2020):
  - Citescore: 13.8
  - Quartile: Q1 in categories *Computer Science Applications* (21/693) and *Computer Science: Software* (17/389)

- **PhD candidate contribution:** research conceptualisation, design and implementation of the experiments, evaluation and verification of the methodology, and manuscript writing.

- **Reproduction rights:** this under-review publication was partially reproduced in Chapter 6.

Conferences

> Vila-Blanco, N.[*,†], Vilas, R. R.[*], Carreira, M. J.[*,†], and Tomás, I.[†,‡] (2020). Towards deep learning reliable gender estimation from dental panoramic radiographs. 9th Starting AI Researcher's Symposium (STAIRS), European Conference on Artificial Intelligence (ECAI), Santiago de Compostela, Spain, 2020. ISSN: 1613-0073

- **Quality indicators:** Workshop in conference ranked Class 2 in GII-GRIN-SCIE (GGS) Conference Rating.

- **PhD candidate contribution:** research conceptualisation, design and implementation of the experiments, evaluation and verification of the methodology, and manuscript writing.

- **Reproduction rights:** this publication was partially reproduced in Chapter 4. The reproduction rights have been granted by the CC BY licence (Figure 8.3)

[*]Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) and Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.
[†]Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Spain.
[‡]Oral Sciences Research Group, Special Needs Unit, Department of Surgery and Medical-Surgical Specialities School of Medicine and Dentistry, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.

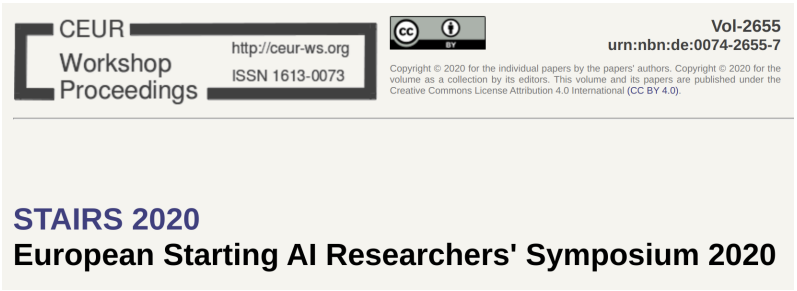


Figure 8.3: Licencing information of the publication [224], from which the contents of Chapter 4 are taken from.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA