



THÈSE

**En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 3 - Paul Sabatier**

**Présentée et soutenue par
Antoine GALIBOURG**

Le 20 juin 2022

Estimation de l'âge dentaire chez le sujet vivant : application des méthodes d'apprentissage machine chez les enfants et les jeunes adultes

Ecole doctorale : **BSB - Biologie, Santé, Biotechnologies**

Spécialité : **ANTHROPOBIOLOGIE**

Unité de recherche :

CAGT - Centre d'Anthropobiologie et de Génomique de Toulouse

Thèse dirigée par

Norbert TELMON et Delphine COMTESSE MARET

Jury

M. Maxime DUCRET, Rapporteur

Mme Delphine TARDIVO, Rapporteur

M. Paul MONSARRAT, Examineur

M. Norbert TELMON, Directeur de thèse

Mme Delphine MARET-COMTESSE, Co-directrice de thèse

Mme Véronique ALUNNI, Présidente

Remerciements aux membres du jury

Monsieur le Professeur Norbert TELMON

Je suis très honoré que vous ayez accepté la direction de ce travail et que vous ayez apporté vos compétences et votre expérience pour guider ces recherches. Je vous remercie également pour votre confiance et pour m'avoir donné la possibilité d'enseigner en master 1 et pour le PASS. Soyez assuré de mon plus profond respect.

Madame le Docteur Delphine MARET - COMTESSE

Je te remercie d'avoir accepté la co-direction de ce travail qui est l'aboutissement des directions de mes master 1 et 2. Tes compétences, ta bienveillance, ton écoute sont un modèle pour moi. Merci de m'avoir guidé, motivé, soutenu dans ce parcours hospitalo-universitaire depuis 10 ans. L'aboutissement de ce travail est une marque de ma plus profonde reconnaissance et de mon très grand respect.

Monsieur le Docteur Paul MONSARRAT

C'est un honneur pour moi que tu aies accepté de co-diriger ce travail. Depuis nos années d'étude, j'ai pu apprécier ton énergie, ta bonne humeur et tes grandes compétences scientifiques et pratiques. Je te remercie de m'avoir fait confiance, de m'avoir formé au codage et à l'intelligence artificielle. Ton dynamisme communicatif et ta disponibilité sont une source de motivation. Sois assuré de ma plus grande estime.

Madame le Professeur Delphine TARDIVO

Je vous remercie sincèrement d'avoir accepté de juger ce travail en qualité de rapporteure et d'y apporter vos compétences médico-légales. Soyez assurée de ma reconnaissance et de mon très grand respect.

Monsieur le Docteur Maxime DUCRET

Je te remercie de l'honneur que tu me fais d'avoir accepté de juger ce travail en tant que rapporteur. C'est un honneur de pouvoir bénéficier de ton expérience en intelligence artificielle. Je te témoigne également mon admiration pour ton apport à la pédagogie en prothèses. Sois assuré de ma reconnaissance et de mon respect.

Madame le Professeur Véronique ALUNNI

Je vous remercie très sincèrement de l'intérêt que vous avez porté à cette thèse et d'avoir accepté de siéger à ce jury. Veuillez trouver ici l'expression de ma gratitude et de mes sentiments les plus respectueux.

Remerciements

A l'équipe du Centre d'Anthropobiologie et de Génomique de Toulouse, dirigée par le Professeur Ludovic Orlando :

Merci pour l'accueil chaleureux au cours de ces 4 années.

José Braga, véritable source d'inspiration, vous êtes à l'origine de ma passion pour l'anthropologie. Je vous remercie pour vos enseignements, vos conseils, vos compétences et votre soutien.

Frédéric, merci pour tes enseignements et pour ta confiance.

Rozenn, ce fut un plaisir de pouvoir travailler avec toi sur Anne d'Alègre.

Clio, je suis impressionnée par tes compétences, ta bonne humeur communicative et ta capacité à fédérer les gens autour de toi.

Morgane, un grand merci pour tes conseils et tes idées innovantes en pédagogie.

Xavier, merci pour ton aide et tes bons conseils.

Jean, pour avoir travaillé avec toi sur tous mes projets de recherche, pour avoir calmement écouté mes projets et trouvé des solutions rationnelles, je te remercie et je tiens à te témoigner mon plus grand respect.

Aux corps enseignant du département de chirurgie dentaire de la faculté de santé de Toulouse :

Je vous remercie pour votre confiance et votre soutien depuis mon parcours étudiant jusqu'à aujourd'hui. C'est un honneur pour moi d'avoir intégré cette belle équipe pleine de talents.

A l'équipe de prothèse de la faculté :

Pr Serge Armand, merci de m'avoir fait confiance, de m'avoir formé et de m'avoir conseillé tout au long de mon parcours. C'était un privilège et un plaisir d'avoir pu travailler à vos côtés.

Pr Philippe Pomar, je vous remercie pour votre écoute bienveillante et de m'avoir guidé et soutenu lors de mon parcours universitaire.

Jean, je me rappellerai toujours ta présence à mes côtés dans les moments durs. Merci pour ton soutien et ta bonne humeur.

Rémi, je te remercie pour ta confiance et ta bienveillance. A la passion commune pour la BD et l'histoire.

Florent, tu es un modèle de réussite universitaire à mes yeux. J'ai un immense respect pour ta capacité de travail, d'empathie et ta capacité à fédérer les gens autour de toi.

Coralie, tu es une bénédiction pour la faculté, je te remercie pour tout ce que tu fais pour la clinique et l'organisation des TP.

Aux autres membres de la faculté :

Cathy, en tant que voisin, je voudrais te témoigner mon plus profond respect et ma joie de travailler à tes côtés. Tes compétences, ton sens éthique et ta bonne humeur sont une source d'inspiration.

Frédéric, ton intérêt pour tous les sujets, tes facilités à trouver des solutions, tes compétences cliniques et tes enseignements sont remarquables. C'est un plaisir de travailler à tes côtés.

Franck, tu es un modèle à suivre pour moi. Ton souci du détail, ta rigueur, ton empathie sont des qualités qui te caractérisent. Merci pour tes conseils précieux qui poussent au dépassement.

Philippe Kemoun, merci pour tes précieux conseils et ton soutien. Le cyclisme aussi est une science.

Marie Gurgel, rayon de soleil du jeudi, c'est un plaisir de travailler avec toi pour tes compétences, ta disponibilité et ta bonne humeur.

Vincent, derrière ta voix de stentor se cache un homme aux qualités encore plus volumineuses. Merci pour tes conseils. Je me rappelle encore tes causeries d'avant match avec l'équipe de rugby de la fac.

Jean-Noël, les qualificatifs à ton égard sont nombreux tellement tu excelles en tout. C'est une véritable chance pour moi de pouvoir travailler à tes côtés. Ta sérénité, ton empathie, tes enseignements et ta façon d'aborder la vie en général sont une source d'inspiration. Merci.

Sarah, merci pour ton écoute, ton soutien, ta bonne humeur. Très heureux de travailler avec toi sur l'institut de simulation.

Sara, un grand merci pour m'avoir recousu le pouce. C'est un plaisir de pouvoir travailler à tes côtés le jeudi.

Karim, un humour caustique et de grandes qualités humaines. Merci pour tout ce que tu fais pour impulser de la nouveauté à la fac et au CHU.

Thibault, la théorie du TFC était donc vraie. Vrai plaisir de travailler avec toi et d'avoir partagé des moments de stress en pays bretonnant.

Luc, tes qualités humaines et de clinicien forcent le respect. Tu donnes une belle impulsion à la prothèse et c'est un véritable plaisir de travailler à tes côtés. Merci pour ton soutien.

Marie et Matthieu, votre bonne humeur et l'énorme travail accompli pour le service donnent envie de se dépasser. Je vous remercie pour votre soutien et votre écoute.

Clément, Antoine, des étudiants devenus de compétents praticiens que j'admire.

Léonor, une personne solaire que j'admire. Je te souhaite le meilleur pour la suite. Merci pour ton soutien.

Géromine, tes qualités humaines et pédagogiques sont impressionnantes. Hâte de travailler avec toi.

Serge, Xavier, Éric et Nicolas, un énorme merci pour votre aide à organiser des TP de qualité.

A l'équipe du DUI :

Philippe, c'est un plaisir d'avoir repris la co-direction du DUI avec toi. Ton soutien, tes conseils sont réconfortants. Au plaisir partagé de l'histoire.

Thierry, Victor et Thibault, vous êtes les rayons de soleil du lundi. C'est un immense plaisir pour moi de travailler avec vous pour faire avancer ce bateau.

Olivier, Louis Philip, Laurent et Philippe, je tenais à vous remercier pour vos enseignements. Je suis fier d'avoir appris sous votre supervision. Merci pour votre soutien.

Mention spéciale pour Victor. Comme dirait notre mentor et cela te va bien : « il ne faut pas confondre compétence et triste mine ». Un immense respect pour ta personne.

A ma belle-famille, merci pour votre présence et votre soutien.

A Hélène et sa famille, véritable source d'inspiration. Je vous remercie pour votre soutien.

A mes parents, merci pour votre patience. Merci pour votre soutien. Merci pour les belles personnes que vous êtes et pour m'avoir appris à être un homme droit.

A Joséphine et Jean.

A Marine, sans toi, ce travail n'aurait pas été possible. Merci pour ton soutien sans faille. Merci de croire en moi. Merci pour tout.

Table des matières :

Contents

Table des matières :	7
Introduction.....	8
1. Contexte de l'estimation de l'âge dentaire chez les vivants	10
1.1. L'âge chronologique	10
1.2. L'estimation de l'âge chronologique, une question qui remonte à l'antiquité.....	12
1.3. Estimation de l'âge : données actuelles.....	13
1.4. Les méthodes médicales de l'estimation de l'âge chronologique	17
1.4.1. L'examen physique.....	18
1.4.2. L'évaluation du développement squelettique	18
1.4.3. L'évaluation du développement dentaire	18
1.4.4. Les méthodes d'atlas.....	19
1.4.5. Les méthodes de mesures directes.....	20
1.4.6. Les méthodes observationnelles.....	21
1.5. Problématiques de l'estimation de l'âge dentaire	24
2. Estimation de l'âge dentaire par différentes méthodes de machine learning	27
3. Discussion sur les méthodes d'intelligence artificielle utilisées dans l'estimation de l'âge dentaire	46
3.1. Définition du champ d'action de l'apprentissage machine.....	48
3.1.1. Méthodologie	48
3.1.2. Méthodes de régression utilisées.....	54
3.1.3. L'apprentissage profond.....	65
3.2. Mise en perspective des résultats.....	73
3.2.1. Etat de la littérature depuis 2021.....	74
3.2.2. Automatisation de l'estimation de l'âge	79
3.2.3. Problématique de la nature des données	88
3.2.4. Usage de l'estimation de l'âge dentaire à partir d'orthopantomogrammes	93
4. Conclusion	96
Bibliographie.....	99

Introduction

En anthropologie, l'analyse des dents permet de retracer des traits d'histoire de vie du genre *homo*. Une molaire temporaire, découverte dans la Drôme, a permis à Slimak *et al.* de reculer de 10 000 ans la date d'arrivée de l'homme moderne en Europe et de questionner les relations ayant existé avec *Homo Neanderthalensis*.¹ La dent, organe le plus résistant du corps humain, peut être l'objet d'analyses variées, allant de la morphologie comparée à l'analyse de l'ADN ancien présent dans le tartre.²

En médecine légale, les données dentaires font partie des trois identifiants déterminants d'un individu, au même titre que les empreintes digitales et l'ADN. Elles peuvent aussi être utilisées pour la détermination de l'âge d'un individu vivant ou mort, pour qualifier un profil biologique, identifié par son sexe et sa stature. Quel est l'âge d'un individu à sa mort ? Quel est l'âge d'un individu lorsque qu'il n'est pas possible de prouver sa date de naissance ? Telles sont des thématiques de recherche actuelles. Dès l'Empire Romain, la législation prévoyait des devoirs et des droits relatifs au citoyen en fonction de son âge et déjà l'éruption des dents permettait de différencier les individus par rapport à un âge clé.³ La dent, marqueur biologique, permet d'inférer l'âge chronologique/biologique d'une personne.

Les techniques d'évaluation de l'âge dentaire ont suivi les progrès de la science. La découverte des rayons X en 1895 a permis d'investiguer la maturation de l'organe dentaire sans être destructif, et différentes méthodes ont été créées. En 1973, Demirjian *et al.* ont mis au point un système d'évaluation de l'âge dentaire basé sur 8 stades de maturation des 7 dents permanentes de la mandibule gauche.⁴ Cette méthode, basée sur des radiographies panoramiques d'enfants Canadiens Français âgés de 3 à 17 ans permet de calculer un score de maturité global à partir des scores de maturité de chacune des dents. Une table de correspondance permet de convertir le score de maturité global en âge dentaire et un graphique de percentiles à 10, 50 et 90 % permet de situer la maturité de l'individu au niveau dentaire pour son âge chronologique.

Cela fera bientôt 50 ans que cette méthode est utilisée et elle reste la méthode d'estimation de l'âge dentaire la plus utilisée.⁵ Ce n'est pourtant pas la méthode la plus précise et elle est critiquée pour l'origine de la population de référence utilisée pour ses tables de références.⁶ Mais elle constitue un étalon de comparaison pour les autres techniques et son système de stades de maturation dentaire offre un bon compromis entre simplicité et robustesse d'utilisation.

Lors de la création de la méthode de Demirjian⁴, les premiers ordinateurs personnels commençaient à peine à voir le jour, mais leurs capacités de calcul et de stockage des données étaient alors limitées. Depuis, les travaux de développement des processeurs et des capacités de traitement des données ont suivi une progression exponentielle jusqu'à aujourd'hui, tout comme l'augmentation du volume de données générées. Statista a réalisé une estimation du volume des données numériques futures et selon leurs projections, en 2025, 4 jours suffiront à produire l'équivalent de la totalité des données numériques créées au cours de l'année 2010.⁷

L'exploitation des données numériques massives (*big data*) nécessite des analyses particulières pour en extraire des informations utiles, d'en découvrir des motifs particuliers. C'est le cas de certaines techniques d'intelligence artificielle qui permettent de réaliser un apprentissage machine à partir de données d'entraînement. Un algorithme est alors capable de réaliser une tâche qui n'a pas été spécifiquement codée comme une régression ou une classification qui sont similaires aux problématiques d'estimation de l'âge dentaire et de classification par rapport à un âge clé.

L'objectif de ce travail a été de comparer la précision de méthodes classiques d'estimation de l'âge avec différentes méthodes d'apprentissage machine sur des individus vivants. Dans une première partie, sont présentés le contexte et les méthodes classiques d'estimation de l'âge dentaire. Les problématiques de ces méthodes sont identifiées et des méthodes d'apprentissage machine sont comparées aux méthodes classiques. Dans une deuxième partie, le fonctionnement des différentes méthodes d'apprentissage machine est présenté. Puis les résultats obtenus sont comparés aux travaux récents d'autres équipes de recherche Enfin les résultats obtenus sont remis en perspective avec leur utilisation pour l'estimation de l'âge dentaire.

1. Contexte de l'estimation de l'âge dentaire chez les vivants

1.1. L'âge chronologique

L'âge chronologique constitue un des quatre paramètres du profil biologique d'un individu avec le sexe, la stature et l'ascendance.⁸ Pour la médecine légale ou l'anthropologie, il correspond au temps écoulé entre la naissance de l'individu et sa mort.

Chez un individu vivant, l'âge chronologique correspond au temps écoulé depuis la naissance de l'individu. C'est un élément d'état-civil dont jouit chaque individu, qui conditionne un encadrement juridique. En France, le code général des impôts, le code de la sécurité sociale, le code civil, le code de l'action sociale et des familles utilisent des critères d'âge maximum ou minimum pour l'application de lois.⁹ Des âges clés servent à protéger ou à responsabiliser les personnes. Les droits et responsabilités d'un individu sont explicitement structurés autour de l'âge chronologique.¹⁰ L'âge de la majorité, l'âge légal du consentement, l'âge du mariage, l'âge minimum du travail, l'âge légal de la retraite, l'âge de la responsabilité civile sont autant de thématiques pour lesquelles l'âge fixe un cadre légal. Tous ces âges clés permettent de définir et réguler l'éducation, le vote, le travail, le mariage, les impôts et les prestations sociales.

L'âge chronologique est une caractéristique individuelle mais il peut aussi être utilisé à un niveau populationnel. Il sert de variable d'entrée pour établir des statistiques visant à décrire et piloter l'action sociale d'un pays.¹¹ C'est également une donnée utilisée en marketing pour identifier et prévoir le comportement des consommateurs.¹²

Comment avoir accès à l'âge chronologique ? Comment est-il établi ? Ce sont des questions qui paraissent anodines car il suffit d'avoir accès à l'acte de naissance de l'individu qui indique sa date de naissance. Le calcul du temps écoulé depuis la date de naissance jusqu'à la date d'intérêt permet d'obtenir l'âge chronologique de l'individu. Si cette donnée semble facile d'accès pour un individu vivant en France, on peut se poser la question de savoir quelle est la proportion des individus sans certificat justificatif de naissance ?

En 2015, le Programme des Nations Unies pour le Développement a adopté 17 Objectifs de Développement Durable. Ces objectifs visent à adopter une action globale commune pour mettre fin à la pauvreté, protéger la planète et garantir la paix des peuples d'ici 2030. Le seizième objectif rappelle qu'un milliard de personnes sont "invisibles" légalement. Ce sont des individus qui ne peuvent pas prouver leur identité. Ce chiffre comprend 625 millions d'enfants de moins de 14 ans dont les naissances n'ont pas été enregistrées.¹³

D'après le décompte du fonds des Nations unies pour l'enfance (United Nations of International Children's : UNICEF), en 2020 **"1 enfant de moins de 5 ans sur 4 n'existe pas officiellement"**.¹⁴ Les pays d'Europe, d'Amérique et d'Asie ont des taux d'enregistrement des naissances supérieurs à 99%. En Afrique sub-saharienne 1 enfant sur 2 ne dispose pas d'acte de naissance. L'Ethiopie et la Somalie ont enregistré respectivement seulement 2,7 et 3,5 % des naissances des enfants de moins de 5 ans.

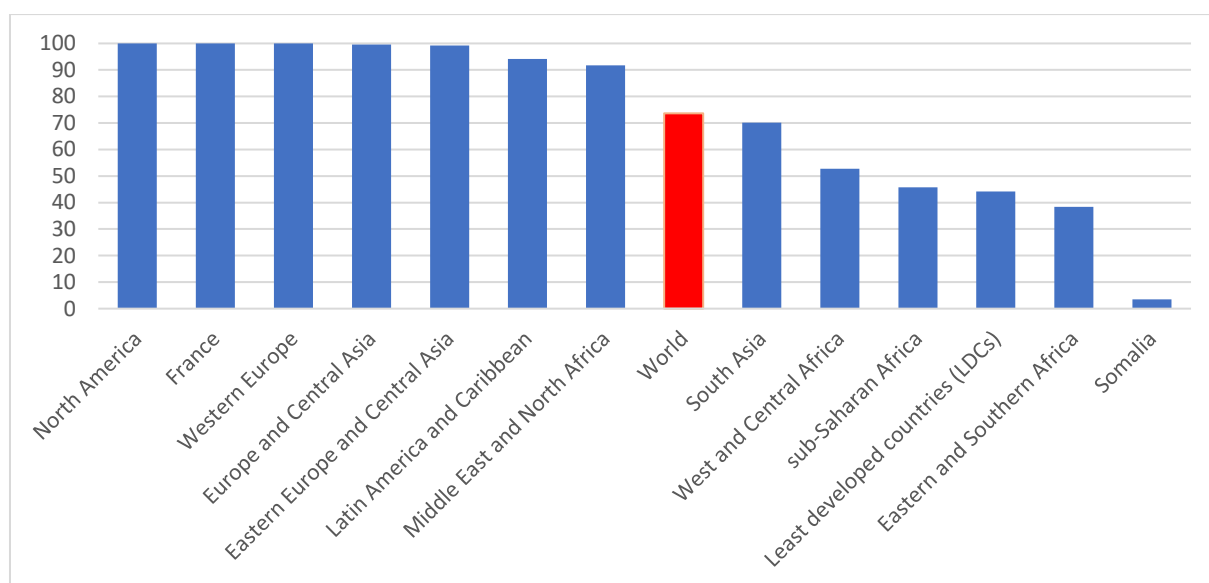


Figure 1: Pourcentage d'enfants de moins de 5 ans dont la naissance est enregistrée. Répartition d'un échantillon de régions et de pays par rapport à la population mondiale.¹⁴

Sachant que l'acte de naissance donne une existence légale à l'individu et lui permet d'accéder aux droits qui en découlent, l'Organisation des Nations Unies a ratifié dans les droits des enfants **l'enregistrement de la naissance comme un droit inaliénable**.¹⁵ Pour les individus n'ayant pas d'acte de naissance, l'estimation de l'âge est donc une des étapes pour obtenir une existence légale.

L'âge chronologique conditionne les droits et devoirs au cours de l'existence d'une personne. Si l'âge chronologique est « évident » pour un Européen, il n'est pas attribué pour 1 milliards de personnes dans le monde. Des raisons culturelles et surtout une absence d'enregistrement et de déclaration des naissances dans certains pays aboutissent à une absence d'âge chronologique d'un enfant sur quatre de moins de cinq ans.

1.2. L'estimation de l'âge chronologique, une question qui remonte à l'antiquité

L'estimation de l'âge chronologique a été décrite dès la période de l'empire romain, il y a 2400 ans, dans le système juridique du droit Romain. Les âges clés de 12 ans pour une fille et 14 ans pour un garçon étaient fixés comme l'âge minimum du mariage et celui de la responsabilité financière. Ces âges pouvaient être estimés par *l'inspectio corpus* qui évaluait aussi bien la maturité physique que mentale de l'individu.^{3,16} L'observation de l'éruption des deuxièmes molaires servait à évaluer l'aptitude au service militaire.¹⁷ En Grèce, Hippocrate avait décrit 7 phases de vie de 7 années (hebdomad) à l'exception de la dernière. Le passage du premier au deuxième hebdomad pouvait être validé à partir d'un indicateur biologique constitué de l'éruption des premières dents permanentes dans la cavité buccale.¹⁶

Dans sa publication, "the Teeth, a test of age", adressée aux membres des deux chambres du parlement en Grande Bretagne, Edwin Saunders revient sur la problématique de l'utilisation des machines à vapeurs lors de la révolution industrielle, qui ont permis d'étendre le temps de production sur la totalité d'une journée.¹⁸ Selon Saunders, les employés étant réfractaires au travail de nuit, les industriels trouvèrent une alternative avec le travail des enfants.¹⁸ Il relate dans cet article des faits observés par Sir Robert Peel :

"Children of the tender age of seven years, and, in some cases, still younger, are frequently admitted to work for thirteen and fourteen hours a-day."

Sir Robert Peel fut à l'origine de la première loi régulant le travail des enfants. En 1819, avec l'application du Cotton Mills and Factories Act, il devenait illégal d'employer un enfant de moins de 9 ans dans une usine de coton, ou de permettre à une jeune personne de moins de seize ans de travailler plus de douze heures par jour. Dans son article de 1837, Saunders propose d'utiliser l'âge dentaire pour remplacer l'utilisation de l'âge physique basé sur la stature afin de lutter contre la fraude liée à l'âge chronologique.¹⁸

Cet article illustre parfaitement l'importance et la problématique de l'estimation de l'âge. **En l'absence de documents prouvant l'âge chronologique ou en cas de doute sur l'âge déclaré par un individu ; comment, pourquoi et qui peut inférer un âge chronologique ?**

1.3. Estimation de l'âge : données actuelles

Actuellement l'estimation de l'âge peut être réalisée sur des individus vivants et morts^{3,8,19,20} Pour les individus décédés, l'estimation de l'âge constitue un des paramètres du profil biologique avec la stature, le sexe et l'ascendance.⁸ Il intéresse les disciplines de médecine légale et d'anthropologie.^{3,8,20} L'estimation de l'âge chez les individus vivants est une technique relativement récente qui a connu un essor avec l'augmentation des flux migratoires ces dernières années.^{3,19,21} Chez les vivants, l'estimation de l'âge peut servir soit à une visée médicale, soit à une visée légale.

Au niveau médical, l'estimation de l'âge d'un individu vivant permet de le situer sur sa courbe de croissance ou de sénescence.²² Dans ce cas, l'estimation de l'âge correspond à l'estimation d'un âge biologique ou physiologique qui communique une information sur l'avance ou le retard de maturité ; ou de sénescence.^{20,22,23} Les disciplines d'orthodontie, de pédodontie, de pédiatrie et de chirurgie orthopédique évaluent *l'âge biologique* pour identifier le niveau de croissance d'un enfant ou d'un adolescent.²⁴

Au niveau légal, l'estimation de l'âge est nécessaire en cas d'absence, perte ou vol de papiers d'identité et lorsqu'il existe de « sérieux doutes » sur la déclaration de l'âge.^{3,8,10,25} Contrairement à l'âge biologique dont la progression est plus ou moins rapide au cours de la vie d'un individu, l'âge légal, basé sur l'âge chronologique, à une progression constante. Pour des questions légales, il existe plusieurs méthodes pour estimer l'âge chronologique. La première consiste à retrouver des traces administratives. La seconde consiste à corrélérer l'âge biologique avec l'âge chronologique.

Un rapport de la European Union Agency for Asylum (EUAA anciennement European Asylum Support Office : EASO) sur la détermination de l'âge en Europe en 2013 présente les droits et responsabilités associées aux limites d'âges légal en Europe. Ils peuvent inclure de manière non exhaustive « les procédures d'immigration, l'âge auquel le mariage est possible, le consentement aux relations sexuelles, le consentement ou refus des soins médicaux, l'entrée dans l'armée ou conscription, la participation à des compétitions sportives, les restrictions d'âge minimum concernant l'emploi, l'âge de la responsabilité pénale et les garanties procédurales pour les enfants dans le cadre des systèmes de justice pénale ». ²⁵ La période la plus critique reste la détermination de l'âge concernant les mineurs et jeunes adultes dont les naissances de 625 millions d'enfants de 14 ans n'ont pas été enregistrées.¹³

A l'échelle mondiale, la **Convention des Droits de l'Enfant de 1989** rédigée par l'UNICEF a permis aux états membres de s'accorder sur des principes et des définitions pour réaffirmer le fait que les enfants ont besoin d'une protection et d'une attention particulières en raison de leur vulnérabilité. L'article

premier définit l'enfant comme tout être humain de moins de 18 ans. L'article 7 précise que l'identité et la date de naissance d'un enfant doivent être enregistrées dès sa naissance. L'article 8 oblige les états à fournir le plus rapidement possible une identité et un âge à l'enfant si ce dernier en est dépourvu. L'article 22 encourage les Etats signataires à fournir l'aide nécessaire pour retrouver l'identité et l'âge des enfants réfugiés.¹⁵ L'UNICEF ne fournit pas de conduite à tenir sur les méthodes de détermination de l'âge et laissent aux Etats Membres le soin d'appliquer les leurs en respectant les principes clés de « non-discrimination », de « l'intérêt supérieur de l'enfant », de « la préservation de l'identité » et de « la protection contre toute forme de violence ». Le Haut-Commissariat des Nations unies aux Droits de l'Homme souligne que la procédure de détermination de l'âge ne doit pas être initiée sans raison et que l'enfant doit être informé de l'âge attribué et du motif de cette procédure.

Au niveau l'Union Européenne, le **rapport de l'EUA sur la détermination de l'âge en Europe en 2013** a identifié les pratiques et les principes généraux de l'estimation de l'âge pour un objectif légal commun aux Etats membres de l'Union Européenne et de la Norvège, de la Suisse, de l'Australie, du Canada, de la Nouvelle-Zélande et des Etats-Unis (34 pays).

Tableau 1: Aperçu de la pratique des 34 pays ayant répondu au questionnaire sur les pratiques d'estimation de l'âge.²⁵

Examens des documents présentés	29 pays
Radiographie du canal carpien (poignet/main)	23 pays
Entretiens oraux de détermination de l'âge	22 pays
Radiographie dentaire	17 pays
Radiographie de la clavicule	15 pays
Observations dentaires	14 pays
Apparence physique	12 pays
Observation de la maturité sexuelle	8 pays
Tests psychologiques	5 pays
Autres formes de détermination : notamment radiographie de l'os iliaque, radiographie de l'épaule droite, radiographie de la colonne vertébrale et examen de l'ADN	3 pays

Le Tableau 1 montre qu'il n'y a pas de méthode commune pour l'estimation de l'âge. Le rapport précise que dans le cadre de l'évaluation de l'âge : « 27 pays utilisent au moins 3 méthodes ; 28 pays utilisent une combinaison de méthodes médicales et non médicales ; 8 pays n'utilisent qu'une combinaison de méthodes non médicales ; et 3 pays n'utilisent qu'une combinaison de méthodes médicales ».²⁵ Les méthodes non médicales regroupent les entretiens oraux, les examens des preuves documentaires et l'évaluation de l'apparence physique et du comportement. Les méthodes médicales regroupent

l'observation de la dentition, l'évaluation du développement physique par un pédiatre, les entretiens et tests psychologiques, les examens de maturité sexuelle et les examens utilisant des rayons ionisants.

En France, plusieurs changements ont eu lieu dans l'évaluation de l'âge des mineurs. En 2019, un nouveau **guide des bonnes pratiques pour l'évaluation de l'âge** a été publié de manière pluri-partenaire avec le ministère de la Justice, le ministère des Solidarités et de la Santé, le ministère de l'Intérieur et le ministère de la Cohésion territoriale et des Relations avec les collectivités locales et territoriales. La méthodologie repose sur un faisceau d'indices qui permet de conclure ou non sur la minorité d'un individu. Les indices peuvent provenir de l'examen des documents d'identité en possession de l'individu, du concours de la préfecture en matière administrative, des entretiens d'évaluation sociale et en dernier recours des examens médicaux.²⁶ **L'article 388 du code civil** précise l'utilisation des examens médicaux dans ce cadre. Les examens radiologiques osseux pour la détermination de l'âge sont envisageables « en l'absence de documents d'identité valables et lorsque l'âge allégué n'est pas vraisemblable, [Ils] ne peuvent être réalisés que sur décision de l'autorité judiciaire et après recueil de l'accord de l'intéressé. Les conclusions de ces examens, qui doivent préciser la marge d'erreur, ne peuvent à elles seules permettre de déterminer si l'intéressé est mineur. Le doute profite à l'intéressé. » L'article 388 précise également l'impossibilité de réaliser l'examen médical visant à évaluer l'âge à partir d'un examen du développement pubertaire des caractères sexuels primaires et secondaires. Le **Conseil constitutionnel** a, par sa décision n° 2018-768 QPC du 21 mars 2019, établi que seul le procureur de la République ou le juge des enfants peuvent faire la demande d'un recours à un examen médical. Il exclut également que les conclusions d'un examen radiologique puissent constituer l'unique élément du faisceau de preuves pour déterminer l'âge de la personne. Enfin le **Haut conseil de la santé publique** a émis un avis relatif à l'évaluation de la minorité d'un jeune étranger isolé le 23 janvier 2014. Il rappelle qu'aucune méthode médicale ne peut donner un âge chronologique précis. Il précise que les examens médicaux doivent être réalisés dans une unité hospitalière de médecine légale par un médecin expert qui déterminera une probabilité d'atteinte d'un âge clé.²⁶

Enfin quand la méthode d'estimation de l'âge médicale est choisie, des recommandations, faites par le groupe d'étude sur le diagnostic médico-légal de l'âge pour l'estimation de l'âge des personnes vivantes dans les procédures pénales (AGFAD) sont disponibles.²⁷ Ces recommandations précisent la séquence des examens et les moyens de présenter les résultats en vue d'une utilisation dans un cadre légal.

L'estimation de l'âge est une pratique établie remontant à 2400 ans. Elle répond aujourd'hui, à un besoin particulier chez l'enfant et le jeune adulte victime de l'absence de déclaration des naissances. Si l'encadrement légal de la pratique est différent selon les pays, elle est réalisée dans le respect des grands principes des droits de l'Homme et de l'Enfant. En France, le recours à l'estimation de l'âge par une méthode médicale est réalisé à la suite d'une demande de l'autorité judiciaire.

1.4. Les méthodes médicales de l'estimation de l'âge chronologique

L'estimation de l'âge chronologique d'un individu peut être effectuée à partir de méthodes médicales ou non médicales. Ces dernières s'appuient principalement sur l'examen des preuves documentaires mais aussi des entretiens oraux. Les examens médicaux sont envisageables en l'absence de documents d'identité valables ou si l'âge allégué semble invraisemblable.²⁶ L'objectif des méthodes médicales est de déterminer un âge biologique le plus proche possible de l'âge chronologique d'une personne.^{3,20} Pour cela un prédicteur biologique de l'âge est défini comme un biomarqueur corrélé à l'âge chronologique.²² Deux approches sont possibles dans l'estimation de l'âge : soit un âge est prédit comme valeur numérique (régression), soit l'individu est positionné par rapport à un âge clé (classification).²⁸

Les méthodes médicales d'estimation de l'âge sont spécifiques pour les enfants et les adultes.²⁹ Pour les enfants, l'estimation médicale de l'âge est basée sur le phénomène de croissance ou développement des différentes structures (osseuses, dentaires, sexuelles)^{17,20,30} Pour les adultes, l'estimation médicale de l'âge est basée sur le phénomène de sénescence observé par des méthodes morpho-histologiques, radiologiques ou biochimiques.³¹ Il existe une zone intermédiaire qui comprend les adolescents et les jeunes adultes pour laquelle les âges de 16, 18 et 21 ans sont utilisés comme âges clés au niveau administratif et légal.

Ce travail s'est focalisé sur **les individus vivants et la tranche d'âge concernant les enfants et les jeunes adultes.**³² Il existe trois méthodes pour estimer l'âge biologique :^{17,20}

- L'examen physique,
- L'évaluation de la maturation squelettique,
- L'évaluation la croissance dentaire.

1.4.1. L'examen physique

L'examen physique comprend tout d'abord les mesures du poids et de la taille. Ces mesures peuvent être confrontées aux percentiles spécifiques concernant la croissance standard des enfants.³³ L'examen physique comprend l'évaluation du développement sexuel. La méthode de Tanner, utilisée par les pédiatres, évalue chez les femmes le développement des seins et des poils pubiens et chez les hommes l'augmentation du volume des testicules et des poils pubiens.²⁰ En France, l'évaluation de l'âge à des fins médico-légales à partir d'un examen pubertaire des caractères sexuels primaires et secondaires est interdite par l'article 388 du Code Civil.²⁶ En outre, Tanner lui-même a précisé que sa méthode ne visait pas à établir un âge chronologique mais un indicateur de développement des caractères sexuels.³⁴

1.4.2. L'évaluation du développement squelettique

Quelques mois seulement après la découverte des rayons X par Röntgen, la possibilité de les utiliser pour estimer l'âge osseux dans un contexte médico-légal est évoquée par Von Ranke. Au cours d'une conférence de la société médicale de Munich, il a proposé que l'âge des enfants puissent être évalué à partir d'une radiographie de la main.¹⁹ Il est intéressant de noter que Von Ranke a voulu utiliser cette méthode personnellement pour vérifier l'âge d'un enfant polonais pianiste prodige qu'il suspectait de mentir sur sa date de naissance.¹⁹ Les méthodes basées sur le développement squelettique peuvent évaluer l'ossification des différentes parties de l'os ; l'aspect morphologique et/ou la taille de l'os ; le moment de la fusion des différentes parties d'un os. Les méthodes les plus utilisées sont celles basées sur la radiographie de la main et du poignet : l'atlas de Greulich et Pyle et la méthode de Tanner-Whitehouse.²⁰

1.4.3. L'évaluation du développement dentaire

L'estimation de l'âge dentaire chez un individu vivant est réalisée à partir de méthodes non destructives et non invasives. Le **développement dentaire** comprend deux parties ; la **formation** et **l'éruption** des dents. La formation des dents correspond aux phénomènes de mise en place puis de minéralisation des tissus constitutifs d'une dent : émail, dentine et cément. L'éruption est le mouvement continu de la dent, allant de l'os alvéolaire jusqu'à l'arcade antagoniste et se poursuit ensuite pour compenser les effets de l'usure.³⁵ Chez l'enfant et le jeune adulte, il peut être évalué par une méthode physique clinique ou radiologique.³⁰

Les méthodes cliniques sont basées sur l'évaluation de la **séquence d'éruption** des dents temporaires ou permanentes dans la cavité buccale. L'émergence d'une dent est un évènement particulier de l'éruption dentaire, qui correspond au moment où la dent perce la gencive.³⁵ Les principales méthodes d'évaluation du développement dentaire sont basées sur des examens radiologiques.^{30,35} Il s'agit des radiographies panoramiques, des radiographies rétro alvéolaires ou des examens de tomodensitométrie de type scanner ou CBCT.³⁶ L'examen radiologique le plus utilisé est la radiographie panoramique qui a un bon compromis dose d'irradiation, champ d'exploration. Il existe une autre méthode plus marginale d'imagerie non irradiante, mais plus contraignante en terme d'accès: l'imagerie par résonance magnétique.^{30,37,38}

Le développement dentaire est évalué par des méthodes radiologiques de type Atlas, des méthodes de mesures directes ou par des méthodes d'évaluation subjective des stades de formation des dents.^{29,30,35,39} L'étude du développement dentaire a d'abord été réalisée à partir de techniques basées sur des dissections et des coupes histologiques.⁴⁰ L'utilisation de la radiographie pour l'estimation de l'âge dentaire a été beaucoup plus tardive que pour l'âge squelettique. Ce n'est qu'en 1941, que Schour et Massler ont utilisé la radiographie ou « roentgenogram intra oral » pour produire un atlas d'estimation de l'âge dentaire.⁴¹ Depuis, un grand nombre de méthodes radiologiques analysant le développement dentaire à partir de la minéralisation de l'organe dentaire ont été décrites. Voici une liste non-exhaustive des principaux auteurs: Gleiser and Hunt (1955)⁴²; Nolla (1960)⁴³; Moorrees, Fanning and Hunt (1963)⁴⁴; Wolanski (1966)⁴⁵; Demirjian, Goldstein and Tanner (1973)⁴; Gustafson and Koch (1974)⁴⁶; Haavikko (1974)⁴⁷; Anderson, Thompson and Popovich (1976)⁴⁸; Demirjian and Goldstein (1976)⁴⁹; Carels et al. (1991)⁵⁰; Mørnstad, Staaf and Welanders (1994)⁵¹; Liversidge and Molleson (1999)⁵²; Teivens and Mørnstad (2001)⁵³; Willems *et al.* (2001)⁵⁴; Blenkin (2009)⁵⁵; and Cameriere, Ferrante and Cingolani (2006)⁵⁶.

1.4.4. Les méthodes d'atlas

La **méthode d'atlas ou schémas de développement dentaire** de Schour et Massler a été la première des méthodes utilisées pour estimer l'âge dentaire.⁴⁰ En 2010, le London Atlas a été créé par AlQahtani *et al.*⁵⁷ Il utilise une combinaison d'éléments basés sur la formation et l'éruption dentaire pour évaluer l'âge dentaire d'individus dont l'âge chronologique est compris entre 23 semaines in utero et 23 ans. Cet atlas permet de classer les individus (Figure 2) sur des intervalles de 3 mois au cours de la première année, puis des intervalles d'un an jusqu'à 16 ans, puis des intervalles supérieurs à 1 an jusqu'à 23 ans. Les avantages de la méthode d'atlas sont sa rapidité et sa simplicité d'utilisation, puisqu'il s'agit de comparer une radiographie panoramique et de trouver le schéma qui correspond

sur l'atlas.²⁹ Les limitations de ces méthodes sont les suivantes. Ce sont les intervalles d'âges entre chaque schéma qui fixe la précision maximale de la méthode. Ainsi pour les intervalles d'un an, la fourchette autour de l'âge moyen ne peut être que de six mois. Ensuite la deuxième limitation porte sur la difficulté d'estimation de certains stades d'éruption avec une radiographie panoramique et avec des dents non alignées d'un quadrant à l'autre.^{29,58}

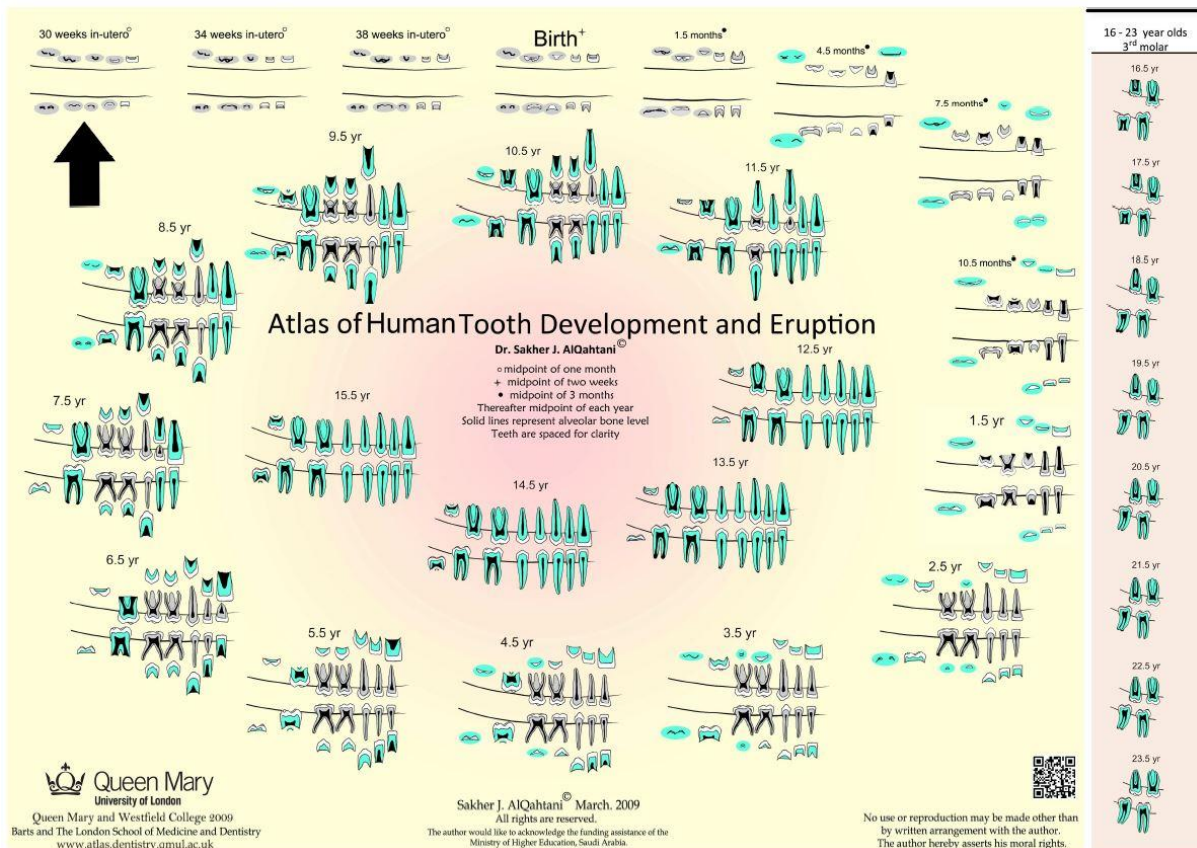


Figure 2: London Atlas de la formation et de l'éruption dentaire⁵⁷

1.4.5. Les méthodes de mesures directes

Les méthodes radiologiques de mesures directes sont basées sur des mesures linéaires, des calculs d'aires ou de volume. Pour **les mesures linéaires**, différents « landmarks » sont utilisés en fonction des méthodes, comme les sommets des apex ouverts des racines, les sommets des cuspidés ou la jonction amélo-dentinaire mésiale et distale.³⁹ Des mesures entre ces landmarks sont ensuite réalisées et permettent grâce à des équations de régression de donner une estimation de l'âge dentaire.^{50,51,56} Les méthodes utilisant les mesures linéaires peuvent être utilisées chez les enfants et les adultes pour l'estimation de l'âge.⁵⁹ Elles peuvent aussi être utilisées pour classer les individus par rapport à un âge clé en s'appuyant sur le développement de la troisième molaire.⁶⁰ En 2006, Cameriere

a proposé une méthode mesurant sur une radiographie panoramique l'ouverture des apex et la longueur des 7 dents mandibulaires gauche permanentes.⁵⁶ Une équation de régression permet d'estimer l'âge dentaire à partir de ces mesures. Une méta analyse de 2021 sur l'utilisation de la méthode pour l'estimation de l'âge chez des enfants a montré une différence moyenne entre l'âge chronologique et l'âge dentaire chez les garçons et les filles de respectivement 0.44 ans (0.26-.063) et 0.34 ans (0.19-0.49) pour des individus âgés de 6 à 16 ans.⁵⁹

Il existe aussi des **calculs de surface ou de volume** qui permettent d'estimer l'âge dentaire à partir du phénomène d'apposition de dentine secondaire.^{61,62} Ces méthodes sont basées sur la segmentation de surfaces ou de volumes de la pulpe et de la dent issues respectivement de radiographies 2D ou 3D. Le ratio permet d'estimer l'âge dentaire à partir d'une équation de régression. La dent choisie pour ces calculs est généralement la canine maxillaire ou mandibulaire.⁶² Ces méthodes peuvent s'utiliser chez les individus adultes. Pour le calcul des surfaces, des radiographies péri-apicales ou panoramiques sont utilisées.^{61,63} Le calcul des volumes est obtenu à partir d'examen de tomodontométries.^{36,62}

1.4.6. Les méthodes observationnelles

La plupart des méthodes radiographiques d'estimation de l'âge sont basées sur des observations par rapport à un schéma dentaire.^{29,64} Ces méthodes sont basées sur l'étude originelle de **Gleiser et Hunt** (1955) dont l'objectif a été de suivre pendant plusieurs années la première molaire mandibulaire permanente droite.⁴² Des radiographies rétro alvéolaires ont été réalisées tous les 3 mois de la naissance à 18 mois, puis tous les 6 mois jusqu'à 10 ans chez 25 filles et 25 garçons. Les auteurs ont réalisé une série de **15 croquis** de contours à partir des images radiographiques correspondant à 15 stades de calcification de la **première molaire mandibulaire droite** (Figure 3)

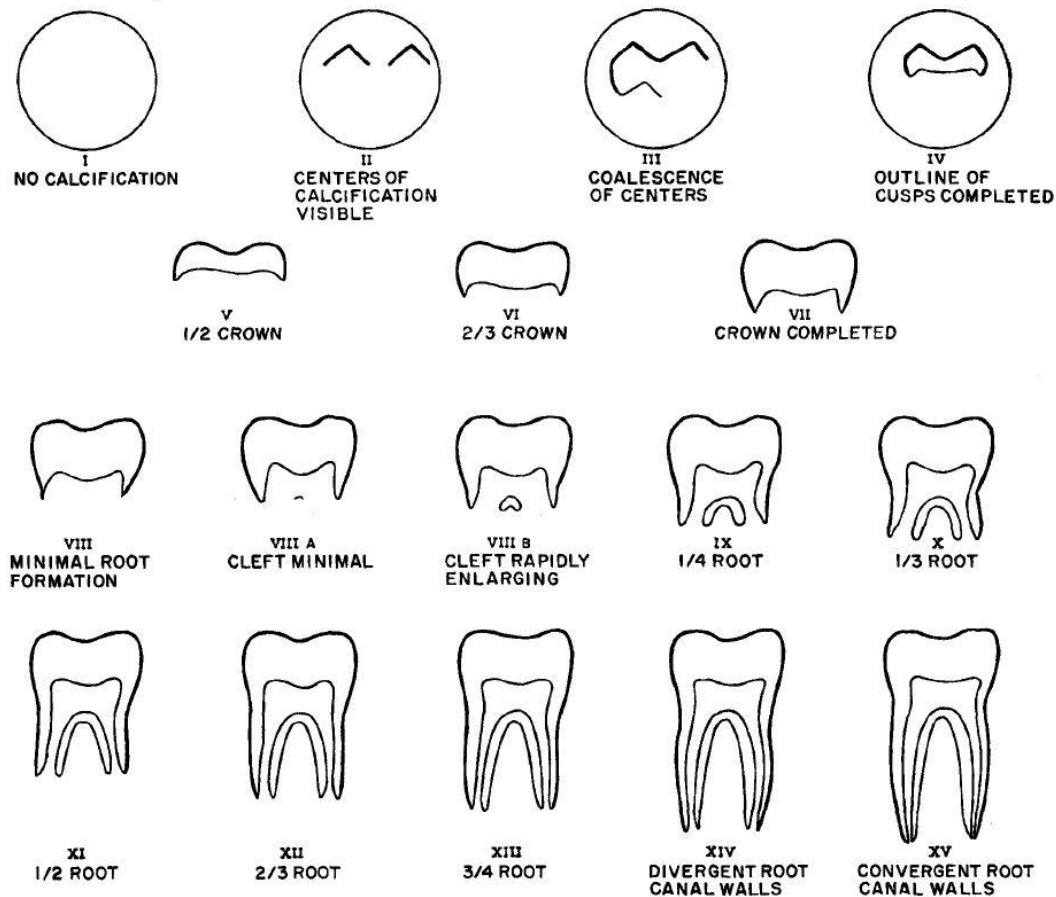


Figure 3: Schéma des 15 stades de calcification de la première molaire mandibulaire permanente Gleiser et Hunt.⁴²

A partir de cette étude, Moorrees, Fanning et Hunt (1963)⁴⁴ ont adapté la méthode aux 8 dents maxillaires et mandibulaires gauche avec 14 stades de développement identifiés. Un autre groupe de travail mené par **Demirjian, Goldstein et Tanner** (1973)⁴ a simplifié la méthodologie précédente en analysant les **7 dents mandibulaires permanentes gauche** à partir de **8 stades** de développement de A à H. L'intérêt réside dans l'association des vues radiographiques, des schémas et des descriptions écrites des 8 stades qui clarifient et facilitent les choix d'un stade (Figure 4).⁴⁹ Pour chaque dent, le stade de maturité correspond à un score numérique de maturité. La somme de ces scores donne le score de maturité final qui est reporté sur une courbe de percentile en fonction du sexe de l'individu pour trouver l'âge dentaire correspondant. Cela permet de situer l'individu sur sa courbe de croissance par rapport à son âge chronologique. *Willems et al.* ont utilisé le même système de notation des stades de développement dentaire tout en exprimant non pas un score de maturité global mais directement un âge exprimé en années.⁶⁵

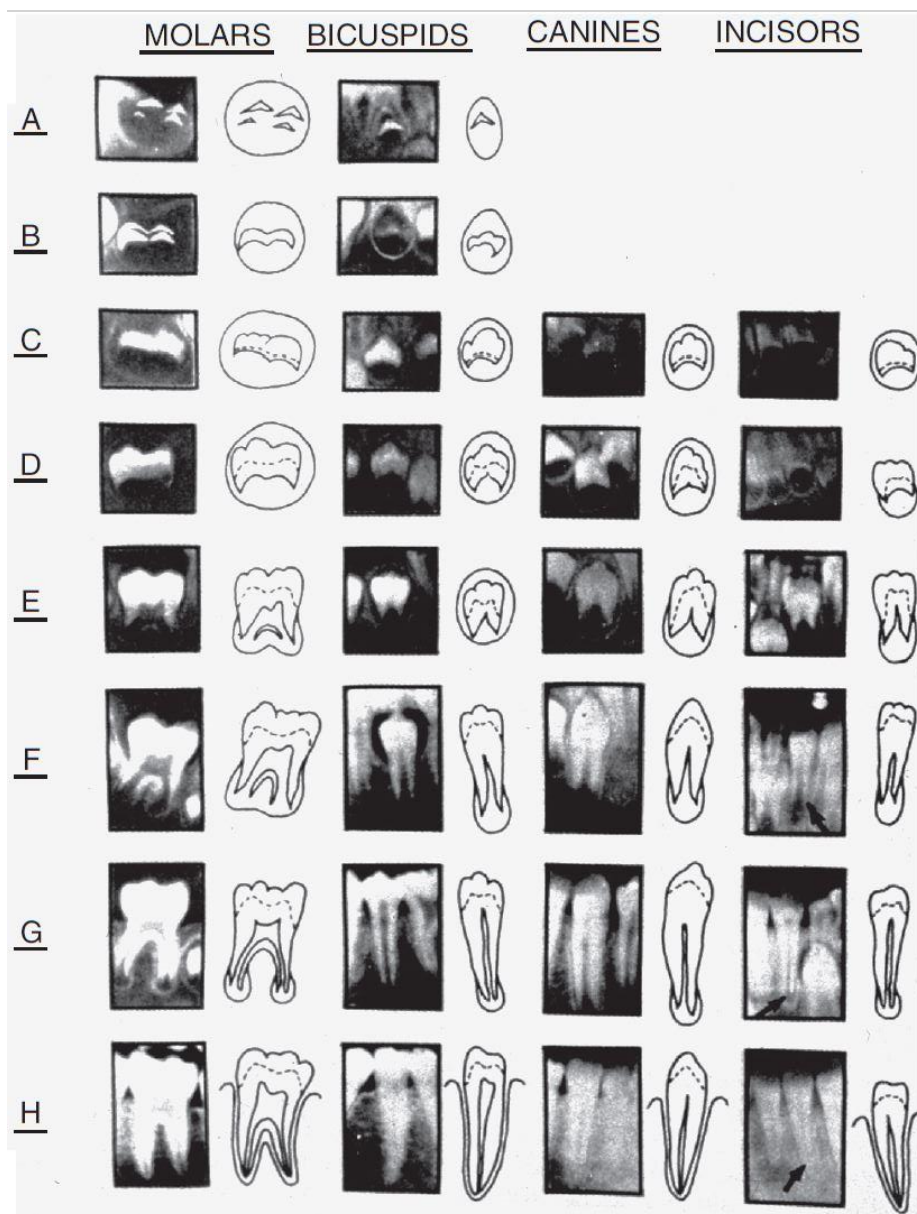


Figure 4: Présentation système de Demirjian permettant de déterminer les 8 stades de développement des dents permanentes mandibulaire.⁴

Une des méthodes les plus utilisées pour l'estimation de l'âge dentaire est celle de **Demirjian**.⁶⁶ Jusqu'en 2020, plus de 150 études évaluant la précision de cette méthode ont été réalisées sur différentes populations.⁶⁷ Malgré sa tendance à surestimer l'âge dentaire, c'est une méthode simple à utiliser et dont la concordance inter-observateurs est excellente à 90 %.^{6,29,67,68} En 2013, Jayaram *et al.* ont mené une méta analyse sur 12 études. Ils ont montré une surestimation moyenne de 0.65 ans chez les filles et 0.60 ans chez les garçons.⁶⁸ La même année, une autre méta analyse a été menée par Yan *et al.* à partir de 26 études. Ils ont montré une meilleure précision de la méthode avec des

surestimations de l'âge pour les filles et les garçons respectivement de 0.39 et 0.35 ans.⁶⁹ En 2017, Esan *et al.* ont réalisé une méta analyse sur 28 études. Ils ont montré une surestimation plus importante de l'âge chronologique pour les filles et les garçons respectivement de 0.72 et 0.62 ans.⁶ Récemment, Hostiuc *et al.* ont montré dans une méta analyse basée sur 89 études une surestimation de l'âge chronologique d'une demie année pour les filles et les garçons.⁶⁷

Pour remédier à cette surestimation de l'âge de la méthode de Demirjian basée sur un échantillon d'origine canadienne-française, **Willems** *et al.* ont proposé d'utiliser de nouveaux scores pour chaque stade à partir d'une population caucasienne Belge.⁵⁴ En 2017, 3 méta analyses ont montré que les scores déterminés par Willems permettent une estimation de l'âge plus précise chez les enfants. Wang *et al.* à partir de 11 études ont montré une surestimation de 0.06 et 0.18 ans respectivement chez les filles et les garçons.⁶⁶ Sehwat *et al.* à partir de 15 études ont montré une surestimation de 0.04 et 0.02 ans respectivement chez les filles et les garçons.⁷⁰ Enfin Esan *et al.* ont montré à partir de 28 études une surestimation plus importante que les deux précédentes.⁶ Dans cette méta analyse l'âge est surestimé de 0.29 et 0.26 ans respectivement chez les filles et les garçons.

1.5.Problématiques de l'estimation de l'âge dentaire

Il n'existe pas de méthode universelle d'estimation de l'âge dentaire.⁶ Les variations de précision au sein de populations différentes ou entre des études ont des causes multiples. Elles peuvent être liées à des paramètres influençant directement la maturité dentaire ou à des raisons méthodologiques.

On peut distinguer des facteurs intrinsèques et extrinsèques à l'individu pouvant impacter le développement dentaire. Parmi les facteurs intrinsèques, on distinguera tous les facteurs portant sur les variations génétiques : la variabilité génétique, le sexe et les variations génétiques liées aux effets du climat ou des migrations à grande échelle. Ce dernier paramètre est la cause de variations géographiques régionales ou à grande échelle.⁷¹ Un facteur endocrinien peut également être la cause de variation de maturation des dents.²⁴ Les facteurs extrinsèques sont les facteurs environnementaux liés à l'écologie humaine, c'est-à-dire le statut socio-économique, la nutrition et les habitudes alimentaires.^{24,68} Il est intéressant de noter qu'avec le temps l'emprise génétique sur le développement dentaire décroît au profit des facteurs environnementaux.⁷¹ Parmi tous ces facteurs, les auteurs s'accordent sur la difficulté à obtenir les informations, à l'exception du sexe. En effet, l'origine géographique des individus est difficile à obtenir aujourd'hui avec les mouvements des

populations. Dans leur méta-analyse, Jayaraman et al. ont mis en évidence l'existence d'un grand degré de variation sur des populations ethniquement similaires et proches géographiquement ; et peu de différences entre deux populations ethniques différentes sur le même territoire.⁶⁸ Il apparaît dans certaines études que l'aspect environnemental aurait un impact plus important que l'origine régionale. Cela souligne l'applicabilité limitée de ces données de référence.

La définition d'un âge biologique, estimation d'un âge chronologique, pourrait également revêtir un intérêt dans la mesure d'une déviation d'un individu par rapport à une norme. Par définition, l'éruption de la dent implique le temps long et peut être ainsi « perturbée » tout au long du processus par des facteurs endogènes et exogènes. En d'autres termes, cet âge biologique dentaire pourrait être un biomarqueur prometteur, reflet du temps long de l'exposition d'un individu à différents facteurs. Il pourrait être intéressant dans des études de cohorte de suivre ces individus et d'évaluer la capacité prédictive d'un tel biomarqueur au développement par exemple de pathologies chroniques.

Les différences observées pour l'estimation de l'âge dentaire peuvent aussi être de nature méthodologique. Les stades de développement sont des marqueurs biologiques discrets qui sont corrélés à un âge chronologique continu.²⁴ Il existe donc un équilibre à trouver entre le nombre de stades et les possibilités pour un observateur de distinguer ces stades. Au niveau statistique ensuite, les approches utilisées par les méthodes radiologiques de Demirjian ou Willems sont issues de régressions linéaires. Cependant l'utilisation d'une régression linéaire implique l'hypothèse d'avoir pour chaque stade de développement radiologique une distribution normale et une variance constante.⁷² Les modèles de Demirjian et Willems excluent la possibilité d'une vitesse de maturité dentaire variable.⁷³ L'autre point statistique problématique pour l'utilisation d'une régression statistique porte sur les variables indépendantes et dépendantes. Dans le cadre de l'estimation de l'âge chronologique à partir de radiographies panoramiques, les variables indépendantes correspondent à chaque dent de l'individu pour laquelle différents stades de développement radiologiques sont identifiés. La variable dépendante subit l'effet des variables indépendantes et correspond ici à l'âge dentaire du sujet. Or plusieurs auteurs ont souligné que les dents permanentes en développement ne donnent pas d'information indépendante sur l'âge.^{24,28,72,74} Braga propose de prendre en compte la séquence de minéralisation de plusieurs dents plutôt que d'utiliser 7 stades de développement dentaire radiologique indépendants.²⁴

Pour pallier ces problèmes statistiques, plusieurs auteurs ont proposé d'utiliser une approche bayésienne.^{21,24,28,74,75} Une autre possibilité consiste à utiliser l'apprentissage machine pour créer un modèle statistique optimisé à partir de données d'apprentissage. Dans cette approche les poids des entrées (prédicteurs) du modèle sont dérivées de manière dynamique et itérative à partir d'un

processus d'optimisation du modèle. L'algorithme permet de trouver les relations entre les différents prédicteurs à partir de sorties connues. C'est la phase d'apprentissage. Ensuite la phase de généralisation permet à l'algorithme de réaliser une régression ou une classification. La notion d'apprentissage machine (machine learning : ML) recouvre l'ensemble des méthodes utilisant des données d'apprentissage pour créer un modèle répondant à une problématique. Dans le domaine de l'estimation de l'âge dentaire, prédire un âge dentaire en tant que valeur numérique correspond à la **régression** et prédire si un individu a atteint ou non un âge clé correspond à une **classification**. Les méthodes de ML ont plusieurs avantages. Premièrement elles permettent de formuler des modèles capables de capturer la complexité des relations entre les prédicteurs. Ensuite elles permettent de se passer des tables de correspondances spécifiques à une population dans la mesure où la méthode s'adapte directement à la population étudiée. Elles permettent alors de prédire directement un âge dentaire pour une utilisation à visée médicale ou légale. Enfin les méthodes de ML ont l'avantage d'intégrer plus de prédicteurs que ceux utilisés dans les méthodes classiques d'estimation de l'âge. Cela a pour conséquences les possibilités d'améliorer la précision des méthodes et/ou d'élargir la plage d'âges estimés. Certains auteurs ont déjà comparé des approches de ML aux méthodes de références pour l'estimation de l'âge dentaire.^{73,76} Mais soit l'étude ne portait que sur une seule méthode de ML⁷³, soit plusieurs méthodes de ML étaient testées entre elles pour prédire l'âge dentaire mais sans comparaison avec une méthode de référence.

L'objectif principal de ce travail a été de comparer différentes méthodes de machine learning par rapport à deux méthodes de référence dans l'estimation de l'âge dentaire chez les enfants : Demirjian et Willems. Pour isoler l'efficacité des méthodes de machine learning dans le traitement statistique des informations fournies par les différents prédicteurs, le même système d'annotation a été utilisé pour identifier les stades de maturation dentaire : 8 stades décrits par Demirjian *et al.* L'objectif secondaire a été de tester la possibilité d'ajouter des prédicteurs aux méthodes de machine learning et de les tester entre elles sur un intervalle d'âge élargi aux jeunes adultes. Pour cela les stades de maturations dentaires des dents de sagesse ont été rajoutés.

Ce travail a abouti à la rédaction d'un article paru en 2021 dans l'International Journal of Legal Medicine.

2. Estimation de l'âge dentaire par différentes méthodes de machine learning

Comparaison de différentes approches d'apprentissage automatique pour prédire l'âge dentaire à l'aide de l'approche par stades de Demirjian

Reproduced with permission from Springer Nature – 26/8/2021



Comparison of different machine learning approaches to predict dental age using Demirjian's staging approach

Antoine Galibourg^{1,2} · Sylvain Cussat-Blanc^{3,4,5} · Jean Dumoncel² · Norbert Telmon^{1,2} · Paul Monsarrat^{1,5,6} · Delphine Maret^{1,2}

Received: 28 July 2020 / Accepted: 9 December 2020 / Published online: 7 January 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

Context Dental age, one of the indicators of biological age, is inferred by radiological methods. Two of the most commonly used methods are using Demirjian's radiographic stages of permanent teeth excluding the third molar (Demirjian's and Willems' method). The major drawbacks of these methods are that they are based on population-specific conversion tables and may tend to over- or underestimate dental age in other populations. Machine learning (ML) methods make it possible to create complex data schemas more simply while keeping the same annotation system. The objectives of this study are to compare (1) the capacity of ten machine learning algorithms to predict dental age in children using the seven left permanent mandibular teeth compared to reference methods and (2) the capacity of ten machine learning algorithms to predict dental age from childhood to young adulthood using the seven left permanent mandibular teeth and the four third molars.

Methods Using a large radiological database of 3605 orthopantomograms (1734 females and 1871 males) of healthy French patients aged between 2 and 24 years, seven left permanent mandibular teeth and the 4 third molars were assessed using Demirjian's stages. Dental age estimation was then performed using Demirjian's reference method and various ML regression methods. Two analyses were performed: with the 7 left mandibular teeth without third molars for the under 16 age group and with the third molars for the entire study population. The different methods were compared using mean error, mean absolute error, root mean square error as metrics, and the Bland-Altman graph.

Results All ML methods had a mean absolute error (MAE) under 0.811 years. With Demirjian's and Willems' methods, the MAE was 1.107 and 0.927 years, respectively. Except for the Bayesian ridge regression that gives poorer accuracy, there was no statistical difference between all ML tested.

Conclusion Compared to the two reference methods, all the ML methods based on the maturation stages defined by Demirjian were more accurate in estimating dental age. These results support the use of ML algorithms instead of using standard population tables.

Keywords Dental age · Machine learning · Artificial intelligence · Demirjian

Introduction

Estimated age is an important characteristic of an individual's biological identity and should be differentiated from

the chronological age. In a living individual, chronological age generally corresponds to calendar age, whereas the biological age or physiological age covers the growth period, which corresponds to the maturation process of

Paul Monsarrat and Delphine Maret contributed equally to this work.

✉ Antoine Galibourg
antoinegalibourg@gmail.com

¹ Dental Faculty, Toulouse Institute of Oral Medicine, Paul Sabatier University, Toulouse University Hospital (Hôpitaux de Toulouse), chemin des Maraîchers, 31062 Toulouse Cedex 09, France

² Laboratoire Anthropologie Moléculaire et Imagerie de Synthèse, UMR 5288, Université Paul Sabatier, Toulouse, France

³ Institute of Advanced Technologies in Living Sciences (ITAV), CNRS – USR3505, Toulouse, France

⁴ Institute of Research in Informatics (IRIT) of Toulouse, CNRS – UMR5505, University of Toulouse, Toulouse, France

⁵ Artificial and Natural Intelligence Toulouse Institute ANITI, Toulouse, France

⁶ STROMALab, CNRS ERL 5311, EFS, ENVT, Inserm, UPS, Université de Toulouse, Toulouse, France

different tissues and organs [1], and the aging phase of an individual [2].

Dental age is one of the indicators of maturity, like skeletal age, or changes in the secondary sexual or somatic characteristics. Dental age is used in clinical practice in orthodontics, pedodontics, pediatrics, and orthopedic surgery. It is also widely used in physical anthropology and forensic science, for the identification of victims of mass disasters, to solve criminal, civil, or administrative issues such as adoption, migration, child abuse, legal consent, asylum proceedings, marriage, social benefits, and age estimation in competitive sports [3–7].

In living individuals, dental age estimation based on dental development is applicable during the period from childhood to young adulthood, that is to say, over a period that extends until all teeth have completed their maturation process, which corresponds to a maximum dental age based on maturation. The estimation methods must therefore be as reliable (range and percentage of the confidence limits) and accurate (mean difference between dental and chronological age) as possible [3, 8–10].

Radiological methods for assessing dental age are based on the continuous process of dental mineralization and allow radio-morphological (i.e., Demirjian's or atlas strategies) or radiometric (i.e., technical camera) observations [11, 12]. The number of tooth development stages described varies from 3 to 27 [13]. The challenge, therefore, lies in choosing a method that has enough stages to follow the continuous growth process as closely as possible and stages that are sufficiently differentiated to be easily discriminated by investigators [14]. Demirjian's dental maturity method was the first to combine outline drawings, written descriptive criteria, radiographic examples of each stage, and rules of choice to decide on borderline stages [13]. Examples where dental age is either early or late in relation to chronological age have been documented when this method is applied to other populations [10]. Population sample (regional background, sexual dimorphism, chronological age distribution), cultural or environmental factors (e.g., socio-economic status, nutrition, dietary habits), and also endocrine factors, or even reasons of statistical methodology [3] may explain such differences. The dental maturity scoring system is nevertheless accepted worldwide for its ease of use and for its excellent inter- and intra-examiner consistency [10, 15]. Beside the initial use of only 7 permanent teeth inside the statistical model (first mandibular incisor to second mandibular molar), the methodology was improved by adding other indicators, such as the third molars, to increase the identifiable age range up to young adults [16–19]. Whether the aim is to infer chronological age or to position an individual in relation to an age threshold, population-specific standards rather than a universal standard or methods developed on other populations need to be used [10]. At the present time, many studies assess the accuracy of a reference method on a population different from that used in the original method [18, 20–32]. Although these studies may reveal overestimation or

underestimation of a method carried out on a sample of the reference population, they do not give a tool that is directly usable in forensic science to answer the questions of age estimation or classification. Depending on the populations studied, the DEM method has a tendency to overestimate dental age and can give average differences between chronological age and dental age ranging from -2.82 to $+0.10$ years for females and from -3.04 to $+0.23$ years for males [33]. Compared to the DEM method, the WILL method overestimates to a lesser extent the age of the children [34].

Machine learning (ML) algorithms are part of artificial intelligence that may represent a powerful tool to develop more accurate algorithmic processes to estimate an individual's dental age (regression problem) or define an age range (classification problem). ML is useful to discover patterns on large scale and multidimensional datasets that would have been difficult for individuals, even well trained, to identify [35]. Machine learning corresponds to a wide range of algorithm families and complexities, from easily interpretable models (e.g., linear regression, polynomial regression) to other models requiring development of additional explainability pipelines (e.g., support vector machine, neural network). The models can be supervised (the train dataset is "labeled") or unsupervised (the model works on its own to discover information, dealing with unlabeled data). These techniques therefore make it possible to provide more specific tools for each population and to determine more accurately the equivalent of the standards table for the population [3]. Machine learning methods can be parametrized at several levels: the hyperparameters are parameters that are determined a priori for each model, whereas the parameters are determined a posteriori, after the training stage. If a forensic utilization is envisaged, these parameters must be identified to comply with Ritz-Timme et al. recommendations for age estimation [36]. ML methods are already used for bone age estimation [37] and are starting to develop for dental age [38–40]. ML methods provide a solution to formulate models capable of capturing complexity without the need for correspondence tables while preserving the same annotation system. ML methods are described as being more accurate than conventional radiological methods [38, 40]. These methods provide mean absolute error estimates of dental age that are less than 0.99 years relative to chronological age [38–40]. The advantages of these methods are the absence of the use of conversion tables specific to a given population, the adaptability of the methods for each population studied, the possibility of directly giving an estimate of the dental age for use in forensic science, the possibility of integrating more predictors than those used in the classical methods, and thus of widening the range of estimation of the dental age that can be estimated from children to the young adult population. Although these methods have been evaluated in comparison with classical reference methods [38, 40], no comparison between the use of ML methods on

different populations has yet been made. In this work, supervised ML methods have been used, since the statistical models have been created through optimization of the model deduced from the “known” dataset [38].

The objectives of this study are to compare (1) the capacity of ten machine learning algorithms to predict dental age in children using the seven left permanent mandibular teeth compared to reference methods and (2) the capacity of ten machine learning algorithms to predict dental age from childhood to young adulthood using the seven left permanent mandibular teeth and the four third molars.

Materials and methods

Data collection

A sample of 3605 panoramic radiographs (age ranging from 2.18 to 24 years) of 1734 girls and 1871 boys was retrospectively collected from the clinical radiological database of the Department of Oral Medicine of the Toulouse University Hospital, corresponding to a large sample of the population of Southern France population irrespective of ethnicity. Radiographs of individuals with systemic diseases or developmental anomalies were excluded. Data were treated as cross-sectional, although a proportion of children from each group was represented by more than one radiograph. According to the guidelines provided by Schmeling et al. [9], we have been working to get an even distribution of subjects across all age groups, a balanced number of boys and girls, and a minimum number of subjects 10 times the number of examined features (7 permanent teeth, 4 third molars and gender, i.e., 12 features examined in this study). As a minimum of 10 boys or girls in each category per year of age is required, the 15 individuals of 3 years old of the sample have been excluded [13]. Since the birth date and the date of the radiological examination are known for each subject, the chronological age is calculated as the number of years elapsed between these two dates and recorded as two decimal points.

Ethical considerations

According to the French health regulations and law governing ethics in human research (Public Health Code), retrospective studies based on the exploitation of usual care data do not need to be submitted to an ethics committee but have to be declared or covered by reference methodology of the French National Commission for Informatics and Liberties (CNIL). A collection and computer processing of personal and medical data was implemented to analyze the results of this research. Toulouse University Hospital signed a commitment of compliance to the reference methodology MR-004 of the CNIL (number: 2206723 v0).

Radiograph analysis

During the assessment of all panoramic radiographs, the observers were blinded for the chronological age. The developmental stages of each of the seven left permanent teeth of the mandible, from central incisor to the second molar, were defined according to the eight radiographic stages defined by Demirjian et al. [11]. Tooth formation stages from A to D represent maturation stages of the crown, and stages from E to H represent maturation stages of the roots. When one or more index teeth were absent, its contralateral homologue was staged. The developmental stages of the third molars were assessed according to the same method by which an extra stage was added for the radiolucent bud (stage I) [41]. Then each alphabetical stage was converted into a numerical development score (stage I, 1; stage A, 2; stage B, 3; stage C, 4; stage D, 5; stage E, 6; stage F, 7; stage G, 8; and stage H, 9). When a tooth or its contralateral was absent, the developmental stage was set to zero (absence, 0). The conditions for estimating dental age by Demirjian’s method are only possible for individuals aged between 3 and 16 years. For this reason, a subgroup was identified for individuals under 16 years of age (U16 group) of the total sample (U24 group). To belong to the U16 group, the 7 mandibular teeth must have a maturity score. To belong to group U24, the stages of maturation of the 7 permanent teeth and all the third molars were considered. In this case, at least 4 permanent teeth from central incisor to second molar had to be staged, but all third molars could be absent.

Statistical analysis

Three months after the first assessment of all samples by the first observer (AG, confirmed and experienced dental practitioner), intra-observer reliability was tested by re-examining 50 panoramic radiographs randomly selected from the U24 group. Inter-observer agreement was tested with a second observer (LG, dental hospital intern) by assessing 50 randomly selected panoramic radiographs. The agreements were quantified using weighted Kappa statistics.

Two reference methods were used for dental age estimation on the U16 group: the one described by Demirjian et al. [11] and the one described by Willems et al. [19]. Demirjian’s method (DEM) to infer dental age is a three-step process. Firstly, each stage of the seven permanent teeth is converted into seven numerical biologic weights with the help of the sex-specific conversion table. Secondly, the weights are added together to give a dental maturity score. Thirdly, sex-specific tables of standards are used to convert the dental maturity score into dental age. Willems’ method (WILL) is easier because all numerical biologic weights of the seven permanent teeth were added together to directly give the dental age.

To validate the different ML strategies compared to the reference methods, the U16 and U24 datasets were randomly divided into a training dataset and test dataset in an 80–20% proportion, respectively. Gender was also used as a predictive feature. The ML approaches used training on two sources of information: the gender and the stages allocated by the main observer to the seven lower left permanent teeth and all third molars. The target values were the chronological ages. The following ML supervised regression algorithms were tested [42]: random forest (RF), support vector machine (SVM), decision tree (DT), Bayesian ridge regression (BRR), k-nearest neighbors (KNN), boosting method AdaBoost (ADAB), polynomial regression (POLYREG), and multi-layer perceptron (MLP). A pipeline was implemented for each ML algorithm, with the possibility of adding a variable normalization technique (min-max, standard scaling, or nothing) and a data reduction technique (principal component analysis or nothing). To test the benefit of combining the different techniques of ML mentioned above, stacking (STACK) [43] and voting (VOTE) [44] methods were also tested. The tuning of hyperparameters to obtain the best model was achieved by exploring multiple combinations using GridSearchCV function and a ten-fold cross-validation (to avoid overfitting, ensuring a 20% validation dataset to be used during hyperparameters optimization). The hyperparameters described in Supplementary Table 1 were tuned. To allow comparisons of the different machine learning methods, the entire procedure, starting with the random creation of new training and test sets, was repeated 20 times.

To allow ML methods to be compared with a reference method for individuals between 16 and 24 years of age, the age assessment based on the dental maturity score from the 4

third molars described by Hofmann et al. on a Central European population was carried out [45].

The accuracy of age predictions was measured by five indicators: the coefficient of determination (R^2), mean error (ME; chronological age minus predicted age), root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE). One-way ANOVA was performed to compare the different indicators across the different ML algorithms, and the multiple pairwise comparisons were adjusted using Tukey's HSD test [46].

The Bland and Altman method [47] was used to graphically assess the degree of agreement between chronological and biological age. All experiments were performed using Scikit-Learn 0.22.1 libraries [42] and Python 3.7.3. The significance level was set to .05.

Results

A total of 3570 orthopantomograms (1719 girls and 1851 boys) were finally considered for the group U24, from which 2230 orthopantomograms (1111 girls and 1119 boys) were in the subgroup U16. Age and sex distribution of the datasets are plotted in Fig. 1. The intra- and inter-observer weighted Kappa were 0.96 and 0.92, respectively.

Overall age prediction performance in the U16 age group

The age prediction performances are presented in Table 1. The mean error was obtained with chronological age minus predicted age; indeed, a negative value implies an

Fig. 1 Age and sex distribution for each category of age per year. X-axis represents age groups, 4 indicates all children from 4 to 4.99 years, etc.

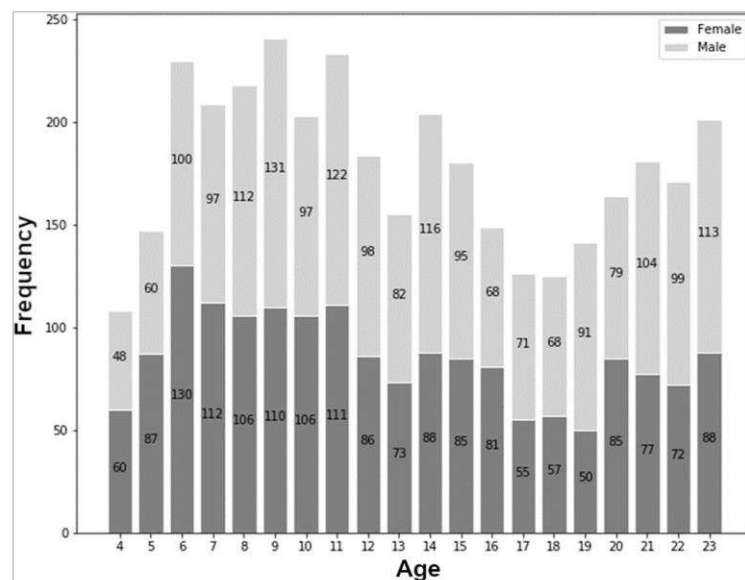


Table 1 Mean error (ME), mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and R^2 values for the U16 group (4–16 years old) assessing performance of machine learning regression methods and reference methods for chronological age estimation

Method	ME \pm SD	MAE \pm SD	MSE \pm SD	RMSE \pm SD	$R^2 \pm$ SD
DEM	-0.705 ± 0.073 b (-0.930 to -0.582)	1.108 ± 0.045 c (1.029 – 1.256)	1.981 ± 0.175 b (1.734 – 2.572)	1.406 ± 0.060 c (1.317 – 1.604)	0.816 ± 0.018 b (0.758 – 0.839)
WILL	-0.220 ± 0.075 c (-0.460 to -0.106)	0.928 ± 0.037 e (0.876 – 1.029)	1.418 ± 0.117 c (1.237 – 1.741)	1.190 ± 0.049 d (1.112 – 1.319)	0.868 ± 0.013 c (0.836 – 0.888)
BRR	-0.002 ± 0.065 a (-0.197 – 0.085)	0.812 ± 0.028 a (0.757 – 0.862)	1.030 ± 0.057 a (0.909 – 1.131)	1.014 ± 0.028 a (0.953 – 1.063)	0.904 ± 0.006 a (0.895 – 0.917)
SVM	0.016 ± 0.068 a (-0.158 – 0.133)	0.729 ± 0.025 d (0.682 – 0.768)	0.901 ± 0.056 a (0.779 – 1.006)	0.949 ± 0.030 b (0.883 – 1.003)	0.916 ± 0.007 a (0.905 – 0.929)
DT	-0.012 ± 0.064 a (-0.174 – 0.106)	0.758 ± 0.034 b, d (0.692 – 0.858)	0.973 ± 0.089 a (0.814 – 1.255)	0.985 ± 0.044 a, b (0.902 – 1.120)	0.910 ± 0.010 a (0.882 – 0.927)
RF	-0.007 ± 0.060 a (-0.183 – 0.110)	0.731 ± 0.025 d (0.683 – 0.782)	0.885 ± 0.054 a (0.775 – 1.008)	0.940 ± 0.029 b (0.880 – 1.004)	0.918 ± 0.006 a (0.905 – 0.930)
KNN	0.009 ± 0.061 a (-0.143 – 0.118)	0.738 ± 0.026 d (0.681 – 0.784)	0.921 ± 0.060 a (0.785 – 1.042)	0.959 ± 0.031 a, b (0.886 – 1.021)	0.915 ± 0.006 a (0.902 – 0.930)
MLP	-0.041 ± 0.099 a (-0.200 – 0.102)	0.742 ± 0.028 d (0.683 – 0.782)	0.907 ± 0.057 a (0.791 – 0.996)	0.952 ± 0.030 b (0.890 – 1.015)	0.916 ± 0.006 a (0.906 – 0.929)
POLYREG	-0.008 ± 0.066 a (-0.191 – 0.106)	0.735 ± 0.024 d (0.692 – 0.783)	0.913 ± 0.058 a (0.792 – 1.031)	0.955 ± 0.031 a, b (0.890 – 1.006)	0.915 ± 0.006 a (0.903 – 0.928)
ADAB	-0.025 ± 0.064 a (-0.225 – 0.075)	0.796 ± 0.026 a, b (0.755 – 0.846)	1.001 ± 0.058 a (0.886 – 1.099)	1.000 ± 0.029 a, b (0.941 – 1.048)	0.907 ± 0.006 a (0.896 – 0.920)
STACK	-0.013 ± 0.066 a (-0.191 – 0.074)	0.733 ± 0.025 d (0.684 – 0.783)	0.904 ± 0.052 a (0.791 – 1.012)	0.950 ± 0.027 b (0.890 – 1.006)	0.916 ± 0.006 a (0.905 – 0.929)
VOTE	0.068 ± 0.292 a (-0.168 – 1.284)	0.770 ± 0.140 a, b, d (0.675 – 1.342)	0.995 ± 0.431 a (0.792 – 2.802)	0.984 ± 0.166 a, b (0.890 – 1.674)	0.908 ± 0.040 a (0.836 – 0.888)

For all metrics, the means \pm standard deviations with (maximum–minimum) of the 20 independent experiments were given. Tukey’s honest significance test for multiple comparison of means was performed at the .05 level. Within the same column, methods sharing the same letter are not significantly different from each other

Abbreviations: ADAB AdaBoost, BRR Bayesian ridge regression, DT decision tree, KNN k-nearest neighbor, MLP multi-layer perceptron, POLYREG polynomial regression, RF random forest, STACK stacking regressor, SVM support vector machine, VOTE voting regressor. Standard deviation (SD) of each method

overestimation of the age estimation method compared to the real age. The two reference methods for estimating dental age, DEM, and WILL were used for U16. The DEM method significantly overestimates dental age more than the WILL method (-0.71 ± 0.07 and -0.22 ± 0.08 , respectively). For all other metrics, the WILL method was more accurate than the DEM method.

All tested machine learning methods were significantly more accurate than the two reference methods for all metrics. ADAB and BRR achieved the lowest performance for MAE (Table 1, Fig. 2a).

Overall age prediction performance in the U24 age group

The age prediction performances are presented in Table 2 and Fig. 2b. The SVM and MLP methods result in a slight overestimation of dental age compared to chronological age (-0.004 ± 0.071 and -0.009 ± 0.148 , respectively), while the other machine learning methods underestimate

it. All metrics show significantly poorer performance for the ADAB and BRR methods. A non-significant trend shows better performance for the STACK, RF, MLP, and SVM methods.

Figure 3 a shows Bland and Altman graphs for the two reference methods DEM and WILL. An overestimation of dental age over the whole U16 group is observed for DEM (mean error -0.70) and WILL (mean error -0.22). The plots show a tendency of both methods to underestimate dental age up to the age of 6 years and to overestimate dental age from the age of 12 years. The outliers are similar for both methods. The onset of dental maturation is earlier for girls compared to boys. A dense cloud of dots is observed for both methods until the age of 8 years, and then alignments of dots parallel to each other are progressively identified more markedly with age. Each of the upward sloping lines corresponds to a combination of stages for the 7 left mandibular teeth. The spread of minimum and maximum chronological age for a combination of maturation stages is therefore visually observed with the extremities of these

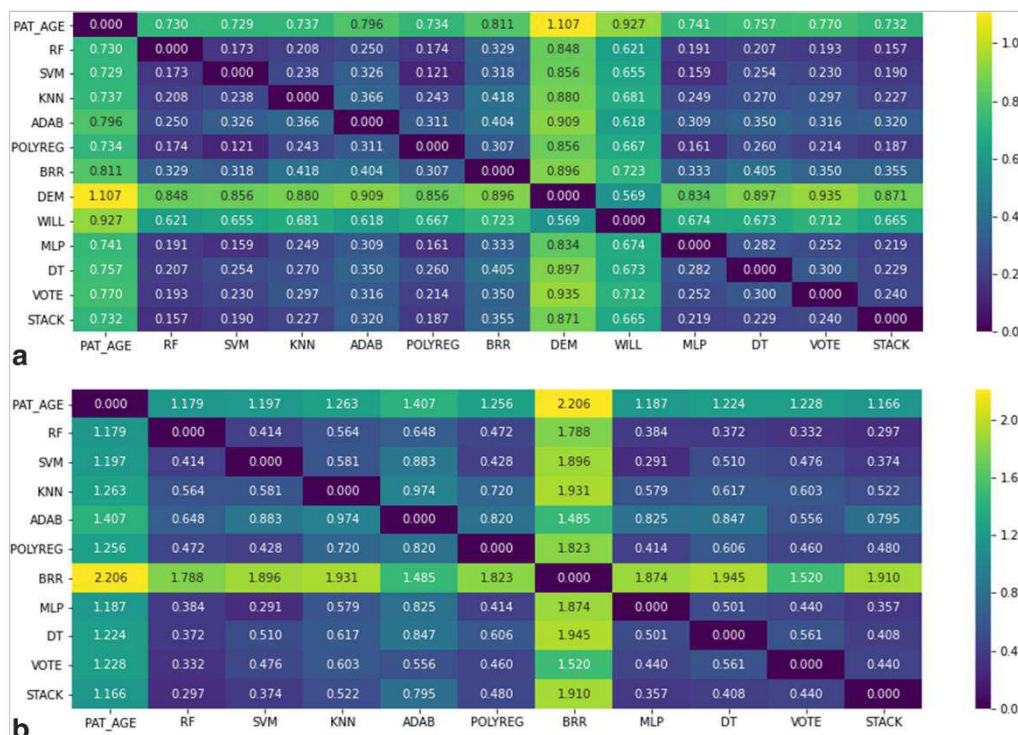


Fig. 2 Heat maps showing the mean of the mean absolute errors (MAE) calculated from the 20 replicates for each pair of dental age estimation methods for the 4–16 years (a) and 4–24 years group (b). Abbreviations: ADAB, AdaBoost; BRR, Bayesian ridge regression; DEM, Demirjian's reference method; DT, decision tree; KNN, k-nearest neighbor; MLP,

multi-layer perceptron; PAT_AGE, chronological age; POLYREG, polynomial regression; RF, random forest; STACK, stacking regressor; SVM, support vector machine; VOTE, voting regressor; WILL, Willems' reference method. Standard Deviation (SD) of each method

alignments. Both reference methods overestimate dental age for the more advanced combinations of tooth maturation and are more centered for the earlier stages.

Figure 3 b shows the results of the stacking learning machine method in the form of two Bland and Altman plots for group U16 and group U24 without and with the third molars, respectively. The dispersion of the measurements is greater for the U24 group from the age of 16 years, with maximal errors from 17 to 20 years, while decreasing after 20 years. Bland and Altman plots are available in supplementary data for all methods.

Discussion

Age estimation using the Demirjian and Willems reference methods

Compared to the world population

The first step in this work was to test our sample with two age estimation methods commonly referred in the literature.

Among the reference methods, Demirjian et al. developed their method as a universal tool to assess dental maturity and predict dental age in children [11]. As expected from the existing literature, the results obtained by DEM method in this study overestimated the age of females by 0.18 years and males by 1.2 years. These results are consistent with the results of the meta-analysis of Jayaraman et al. which shows that the method overestimates the age on average of females by 0.65 years (−0.10 to 2.82 years) and males by 0.60 years (−0.23 to 3.04 years) [33]. Several authors have tried to overcome this problem, such as Willems et al. by distinct maturity score tables that give the age directly in years [19]. The results obtained by WILL method in this study underestimated the age of females by 0.36 years and overestimated the age of males by 0.77 years. These results are consistent with the results of the meta-analysis of Sehrawat et al. which shows that the method overestimates the age on average of females by 0.07 years (−0.53 to 1.01 years) and males by 0.16 years (−1.1 to 0.7 years) [34]. Although the meta-analysis of the results of these reference methods is consistent with our results, we can also compare them with a population that is geographically quite similar.

Table 2 Mean error (ME), mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and R^2 values for the U24 group (4–24 years old), comparing machine learning regression methods for chronological age estimation

	ME \pm SD	MAE \pm SD	MSE \pm SD	RMSE \pm SD	$R^2 \pm$ SD
BRR	0.017 \pm 0.117 <i>a</i> (–0.217–0.228)	2.206 \pm 0.064 <i>b</i> (2.094–2.358)	7.822 \pm 0.529 <i>b</i> (6.932–9.202)	2.795 \pm 0.094 <i>b</i> (2.633–3.034)	0.759 \pm 0.017 <i>b</i> (0.731–0.789)
SVM	–0.004 \pm 0.071 <i>a</i> (–0.135–0.152)	1.197 \pm 0.037 <i>d, e</i> (1.129–1.258)	2.758 \pm 0.172 <i>d, e</i> (2.512–3.042)	1.660 \pm 0.052 <i>c, f</i> (1.585–1.744)	0.915 \pm 0.005 <i>c, d</i> (0.906–0.928)
DT	0.011 \pm 0.071 <i>a</i> (–0.156–0.142)	1.224 \pm 0.042 <i>c, d</i> (1.137–1.308)	2.832 \pm 0.216 <i>c, d, e</i> (2.432–3.308)	1.682 \pm 0.064 <i>c, d</i> (1.560–1.819)	0.913 \pm 0.006 <i>c, d</i> (0.898–0.925)
RF	0.012 \pm 0.065 <i>a</i> (–0.115–0.144)	1.179 \pm 0.038 <i>e</i> (1.113–1.258)	2.595 \pm 0.165 <i>d</i> (2.329–3.024)	1.610 \pm 0.051 <i>f</i> (1.526–1.728)	0.920 \pm 0.005 <i>c</i> (0.906–0.928)
KNN	0.044 \pm 0.081 <i>a</i> (0.087–0.224)	1.263 \pm 0.044 <i>c</i> (1.201–1.352)	3.087 \pm 0.231 <i>c</i> (1.667–1.898)	1.756 \pm 0.065 <i>e</i> (1.667–1.898)	0.905 \pm 0.007 <i>e</i> (0.886–0.913)
MLP	–0.009 \pm 0.148 <i>a</i> (–0.279–0.234)	1.187 \pm 0.043 <i>d, e</i> (1.127–1.275)	2.701 \pm 0.200 <i>d</i> (1.568–1.778)	1.643 \pm 0.060 <i>c, f</i> (1.568–1.778)	0.917 \pm 0.006 <i>c</i> (0.902–0.926)
POLYREG	0.003 \pm 0.069 <i>a</i> (–0.135–0.152)	1.256 \pm 0.039 <i>c</i> (1.176–1.309)	2.995 \pm 0.228 <i>c, e</i> (2.586–3.437)	1.729 \pm 0.066 <i>d, e</i> (1.608–1.854)	0.908 \pm 0.007 <i>d, e</i> (0.894–0.920)
ADAB	0.031 \pm 0.082 <i>a</i> (–0.090–0.207)	1.407 \pm 0.049 <i>a</i> (1.338–1.504)	3.455 \pm 0.212 <i>a</i> (3.152–3.939)	1.858 \pm 0.057 <i>a</i> (1.775–1.985)	0.893 \pm 0.006 <i>a</i> (0.875–0.903)
STACK	0.005 \pm 0.064 <i>a</i> (–0.119–0.130)	1.166 \pm 0.037 <i>e</i> (1.095–1.232)	2.594 \pm 0.175 <i>d</i> (2.300–2.985)	1.610 \pm 0.054 <i>f</i> (1.516–1.728)	0.920 \pm 0.005 <i>c</i> (0.908–0.929)
VOTE	0.014 \pm 0.068 <i>a</i> (–0.093–0.140)	1.228 \pm 0.041 <i>c, d</i> (1.167–1.296)	2.770 \pm 0.189 <i>d, e</i> (2.538–max.233)	1.663 \pm 0.056 <i>c, f</i> (1.593–1.798)	0.915 \pm 0.006 <i>c, d</i> (0.900–0.923)

For all metrics, the means \pm standard deviations with (maximum–minimum) of the 20 experiments were given. Tukey’s honest significance test for multiple comparison of means was performed at the .05 level. Within the same column, methods sharing the same letter are not significantly different from each other

Abbreviations: ADAB AdaBoost, BRR Bayesian ridge regression, DT decision tree, KNN k-nearest neighbor, MLP multi-layer perceptron, POLYREG polynomial regression, RF random forest, STACK stacking regressor, SVM support vector machine, VOTE voting regressor. Standard deviation (SD) of each method

Compared to the population of Southwest France

The results obtained on a sample of population in the South of France with the two reference methods (i.e., Demirjian and Willems) are also in agreement with a previous study by Urzel and Bruznek, who used on a similar population, the same WILL age estimation method, and a variant of the DEM method revised by Demirjian and Goldstein [17, 24]; both methods overestimate dental age, and WILL method is more accurate than DEM method. For the WILL method, the results found by Urzel and Bruznek are more accurate, with a slight underestimation by 0.09 years for females and an overestimation by 0.14 years for males. These results can be explained by an age group distribution in this study with more individuals in the age groups below 7 years and above 13 years, namely, the age categories with the higher age differences in our study.

Age estimation using the machine learning methods

The second step of present work was to develop new standards based on machine learning strategies that can improve on the accuracy of actual table standards. Instead of developing a fixed standards table, machine learning may perceive

complex data schemas. These strategies consider the workflow of ML methods (several algorithms may be chained) as well as the best hyperparameters determined a priori during training of the model.

In the literature, three main studies have used machine learning to estimate dental age and are discussed in this section: (a) the study by Tao et al. [38] with a multi-layer perceptron, using the same annotation system for the stages of tooth maturation; (b) the study by Štepanovský et al. [39] with several machine learning methods (multiple linear regression, support vector machine, multi-layer perceptron, radial basis function neural network, k-nearest neighbors, and k-star); and c) the study by Villa-Blanco et al. [40] with a convolutional neural network.

Comparison with reference methods

The first part of the evaluation of the ML methods was to compare them with the reference methods DEM and WILL under the same conditions of use of the latter, i.e., for the U16 population without using the third molars. All ML methods for estimating dental age outperformed the DEM and WILL reference methods. Among all machine learning methods

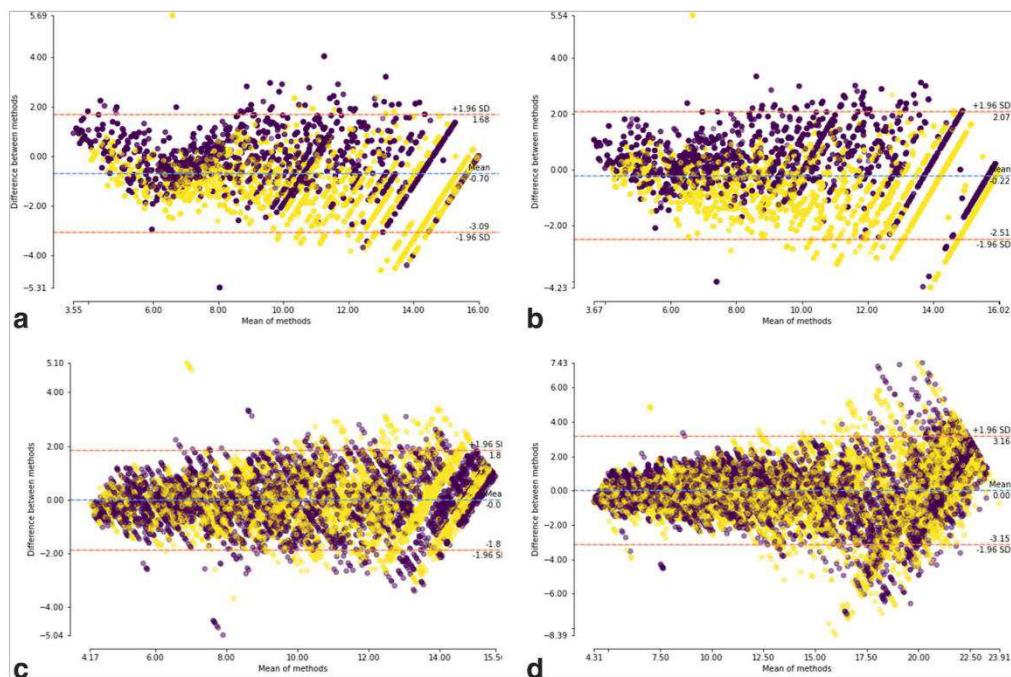


Fig. 3 Bland and Altman plot of difference in years between chronological age and dental age estimated **a** Demirjian's method (for the 4–16 years group), **b** Willems' method (for the 4–16 years group), **c** stacking regressor machine learning method (for the 4–16 years group), and **d** stacking regressor machine learning method (for the 4–24 years

group). Negative values indicate that dental age is higher than chronological age. Purple points represent male individuals and yellow points represent female individuals. These graphs represent the 20 independent experiments performed

tested (4–16 years), the maximum mean differences obtained were an overestimation of 15 days for the multi-layer perceptron method and an underestimation of 25 days for the support vector machine, whereas DEM and WILL overestimated the age by a mean of 257 days and 80 days, respectively. These results are in agreement with the results of other studies on dental age estimation [38, 40].

The study by Metsäniitty et al. [18] assessed dental age in a U24 population using the 7 permanent mandibular teeth and the 4 third molars. The development of these teeth was staged respectively according to DEM method and Köhler et al. [48] The reported RMSE was 1.175 years for females and 1.095 years for males. These results are more accurate than those obtained for U24. This may be explained on the one hand by the third molar staging system using 10 stages of mineralization and on the other hand by the under-representation of individuals older than 17 years old in the sample.

Comparison with a study using a multi-layer perceptron

The study by Tao et al. [38] assessed one particular ML method that is a multi-layer perceptron, using the same annotation system for the stages of tooth maturation, and 7 permanent teeth. The MAE, MSE, and RMSE results of Tao et al. for MLP are 0.990,

1.775, and 1.332 years for males and 1.261, 2.616, and 1.617 years for females, compared to 0.742, 0.907, and 0.952 years for male/female in the present study in U16. These differences can be explained by the broader age boundaries used by Tao et al. For U24, using 7 permanent teeth + 4 third molars, the MLP MAE, MSE, and RMSE results are 1.187, 2.701, and 1.643 years.

Comparison with a study using a different annotation system

The study by Štepanovský et al. used several machine learning methods (multiple linear regression, support vector machine, multi-layer perceptron, radial basis function neural network, k-nearest neighbors, and k-star) [39]. However, Štepanovský used an annotation system described by Moorrees et al. employing more stages of mineralization of the dental organ [1]. The population sample of Štepanovský et al. [39] was 662 males and 314 females aged 2.7 to 20.5 years, with 16 teeth annotated on the left side of the maxilla and mandible on panoramic radiographs (14 stages of maturation). The MAE and RMSE results are under 0.94 and 1.27 years, respectively. The first reason for this better accuracy may be the frequency distribution of the population: 10% of the population was aged over 15 years compared to the 35% of the present study; the 15–20 age group exhibits the largest variations. The second

possible reason is the use of more dental maturation stages and more teeth annotated than in our study: 14 versus 8 stages and 16 versus 11, respectively. Thus, the population of Štepanovský et al. may be compared to the U16 group of this study where the MAE and RMSE of all methods tested are under 0.811 and 1.014 years. Indeed, using a simpler annotation standard does not result in a significant performance loss compared to a higher number of teeth annotations.

Comparison with a study using a deep learning approach

The study by Villa-Blanco et al. used a convolutional neural network (CNN, a deep learning strategy) [40]. The authors succeeded in developing a fully automatic method of image analysis. Without using any annotation system, panoramic radiography is used straightforwardly. For the U16 group, the CNN obtained a ME and a MAE of -0.020 ± 0.97 years and 0.75 years, respectively. For group U25, the CNN achieved a MAE of 1.21 ± 1.17 years. These results are similar with those obtained for U16 and U24, respectively.

The use of a CNN proposed by Vila-Blanco et al. is particularly attractive because it is fast, easy to use, and less subjected to human interpretation. It presents a disadvantage according to the explainability of the results. Only areas of interest on the image are identified using a Grad-CAM methodology [49], whose heat map is mainly centered on the teeth. Nevertheless, the authors underline that Grad-CAM heat map remains “significantly variable according to the images.”

Predictors used in ML methods

A strong advantage in forensic medicine relies on the combination of a proven annotation system (i.e., the maturation stages described by Demirjian) to create a statistical model adapted to the population to study. It was hypothesized that supervised machine learning is capable of interpreting data whose meaning is not accessible to us [35] and that the amount of data matters more than algorithms [50]. In the present study, sexual dimorphism (gender) has also been used as a predictor for ML, to avoid losing power by developing separate models for boys and girls.

Bland and Altman graphs to interpret the results of dental age estimation

The Bland and Altman graphs are informative at several levels. It can be visually observed that the reference methods tend to overestimate dental age for ages above 14 years. The ML methods show a more homogeneous distribution of points around the x -axis. We also observe for the same chronological age, the dispersion of the dental ages, which reflects both the individual variability of the maturation of the teeth, the imprecision of the machine learning methods, and the number of maturation stages.

Thus, it is observed that even using third molars, the variability increases between 16 and 21 years of age (consequently the RMSE and MSE increase due to the square penalty). The Bland and Altman graph also allows us to observe the individual variability of a combination of maturation stages that corresponds to the alignment of points whose extremities correspond to individuals in the sample who are either early or late in development relative to the population sample. Finally, such graphs have an interest in the explicability of the ML methods used. Explicability of results is essential in forensic medicine. To be used in legal proceedings, it is necessary to provide elements of explanation. Some methods behave like black boxes (MLP, RF), while others are perfectly explainable (polynomial regression, decision tree). However, the methods the least easy to interpret give the best results on all the metrics (SVM, RF, MLP). New tools are being developed to break the black box and bring tangible elements of interpretation.

Drawbacks of machine learning methods tested

There are two approaches in forensic for the dental age. The first one estimates dental age and the second one aims at classifying individuals with respect to a key age [6]. A limitation of our study was to compare our results with those of the literature using reference methods to estimate dental age for the 16–24 age groups which is rather studied using a classification method. Prieto et al. [51] provided results comparable to ours on a Spanish population aged 14–21 years. They observe a mean error of -0.1 ± 1.3 and -0.07 ± 1.22 years respectively for the left and right third molar. Our results extracted from U24 for the 16–24-year-old subgroup give higher mean errors (Supplementary Table 2). Estimating the dental age for individuals aged from 16 to 24 years from the regression equation proposed by Hofmann et al. [45] gives results comparable (RMSE = 2.061 years) with those obtained with ML methods (Supplementary Table 2).

The use of dental age estimation methods, including machine learning methods in a forensic context, must consider individual variability. This is particularly important since the stages of dental maturation were developed in 1973 on a Canadian population [11]. Whatever the statistical approach, the variability of the prediction of dental age represents individual variability. Adding geographic origin, socio-nutritional conditions and endocrine factors using ML algorithm may help to reduce such individual variability [3] and improve dental age estimation. There is a trend towards maximum prediction accuracy for ML algorithms. This limit in accuracy may be due to the limitations of the annotation system and/or to inter-individual variability. A solution would be to go beyond the stages or to use additional predictors. Deep learning methods could allow a more detailed analysis of the maturation stages by breaking down the inter- or intra-examiner reproducibility. These methods are already used with CNN to analyze an entire panoramic radiograph to determine dental/bone age [40].

Conclusion

Machine learning methods were used in this study to assess the predictability of a dental maturation staging system based on dental panoramics. On the population studied in U16, as a reference, WILL method was significantly more accurate than DEM, and all ML methods were more accurate than the best reference method. On the population studied in U24, all ML methods rendered similar results, except for one (BRR). This study confirms that ML methods outperform the classic approaches for age estimation based on radiographic dental staging, from childhood up to early adulthood.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00414-020-02489-5>.

Code availability The code generated during and/or analyzed during the current study is available from the corresponding author on reasonable request.

Authors' contributions Antoine Galibourg: Conceptualization, software, investigation, and writing—original draft

Sylvain Cussat-Blanc: Methodology and software

Jean Dumoncel: Writing—review and editing, software

Norbert Telmon: Validation and supervision

Paul Monsarrat: Formal analysis, methodology, software, writing—review and editing

Delphine Maret: Writing—review and editing, methodology

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflicts of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

- Moorrees CFA, Fanning EA, Hunt EE (1963) Age variation of formation stages for ten permanent teeth. *J Dent Res* 42:1490–1502. <https://doi.org/10.1177/00220345630420062701>
- Jylhävä J, Pedersen NL, Hägg S (2017) Biological age predictors. *EBioMedicine* 21:29–36. <https://doi.org/10.1016/j.ebiom.2017.03.046>
- Braga J, Heuze Y, Chabadel O, Sonan NK, Gueramy A (2005) Non-adult dental age assessment: correspondence analysis and linear regression versus Bayesian predictions. *Int J Legal Med* 119:260–274. <https://doi.org/10.1007/s00414-004-0494-8>
- Cunha E, Baccino E, Martrille L, Ramsthaler F, Prieto J, Schuliar Y, Lynnerup N, Cattaneo C (2009) The problem of aging human remains and living individuals: a review. *Forensic Sci Int* 193:1–13. <https://doi.org/10.1016/j.forsciint.2009.09.008>
- Willems G (2001) A review of the most commonly used dental age estimation techniques. *Journal of Forensic Odonto-Stomatology* 19:9–17
- Sironi E, Gallidabino M, Weyermann C et al (2018) Bayesian networks of age estimation and classification based on dental evidence: a study on the third molar mineralization. *J Forensic Legal Med* 55:23–32. <https://doi.org/10.1016/j.jflm.2018.02.005>
- Timme M, Steinacker JM, Schmeling A (2017) Age estimation in competitive sports. *Int J Legal Med* 131:225–233. <https://doi.org/10.1007/s00414-016-1456-7>
- Galić I, Vodanović M, Cameriere R, Nakaš E, Galić E, Selimović E, Brkić H (2011) Accuracy of Cameriere, Haavikko, and Willems radiographic methods on age estimation on Bosnian-Herzegovian children age groups 6–13. *Int J Legal Med* 125:315–321. <https://doi.org/10.1007/s00414-010-0515-8>
- Schmeling A, Geserick G, Reisinger W, Olze A (2007) Age estimation. *Forensic Sci Int* 65:178–181. <https://doi.org/10.1016/j.forsciint.2006.05.016>
- Esan TA, Yengopal V, Schepartz LA (2017) The Demirjian versus the Willems method for dental age estimation in different populations: a meta-analysis of published studies. *PLoS One* 12:1–23. <https://doi.org/10.1371/journal.pone.0186682>
- Demirjian A, Goldstein H, Tanner JM (1973) A new system of dental age assessment. *Hum Biol* 45:211–227
- Thevissen PW, Fieuws S, Willems G (2011) Third molar development: measurements versus scores as age predictor. *Arch Oral Biol* 56:1035–1040. <https://doi.org/10.1016/j.archoralbio.2011.04.008>
- Liversidge HM (2012) The assessment and interpretation of Demirjian, Goldstein and Tanner's dental maturity. *Ann Hum Biol* 39:412–431
- Lucas VS, Andiappan M, McDonald F, Roberts G (2016) Dental age estimation: a test of the reliability of correctly identifying a subject over 18 years of age using the gold standard of chronological age as the comparator. *J Forensic Sci* 61(5):1238–1243. <https://doi.org/10.1111/1556-4029.13132>
- Jayaraman J, Roberts GJ, Wong HM, King NM (2018) Dental age estimation in southern Chinese population using panoramic radiographs: validation of three population specific reference datasets. *BMC Med Imaging* 18:4–11. <https://doi.org/10.1186/s12880-018-0250-z>
- Chaillet N, Nyström M, Demirjian A (2005) Comparison of dental maturity in children of different ethnic origins: international maturity curves for clinicians. *J Forensic Sci* 50:1–11. <https://doi.org/10.1520/jfs2005020>
- Demirjian A, Goldstein H (1976) New systems for dental maturity based on seven and four teeth. *Ann Hum Biol* 3:411–421. <https://doi.org/10.1080/03014467600001671>
- Metsäniitty M, Waltimo-Sirén J, Ranta H, Fieuws S, Thevissen P (2019) Dental age estimation in Somali children and sub-adults combining permanent teeth and third molar development. *Int J Legal Med* 133:1207–1215. <https://doi.org/10.1007/s00414-019-02053-w>
- Willems G, Van Olmen A, Spiessens B, Carels C (2001) Dental age estimation in Belgian children: Demirjian's technique revisited. *J Forensic Sci* 46:893–895
- Alassiry A, Alshomrani K, Al Hasi S et al (2019) Dental age assessment of 3–15-year-old Saudi children and adolescents using Demirjian's method—a radiographic study. *Clinical and Experimental Dental Research* 5:336–342. <https://doi.org/10.1002/cre2.186>
- Prasad H, Kala N (2019) Accuracy of two dental age estimation methods in the Indian population - a meta-analysis of published studies. *Journal of Forensic Odontostomatology* 3:2–11

22. Ranasinghe S, Perera J, Taylor JA, Tennakoon A, Pallewatte A, Jayasinghe R (2019) Dental age estimation using radiographs: towards the best method for Sri Lankan children. *Forensic Sci Int* 298:64–70. <https://doi.org/10.1016/j.forsciint.2019.02.053>
23. Yang Z, Geng K, Liu Y, Sun S, Wen D, Xiao J, Zheng Y, Cai J, Zha L, Liu Y (2019) Accuracy of the Demirjian and Willems methods of dental age estimation for children from central southern China. *Int J Legal Med* 133:593–601. <https://doi.org/10.1007/s00414-018-1924-3>
24. Urzel V, Bruzek J (2013) Dental age assessment in children: a comparison of four methods in a recent french population. *Journal of Forensic Science* 58:1341–1347. <https://doi.org/10.1111/1556-4029.12221>
25. Alqadi MA, Abuaffan AH (2019) Validity of the Demirjian and Fishman methods for predicting chronological age amongst Yemeni children. *Sultan Qaboos University Medical Journal* 19: 26–31. <https://doi.org/10.18295/squmj.2019.19.01.006>
26. Balgi P, Katge F, Singh S, Jayaraman J (2020) Comparative assessment to determine the accuracy of Demirjian and DARLInG methods of dental age estimation in western Indian children. *Legal Medicine (Tokyo)* 42:101629. <https://doi.org/10.1016/j.legalmed.2019.101629>
27. Hegde S, Patodia A, Shah K, Dixit U (2019) The applicability of the Demirjian, Willems and Chaillet standards to age estimation of 5-15 year old Indian children. *The Journal of Forensic Odontostomatology* 37:40–50
28. Kar May L, Mei Shian AY, Durward C, Jayaraman J (2020) A method of estimating age of undocumented children and young adults of different socioeconomic status in Cambodia. *Heliyon* 6: e03476. <https://doi.org/10.1016/j.heliyon.2020.e03476>
29. Kermani M, Tabatabaei Yazdi F, Abed Haghighi M (2019) Evaluation of the accuracy of Demirjian's method for estimating chronological age from dental age in Shiraz, Iran: using geometric morphometrics method. *Clinical and Experimental Dental Research* 5:191–198. <https://doi.org/10.1002/cre2.169>
30. Lan LM, Yang ZD, Sun SL et al (2019) Application of Demirjian's and Cameriere's method in dental age estimation of 8-16 year old adolescents from Hunan Han nationality. *Fa Yi Xue Za Zhi* 35: 406–410. <https://doi.org/10.12116/j.issn.1004-5619.2019.04.005>
31. Mohanty I, Panda S, Dalai RP, Mohanty N. Predictive accuracy of Demirjian's, modified Demirjian's and India specific dental age estimation methods in Odisha (Eastern Indian) population. *The Journal of Forensic Odontostomatology* 37:32–39
32. Moness Ali AM, Ahmed WH, Khattab NM (2019) Applicability of Demirjian's method for dental age estimation in a group of Egyptian children. *BDJ Open* 5:2. <https://doi.org/10.1038/s41405-019-0015-y>
33. Jayaraman J, Wong HM, King NM, Roberts GJ (2013) The French–Canadian data set of Demirjian for dental age estimation: a systematic review and meta-analysis. *J Forensic Legal Med* 20: 373–381. <https://doi.org/10.1016/j.jflm.2013.03.015>
34. Schrawat JS, Singh M (2017) Willems method of dental age estimation in children: a systematic review and meta-analysis. *J Forensic Legal Med* 52:122–129. <https://doi.org/10.1016/j.jflm.2017.08.017>
35. Fabris F, de Magalhães JP, Freitas AA (2017) A review of supervised machine learning applied to ageing research. *Biogerontology* 18:171–188. <https://doi.org/10.1007/s10522-017-9683-y>
36. Ritz-Timme S, Cattaneo C, Collins MJ, Waite ER, Schütz HW, Kaatsch HJ, Bormann HIM (2000) Age estimation: the state of the art in relation to the specific demands of forensic practise. *Int J Legal Med* 113:129–136. <https://doi.org/10.1007/s004140050283>
37. Dallora AL, Anderberg P, Kvist O, Mendes E, Diaz Ruiz S, Sanmartin Berglund J (2019) Bone age assessment with various machine learning techniques: a systematic literature review and meta-analysis. *PLoS One* 14:1–22. <https://doi.org/10.1371/journal.pone.0220242>
38. Tao J, Wang J, Wang A et al (2020) Dental age estimation: a machine learning perspective. In: Hassanien a., Azar a., Gaber T., Bhatnagar R., F. Tolba M. (eds) *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019)*. AMLTA 2019. *Advances in Intelligent Systems and Computing*, vol 921. Springer, Cham. pp. 722–733. https://doi.org/10.1007/978-3-030-14118-9_71
39. Štepanovský M, Ibrová A, Buk Z, Velemínská J (2017) Novel age estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods. *Forensic Sci Int* 279:72–82. <https://doi.org/10.1016/j.forsciint.2017.08.005>
40. Vila-Blanco N, Carreira MJ, Varas-Quintana P, Balsa-Castro C, Tomas I (2020) Deep neural networks for chronological age estimation from OPG images. *IEEE Trans Med Imaging* 39:2374–2384. <https://doi.org/10.1109/TMI.2020.2968765>
41. Duangto P, Iamaroon A, Prasitvattanaseree S, Mahakkanukrauh P, Janhom A (2017) New models for age estimation and assessment of their accuracy using developing mandibular third molar teeth in a Thai population. *Int J Legal Med* 131:559–568. <https://doi.org/10.1007/s00414-016-1467-4>
42. Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830 <https://www.jmlr.org/papers/v12/> (accessed 1 December 2020)
43. Wolpert DH (1992) Stacked generalization. *Neural Netw* 5(2):241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
44. Mu X, Lu J, Watta P, Hassoun MH (2009) Weighted voting-based ensemble classifiers with application to human face recognition and voice recognition. In: 2009 International Joint Conference on Neural Networks [Internet]. IEEE 2168–2171. <https://doi.org/10.1109/IJCNN.2009.5178708>
45. Hofmann E, Robold M, Proff P, Kirschnick C (2017) Age assessment based on third molar mineralisation: an epidemiological-radiological study on a central-European population. *J Orofac Orthop* 78:97–111. <https://doi.org/10.1007/s00056-016-0063-z>
46. Tukey JW (1949) Comparing individual means in the analysis of variance. *Biometrics* 5:99–114
47. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
48. Köhler S, Schmelzke R, Loitz C, Püschel K (1994) Die entwicklung des weisheitszahnes als kriterium der lebensaltersbestimmung [development of wisdom teeth as a criterion of age determination]. *Ann Anat* 176:339–345
49. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2020) Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis* 128:336–359. <https://doi.org/10.1007/s11263-019-01228-7>
50. Halevy A, Norvig P, Pereira F (2009) The unreasonable effectiveness of data. *IEEE Intell Syst* 24:8–12. <https://doi.org/10.1109/MIS.2009.36>
51. Prieto JL, Barbería E, Ortega R, Magaña C (2005) Evaluation of chronological age based on third molar development in the Spanish population. *Int J Legal Med* 119:349–354. <https://doi.org/10.1007/s00414-005-0530-3>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Data

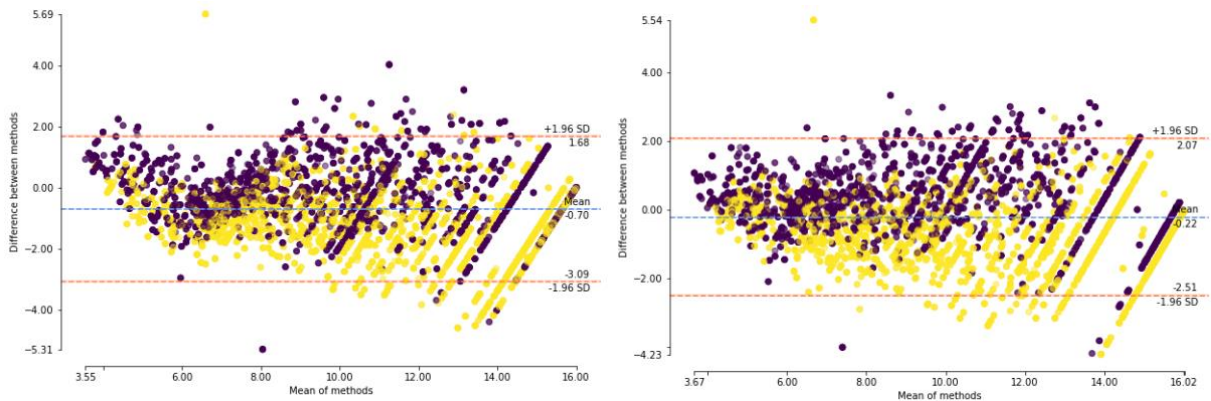


Figure S1: Bland and Altman plot of the difference in years between chronological age and dental age estimated for the 4-16 years age group (group U16) by Demirjian's method (DEM) on the left, and Willem's method (WILL) on the right. Negative values indicate lower dental age than chronological age. Purple points represent male individuals and yellow points represent female individuals.

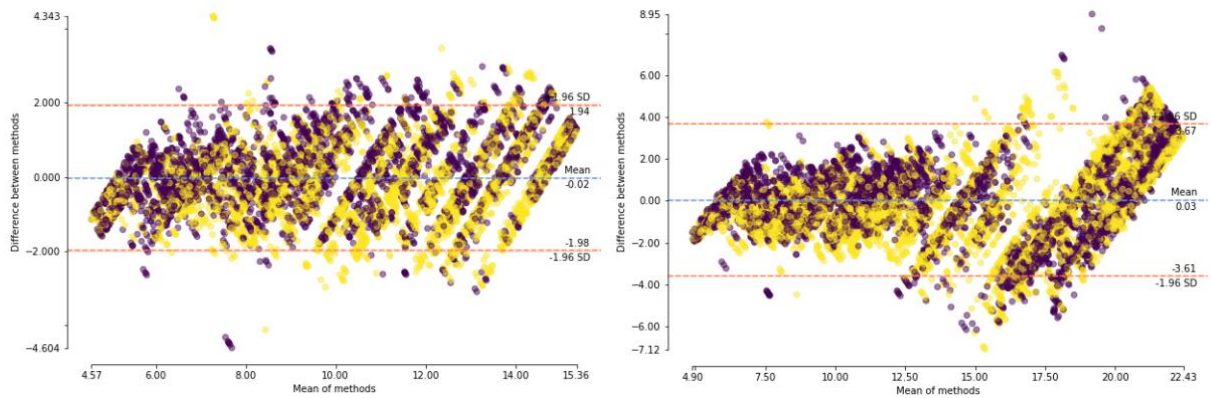


Figure S2: Bland and Altman plot difference in years between chronological age and dental age estimated by Adaboost Regressor machine learning method (ADAB) for the 4-16 years age group (group U16) on the left, and the 4-24 years age group (group U24) on the right. Negative values indicate lower dental age than chronological age. Purple points represent male individuals and yellow points represent female individuals.

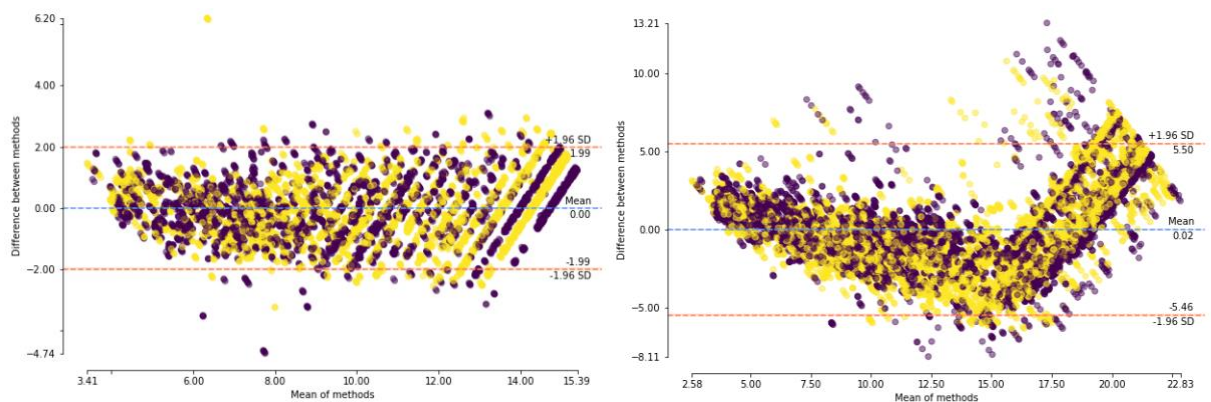


Figure S3: Bland and Altman plot difference in years between chronological age and dental age estimated by Bayesian Ridge Regression machine learning method (BRR) for the 4-16 years age group (group U16) on the left, and the 4-24 years age group (group U24) on the right. Negative values indicate

lower dental age than chronological age. Purple points represent male individuals and yellow points represent female individuals.

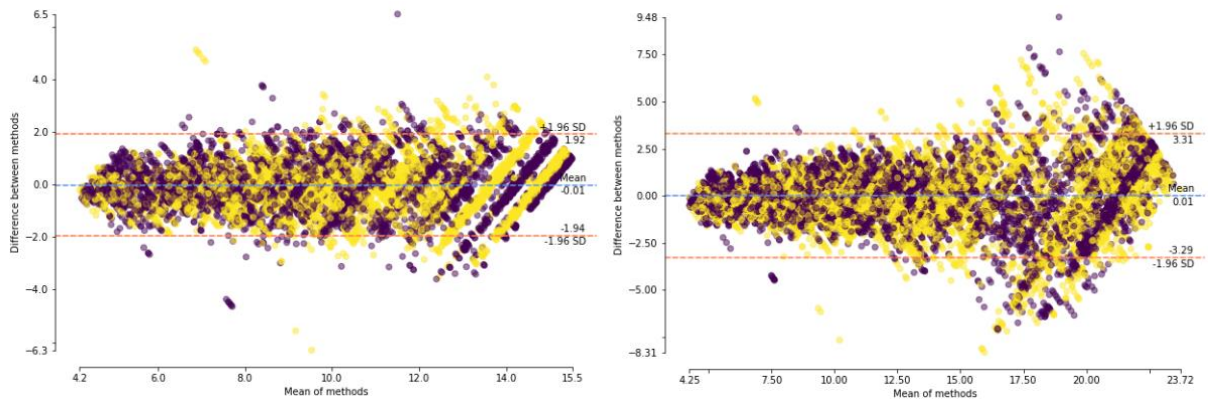


Figure S4: Bland and Altman plot difference in years between chronological age and dental age estimated by Decision Tree regressor machine learning method (DT) for the 4-16 years age group (group U16) on the left, and the 4-24 years age group (group U24) on the right. Negative values indicate lower dental age than chronological age. Purple points represent male individuals and yellow points represent female individuals.

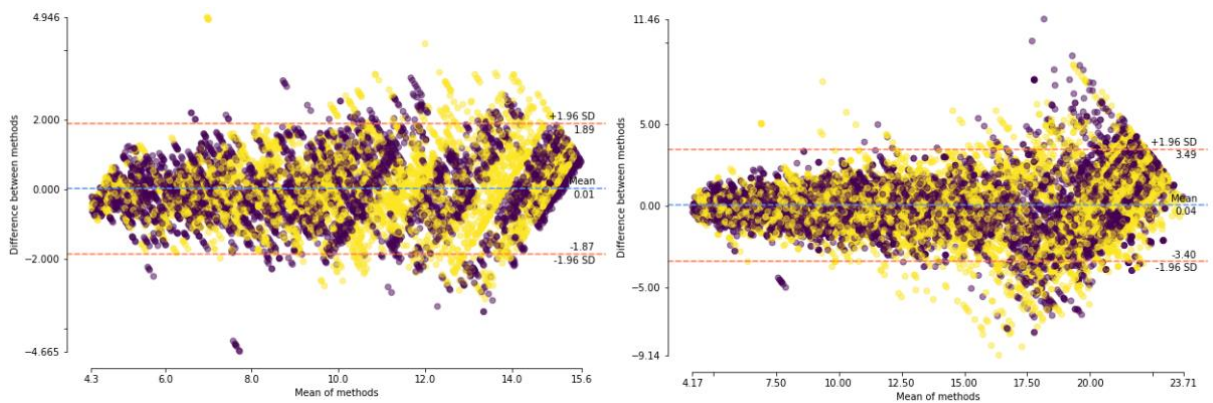


Figure S5: Bland and Altman plot difference in years between chronological age and dental age estimated by K-Nearest Neighbor regressor machine learning method (KNN) for the 4-16 years age group (group U16) on the left, and the 4-24 years age group (group U24) on the right. Negative values indicate lower dental age than chronological age. Purple points represent male individuals and yellow points represent female individuals.

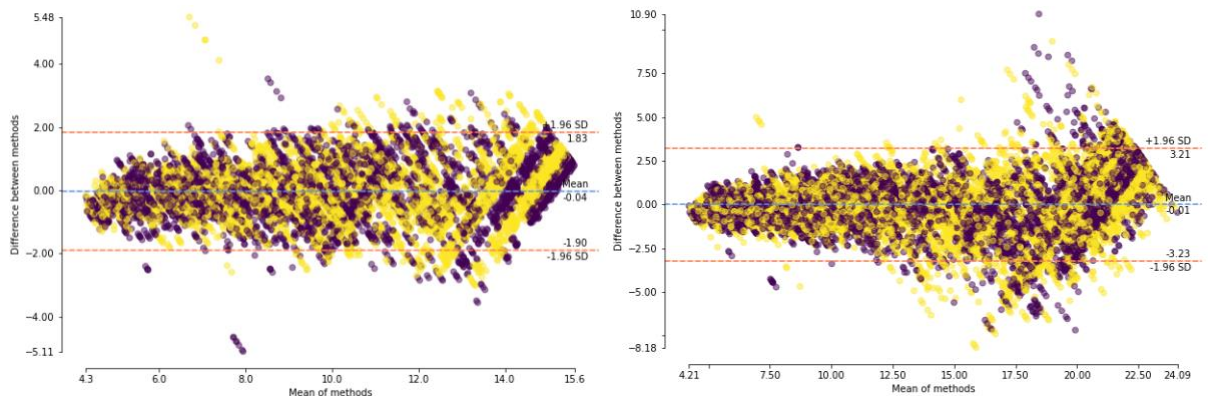


Figure S6: Bland and Altman plot difference in years between chronological age and dental age estimated by Multi-Layer Perceptron regressor machine learning method (MLP) for the 4-16 years age

group (group U16) on the left, and the 4-24 years age group (group U24) on the right. Negative values indicate lower dental age than chronological age. Purple points represent male individuals and yellow points represent female individuals.

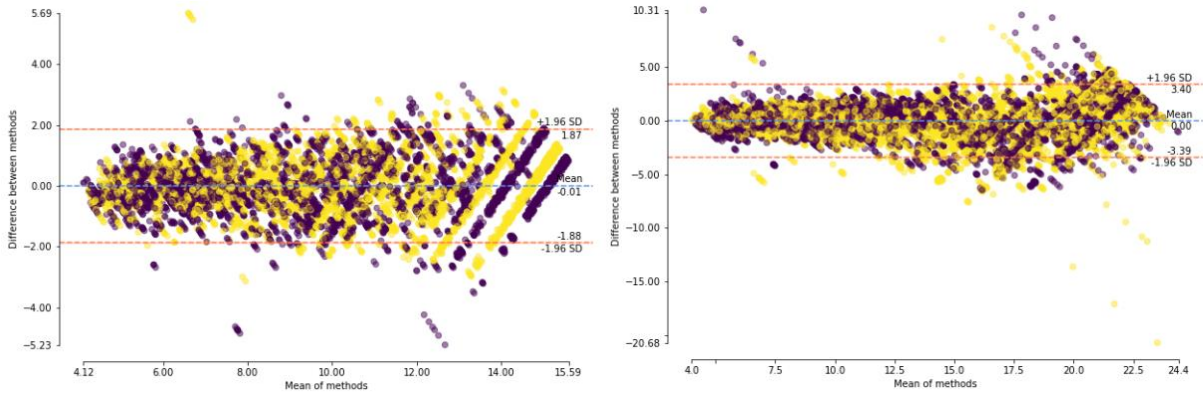


Figure S7: Bland and Altman plot difference in years between chronological age and dental age estimated by Polynomial Regressor machine learning method (POLYREG) for the 4-16 years age group (group U16) on the left, and the 4-24 years age group (group U24) on the right. Negative values indicate lower dental age than chronological age. Purple points represent male individuals and yellow points represent female individuals.

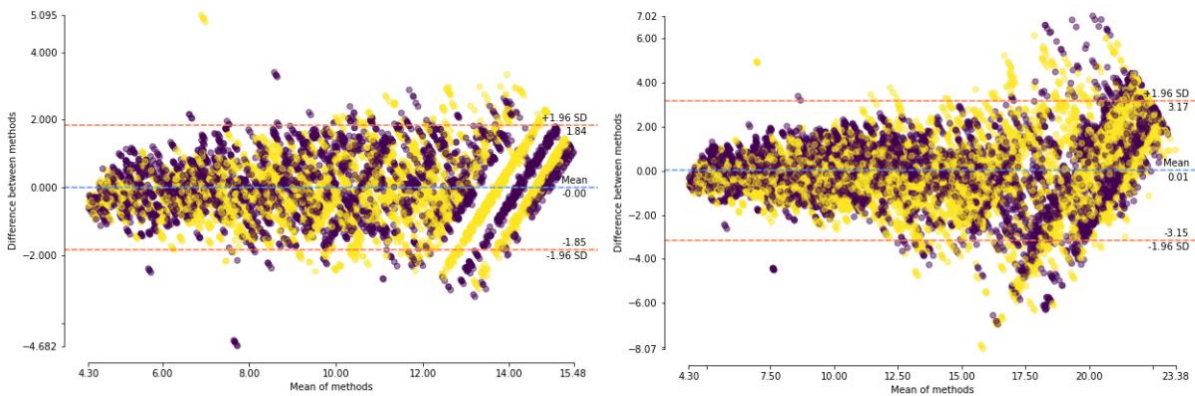


Figure S8: Bland and Altman plot difference in years between chronological age and dental age estimated by Random Forest regressor machine learning method (RF) for the 4-16 years age group (group U16) on the left, and the 4-24 years age group (group U24) on the right. Negative values indicate lower dental age than chronological age. Purple points represent male individuals and yellow points represent female individuals.

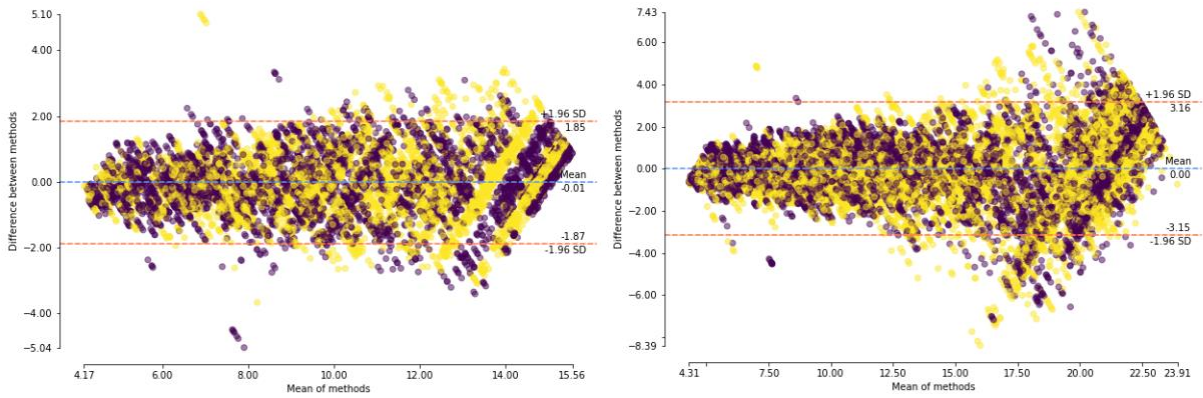


Figure S9: Bland and Altman plot difference in years between chronological age and dental age estimated by Stacking Regressor machine learning method (STACK) for the 4-16 years age group (group U16) on the left, and the 4-24 years age group (group U24) on the right. Negative values indicate lower dental age than chronological age. Purple points represent male individuals and yellow points represent female individuals.

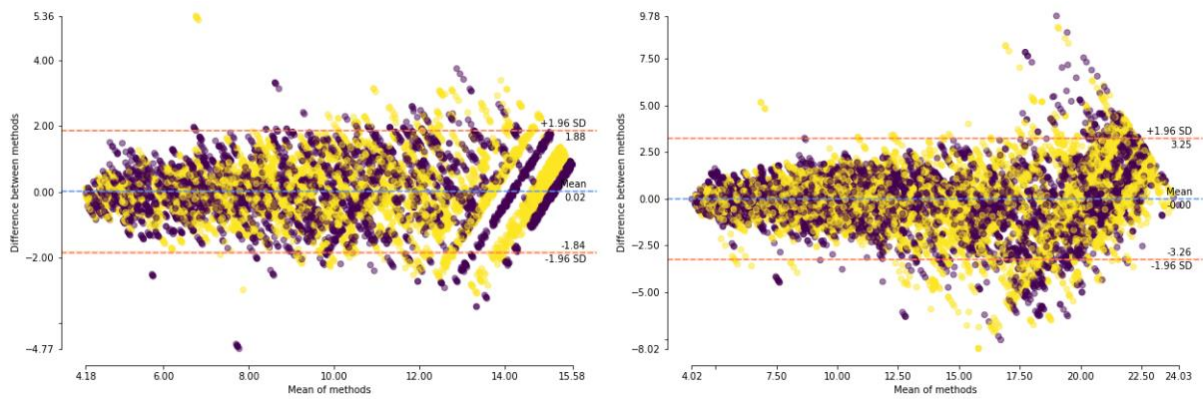


Figure S10: Bland and Altman plot difference in years between chronological age and dental age estimated by Support Vector Machine regressor machine learning method (SVM) for the 4-16 years age group (group U16) on the left, and the 4-24 years age group (group U24) on the right. Negative values indicate lower dental age than chronological age. Purple points represent male individuals and yellow points represent female individuals.

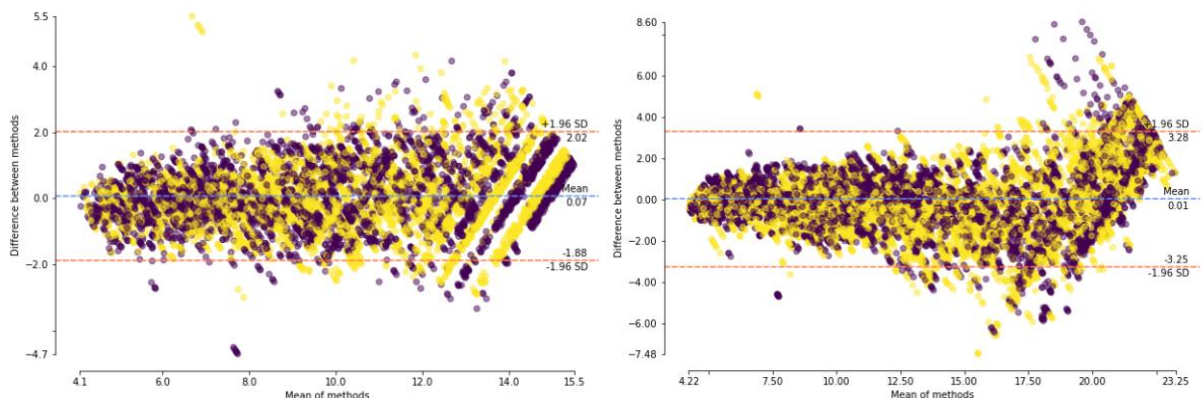


Figure S11: Bland and Altman plot difference in years between chronological age and dental age estimated by Voting Regressor machine learning method (VOTE) for the 4-16 years age group (group U16) on the left, and the 4-24 years age group (group U24) on the right. Negative values indicate lower dental age than chronological age. Purple points represent male individuals and yellow points represent female individuals.

Supplementary Table 1: List of the tuned hyperparameters for each Machine Learning algorithm. For each hyperparameter, the values inside square brackets were explored by Grid Search.

Random Forest (RF)	<p><i>max_depth</i>: the maximum depth of a tree [6 to 9] <i>max_features</i>: the number of features to consider when looking for the best split at a node [sqrt or auto] <i>min_samples_split</i>: the minimum number of samples required to split an internal node [5 to 25] <i>min_samples_leaf</i>: the minimum number of samples required to be at a leaf node [2 to 4] <i>bootstrap</i>: whether bootstrap samples are used when building trees [true, false]</p>
Support Vector Machine (SVM)	<p><i>kernel</i>: the kernel type used in the algorithm [polynomial, linear, rbf] <i>gamma</i>: kernel coefficient [0.0001, 0.0005, 0.001, 0.01, 0.05, 0.1] <i>C</i>: the regularization parameter. The higher the parameter the lower the regularization strength [1 to 10]</p>
K-Nearest Neighbors (KNN)	<p><i>n_neighbors</i>: the number of neighbors or K to use [3 to 20] <i>weights</i>: the weight function used in prediction [uniform or distance] <i>algorithm</i>: type of algorithm used to compute the nearest neighbors [ball tree, kd tree or brute]</p>
Adaboost (ADAB)	<p><i>learning_rate</i>: the learning rate shrinks the contribution of each regressor by its float number [0.01, 0.025, 0.05, 0.1, 0.3, 1] <i>loss</i>: the loss function to use when updating the weights after each boosting iteration [linear, square or exponential] <i>n_estimators</i>: the maximum number of estimators at which boosting is terminated [50 to 200]</p>
Polynomial regression (POLYREG)	<p><i>poly_degree</i>: the degree of polynomial regression [1 to 4]</p>
Bayesian Ridge Regression (BRR)	<p><i>alpha_1</i>: shape parameter for the Gamma distribution prior over the alpha parameter [$1e^{-15}$, $1e^{-10}$, $1e^{-8}$, $1e^{-4}$, $1e^{-3}$, $1e^{-2}$, 1, 5, 10, 20] <i>alpha_2</i>: inverse scale parameter (rate parameter) for the Gamma distribution prior over the alpha parameter [$1e^{-15}$, $1e^{-10}$, $1e^{-8}$, $1e^{-4}$, $1e^{-3}$, $1e^{-2}$, 1, 5, 10, 20] <i>lambda_1</i>: shape parameter for the Gamma distribution prior over the lambda parameter [$1e^{-15}$, $1e^{-10}$, $1e^{-8}$, $1e^{-4}$, $1e^{-3}$, $1e^{-2}$, 1, 5, 10, 20] <i>lambda_2</i>: inverse scale parameter (rate parameter) for the Gamma distribution prior over the lambda parameter [$1e^{-15}$, $1e^{-10}$, $1e^{-8}$, $1e^{-4}$, $1e^{-3}$, $1e^{-2}$, 1, 5, 10, 20]</p>
Multi-Layer Perceptron (MLP)	<p><i>activation</i>: the activation layer for the hidden layers [relu] <i>alpha</i>: penalty regularization term parameter [0.005, 0.01]</p>

	<p><i>hidden_layer_sizes</i>: the i^{th} element represents the number of neurons in the i^{th} hidden layer ([0,2,4,8,16,32,64,128,256,512] for 4 to 6 hidden layers)</p> <p><i>solver</i>: the solver for weight optimization [sgd or adam]</p>
	<p><i>criterion</i>: the function to measure the quality of a split in the tree [mse, Friedman mse or mae]</p> <p><i>max_depth</i>: the maximum depth of the tree [6 to 9]</p> <p><i>max_features</i>: the number of features to consider when looking for the best split at a node [auto, sqrt, log2]</p>
Decision Tree (DT)	<p><i>min_samples_split</i>: the minimum number of samples required to split an internal node [5 to 25]</p> <p><i>min_samples_leaf</i>: the minimum number of samples required to be at a leaf node [2 to 4]</p> <p><i>splitter</i>: the strategy used to choose the split at each node [best or random]</p>
Voting (VOTE)	<p><i>estimators</i>: Invoking the fit method on the VotingRegressor will fit clones of those original estimators that will be stored in the class attribute [RF, SVM, KNN, ADAB, POLYREG, BRR, MLP]</p>
Stacking (STACK)	<p><i>estimators</i>: the list of base estimators that will be stacked together [RF, SVM, KNN, ADAB, POLYREG, BRR, MLP]</p> <p><i>final_estimator</i>: a regressor that will be used to combine the base estimators [GradientBoostingRegressor]</p>

Supplementary Table 2: Mean Error (ME), Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and R² values for the 16-24 years old group assessing performance of machine learning regression methods and reference method for chronological age estimation. For all metrics, the means of the 20 independent experiments were given. Tukey’s honest significance test for multiple comparison of means was performed at the .05 level. Within the same column, methods sharing the same letter are not significantly different from each other.

Abbreviations: ADAB: AdaBoost, BRR: Bayesian Ridge Regression, DEM-Hofmann: Demirjian staging and Hofmann age regression for third molars, DT: Decision Tree, KNN: K-Nearest Neighbor, MLP: Multi-Layer Perceptron, POLYREG: Polynomial Regression, RF: Random Forest, STACK: Stacking Regressor, SVM: Support Vector Machine, VOTE: Voting Regressor..

<i>Method</i>	<i>ME</i>	<i>MAE</i>	<i>MSE</i>	<i>RMSE</i>	<i>R²</i>
<i>DEM-Hofmann</i>	-0.506	1.639	4.248	2.061	0.213
<i>BRR</i>	2.247	2.828	12.271	3.500	-1.216
<i>SVM</i>	0.527	1.689	4.381	2.092	0.209
<i>DT</i>	0.551	1.770	4.668	2.160	0.157
<i>RF</i>	0.657	1.707	4.295	2.072	0.224
<i>KNN</i>	0.606	1.834	5.216	2.280	0.059
<i>MLP</i>	0.557	1.695	4.416	2.099	0.203
<i>POLYREG</i>	0.603	1.720	4.586	2.139	0.171
<i>ADAB</i>	1.144	2.004	5.773	2.402	-0.041
<i>STACK</i>	0.515	1.715	4.354	2.086	0.214
<i>VOTE</i>	0.896	1.712	4.701	2.167	0.152

3. Discussion sur les méthodes d'intelligence artificielle utilisées dans l'estimation de l'âge dentaire

Cet article a mis en évidence la supériorité de toutes les méthodes de ML testées par rapport aux méthodes classiques de Demirjian et Willems dans l'estimation de l'âge dentaire. Il peut néanmoins sembler difficile de désigner la méthode de ML la plus précise. Elles sont présentées sous le terme générique de méthodes de machine learning ou bien encore comme des méthodes d'intelligence artificielle. Il existe un biais d'utilisation de ces termes qui englobent une grande variété de méthodes et de techniques qui n'ont pas les mêmes objectifs, ni les mêmes finalités.

L'expression d'intelligence artificielle est née lors de la conférence de Dartmouth en 1956.⁷⁷ Pendant 2 mois, des scientifiques comme Marvin Minsky et John McCarthy se sont réunis lors d'un camp d'été sur le campus de la faculté de Dartmouth. L'expression « Intelligence Artificielle » y a été inventé et les objectifs de l'IA y ont été définis. Cette conférence est l'initiation de la première révolution de l'IA avec un engouement pour le développement d'une approche basée sur la logique mathématique et l'exploration arborescente. Les participants de cette conférence sont identifiés comme les pères fondateurs de l'IA. Il n'en reste pas moins que la définition de l'IA reste difficile à établir puisqu'il existe de nombreuses manières de la définir. L'inventeur de ce terme, John McCarthy, disait lui-même : « Dès que ça marche, plus personne n'appelle ça de l'IA ». ⁷⁸ En effet, si le terme est relativement récent, l'IA était déjà utilisée depuis plusieurs centaines d'années. D'ailleurs, qui aujourd'hui s'extasie devant une calculatrice lors de la réalisation d'additions ou de soustractions, tandis qu'en 1642, la Pascaline, première calculatrice, créée par Blaise Pascal interpellait. Tout procédé algorithmique utilisé par une machine pourrait alors prétendre à la dénomination d'intelligence artificielle. Yann LeCun, qui a participé à la deuxième révolution de l'IA avec le développement des réseaux neuronaux profonds, y ajoute donc la notion d'apprentissage. Il définit alors l'IA comme une « machine qui a la capacité d'accomplir des tâches généralement assurées par les animaux ou les humains : percevoir, raisonner et agir. Elle [l'IA] est inséparable de la capacité à apprendre, telle qu'on l'observe chez les êtres vivants. Les systèmes d'IA ne sont que des circuits électroniques et des programmes informatiques très sophistiqués. » ⁷⁹

L'intelligence artificielle, l'apprentissage machine et l'apprentissage profond peuvent alors être replacés dans leur contexte. Ces trois disciplines appartiennent à l'informatique en tant qu'étude des processus algorithmiques et des machines de calcul. L'intelligence artificielle est développée sous deux grandes familles de programmation. La première, par l'approche symbolique, est basée sur la logique et correspond à la première révolution de l'intelligence artificielle. Les explorations arborescentes font

partie de cette approche. La deuxième famille de programmation correspond à l'approche numérique. C'est la deuxième révolution de l'intelligence artificielle, basée sur l'expérience passée, le connexionnisme et les réseaux de neurones artificiels. L'apprentissage machine est une sous-partie de l'intelligence artificielle. Cette discipline se concentre sur la façon dont les ordinateurs apprennent à partir de données. Enfin une sous-partie de l'apprentissage machine est constituée de l'apprentissage profond. Ces trois disciplines des sciences informatiques sont également au carrefour des sciences mathématiques et des sciences des données qui portent sur l'extraction des connaissances à partir de l'exploration d'un ensemble de données.⁸⁰ (Figure 5: Positionnement de l'intelligence artificielle dans les champs des sciences informatiques, des sciences des données et de mathématiques.)

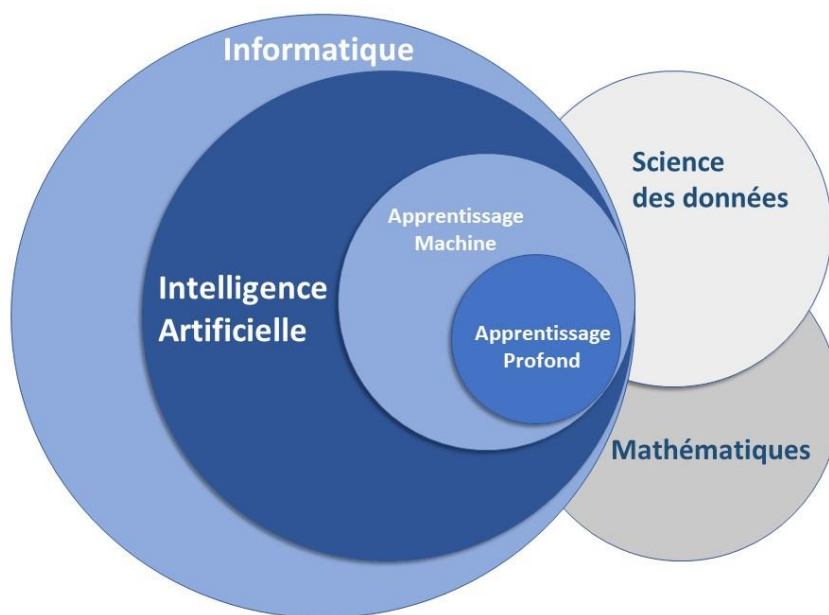


Figure 5: Positionnement de l'intelligence artificielle dans les champs des sciences informatiques, des sciences des données et de mathématiques.⁸⁰

Ces définitions préalables permettent d'identifier les domaines sur lesquels ont porté ce travail. Une partie sur l'apprentissage machine a permis de comparer les performances de différentes méthodes de ML par rapport aux méthodes classiques dans l'évaluation de l'âge dentaire (Demirjian et Willems). La deuxième partie du travail, a consisté à évaluer les méthodes d'apprentissage profond. Dans la suite du manuscrit, les méthodologies de l'apprentissage machine et profond ont été présentées. Puis les résultats obtenus ont été mis en perspective avec les derniers résultats de littérature ainsi que les questionnements qu'ils ont soulevés.

3.1. Définition du champ d'action de l'apprentissage machine

3.1.1. Méthodologie

Les méthodes de machine learning sont basées sur un apprentissage supervisé ou non supervisé. (Figure 6) Dans l'**apprentissage non-supervisé**, l'algorithme vise à identifier des regroupements naturels dans les données d'entraînement (exemple de la clusterisation). Ce sont des tâches difficilement ou non réalisables par un humain. Un **apprentissage supervisé** nécessite une phase d'entraînement, où l'algorithme utilise des exemples annotés dont il tire un modèle ou une fonction de prédiction de façon à pouvoir prédire avec le plus de précision possible les éléments connus.

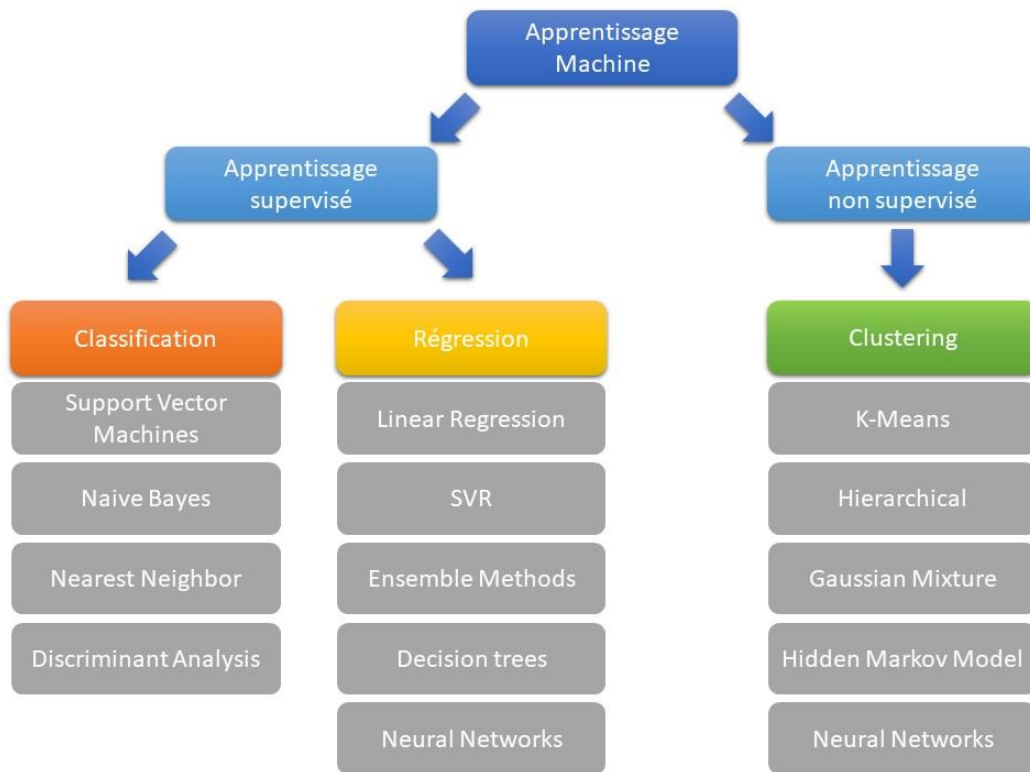


Figure 6: méthodes de machine learning disponibles pour réaliser un apprentissage non supervisé ou supervisé de classification ou de régression.

Cette méthode vise à accomplir des tâches que le professionnel de santé peut déjà accomplir sans problème (par exemple lorsque la tâche est très répétitive), ou bien aider le professionnel de santé à intégrer une masse de données complexe qu'il n'aurait pas pu faire seul. Il constitue donc une aide pour l'utilisateur. En médecine, son utilisation est importante à plusieurs niveaux, mais des exemples emblématiques permettent la détection automatique des rétinopathies diabétiques et l'aide au diagnostic des cancers de la peau, en utilisant un type d'apprentissage machine appelé réseaux

neuronaux.⁷⁷ L'apprentissage supervisé permet de réaliser deux tâches dont la **classification**. Dans ce cas le résultat est ordonné lors d'un choix binaire ou parmi plusieurs classes. Par exemple, partant d'une radiographie panoramique, il sera possible de dire si un individu est majeur ou mineur ? L'autre tâche d'apprentissage supervisé est de prédire une valeur numérique. Par exemple, quel est l'âge d'un individu connaissant les stades de maturation dentaire radiographiques des 7 dents mandibulaires permanentes gauche ? Cette tâche prend le nom de **régression**. Les éléments qui communiquent des informations pour la régression sont appelés **prédicteurs** et les solutions associées à ces derniers sont appelées **labels**.

L'apprentissage machine peut être basé soit sur des exemples ou « instances » soit sur des modèles.⁸¹ **L'apprentissage basé sur des instances** ou apprentissage basé sur la mémoire compare les nouvelles données avec les données stockées en mémoire. **L'apprentissage basé sur un modèle** généralise une règle à partir des données connues pour pouvoir traiter les nouvelles données. Dans ce travail, les deux types d'apprentissage ont été explorés.⁷⁹ Comment estimer l'âge d'un individu vivant ? Tel est l'objectif de ce travail de thèse.

Pour ce faire, un apprentissage supervisé, ayant pour but de réaliser une tâche de régression, a été utilisé. Quatre étapes ont été nécessaires.

3.1.1.1. Acquisition et préparation des données d'entraînement

La première étape de l'apprentissage machine supervisé consiste en **l'acquisition des données d'entraînement**. Ces données sont ensuite normalisées, transformées et nettoyées pour obtenir un ensemble de données d'entraînement utilisable par les algorithmes. Les données sont annotées si nécessaire.⁸² C'est-à-dire que les prédicteurs sont associés à un label, à une cible qui représente la sortie souhaitée de l'algorithme de ML. Dans ce travail, un échantillon de 3 605 radiographies panoramiques (âge compris entre 2,18 et 24 ans) de 1 734 filles et 1 871 garçons a été collecté rétrospectivement à partir de la base de données radiologique clinique du département de médecine orale du CHU de Toulouse. Ces radiographies ont ensuite été annotées de manière à fournir pour le groupe U16 (âgé de moins de 16 ans) : 7 prédicteurs correspondants aux stades de maturation radiographiques des 7 dents permanentes mandibulaires gauche ; et pour le groupe U24 (âgé de moins de 24 ans) : 11 prédicteurs correspondants aux 7 prédicteurs précédents auxquels s'ajoutent 4 prédicteurs correspondants aux 4 troisièmes molaires. Pour les deux groupes, un prédicteur supplémentaire correspondant au sexe a été ajouté. Pour chaque prédicteur a été déterminé le stade de maturation dentaire selon la classification de Demirjian. Le label, c'est-à-dire la sortie supervisée

du système a consisté à calculer l'âge de l'individu lors la radiographie en soustrayant la date de réalisation de l'examen à sa date de naissance.

3.1.1.2. Entraînement des algorithmes

La deuxième étape concerne la **phase d'entraînement**. Plusieurs modèles de machine learning sont sélectionnés et sont entraînés de la même manière pour fournir chacun une **fonction de prédiction**.⁸² Elle permet, à partir des prédicteurs, de prédire la sortie ou label d'un des éléments de la base de données d'entraînement. Une **fonction de coût** permet de mesurer l'erreur d'apprentissage et de modifier les paramètres des fonctions de prédiction pour améliorer son ajustement aux données d'apprentissage. Prenons l'image d'une cible, Figure 7. La fonction de prédiction envoie la flèche et la fonction de coût mesure l'écart qu'il existe entre la prédiction et le centre de la cible qui correspond au label pendant la phase d'entraînement. Cette étape consiste à trouver les **paramètres** qui minimisent la fonction de coût. Dans un espace 2D, il est possible de trouver une fonction de prédiction polynomiale de degré $n-1$ pour décrire n données connues.⁷⁹ Dans ce cas de figure, la fonction de coût est nul, l'apprentissage est parfait. Mais les chances sont quasiment nulles de prédire la position d'une nouvelle donnée. L'apprentissage machine se trouve dans un état de **surapprentissage** ou overfitting du modèle sur les données d'entraînement. Pour cela des données doivent être séparées des données d'entraînement afin qu'elles restent inconnues de l'algorithme. Les données annotées sont donc séparées en 3 parties : une partie d'entraînement (**training set**), une partie de validation pour modifier les paramètres de l'algorithme (**validation set**), et une partie de test qui reste inconnue pour évaluer la performance du modèle entraîné (**test set**). (Figure 8)

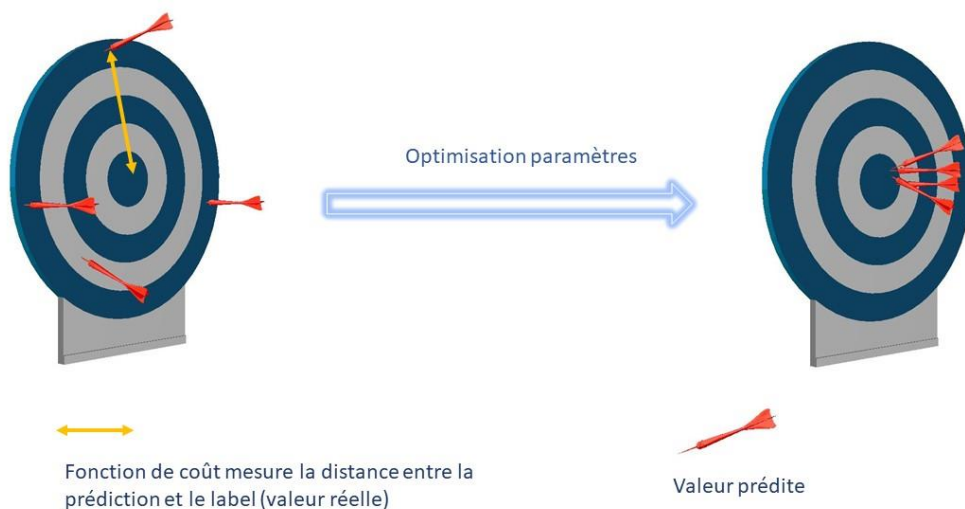


Figure 7 : Représentation visuelle de la fonction de coût permettant de mesurer la distance entre une valeur prédite et la valeur réelle.

Une méthode de validation croisée (**cross validation**) a été utilisée pour empêcher l'adaptation excessive des algorithmes à l'ensemble de données d'entrée et par conséquent d'améliorer les capacités de généralisation des algorithmes (Figure 1Figure 8). Tout d'abord, l'ensemble de données a été divisé en 2 parties : 80 % de données annotées pour l'apprentissage et la validation du modèle et 20 % de données pour tester le modèle. La première de ces parties était à son tour divisée en deux autres parties : apprentissage (80%) et validation (20%). Ce processus est répété 10 fois pour balayer la totalité l'ensemble apprentissage/validation. Enfin, les performances des algorithmes ont pu être évaluées dans le jeux de données de test. Toutes les divisions des ensembles de données ont été effectuées en préservant la distribution des âges et la proportion d'hommes et de femmes, ce qui signifie que les ensembles de données de formation, d'évaluation et de test avaient les mêmes caractéristiques (randomisation).



Figure 8 : Décomposition du set de données annotée en des sets d'apprentissage, de validation et de test avec la méthode de cross validation.

A ce niveau, les **paramètres** des algorithmes de machine learning ont été optimisés par la phase d'apprentissage pour réaliser le plus précisément possible la tâche de régression, dans notre cas l'estimation de l'âge dentaire. Mais chaque algorithme possède des **hyperparamètres** qui peuvent être réglés en amont de la phase d'apprentissage. Le même type de modèle d'apprentissage automatique peut nécessiter des contraintes, des poids ou des taux d'apprentissage différents pour généraliser

différents modèles de données. L'optimisation des hyperparamètres permet de trouver un ensemble d'hyperparamètres qui donne un modèle optimal minimisant une fonction de perte prédéfinie sur des données indépendantes.⁸³ Pour continuer sur l'image de la cible (Figure 9), l'algorithme correspond au moyen d'envoyer le projectile sur la cible. Il existe par exemple, un algorithme de ML de type « jet manuel » et un autre algorithme de ML de type « jet par une arme de projection ». Le réglage de l'hyperparamètre du jet par arme de projection sera de choisir entre un arc, une arbalète ou un lance pierre. Pour chaque arme de projection le réglage des paramètres correspond au réglage de la mire qui est différent entre chaque arme. Il existe plusieurs manières de régler ou optimiser les hyperparamètres telles que la recherche sur grille (grid search), la recherche aléatoire, l'optimisation bayésienne, le gradient de descente stochastique. Dans notre cas, nous avons utilisé la méthode de **grid search** qui consiste à utiliser l'ensemble de validation créé lors de la cross validation et de balayer de manière exhaustive un sous-ensemble spécifié manuellement de l'espace des hyperparamètres.⁸⁴ Il suffit ensuite d'évaluer l'algorithme pour chaque valeur d'hyperparamètre possible. Dans l'exemple de la cible, pour évaluer l'hyperparamètre de l'arme de jet, le lance pierre sera d'abord choisi, sa mire sera réglée, puis le même protocole sera appliqué sur l'arbalète et l'arc.

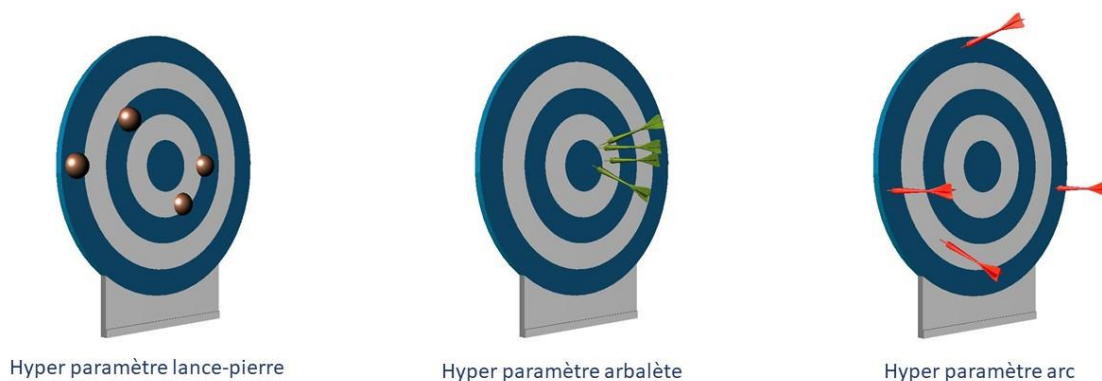


Figure 9: Exemple d'un choix d'hyper paramètre pour un algorithme de ML "jet de projectile" parmi 3 possibilités .

3.1.1.3. Choix de l'algorithme

Une fois les algorithmes entraînés et testés sur l'ensemble des données annotées disponibles, les performances des algorithmes peuvent être analysées. Le seul critère d'examen d'un modèle de ML est sa performance, c'est-à-dire la précision des décisions du modèle. Il faut donc un moyen pour mesurer la distance entre une itération particulière du modèle et les valeurs réelles. Les **fonctions de perte ou de coût** mesurent la distance entre une valeur estimée et sa valeur réelle. Dans notre cas de régression pour prédire l'âge dentaire, la valeur prédite est une valeur continue. Il existe l'erreur

moyenne absolue (mean absolute error : MAE), l'erreur quadratique moyenne (mean squared error : MSE), et la racine carrée de l'erreur quadratique moyenne (root mean squared error : RMSE). Souvent la métrique évaluant la performance d'une méthode d'estimation de l'âge dentaire est basée sur l'erreur moyenne. Mais comme cela a été souligné dans l'article, les méthodes classiques ont tendance soit à sous-estimer, soit à surestimer l'âge dentaire par rapport à l'âge chronologique. Il en résulte qu'une méthode peut être particulièrement performante avec une moyenne proche de zero avec une partie de l'échantillon sous-estimé et une partie dont l'âge est surestimé. Pour remédier à ce problème voici les métriques utilisées dans l'article :

L'erreur moyenne absolue (MAE):

L'erreur absolue moyenne (également appelée perte L1) est l'une des fonctions de perte les plus simples et les plus robustes utilisées pour les modèles de régression. Cette fonction ne tient pas compte de la direction des valeurs positives ou négatives.

$$MAE = \frac{\sum_{i=1}^n |\hat{\text{âge dentaire}}(i) - \text{âge chronologique}(i)|}{n}$$

L'erreur quadratique moyenne au carré (MSE) :

L'erreur quadratique moyenne (également appelée perte L2) est une fonction utilisée qui a l'avantage de pénaliser plus largement les erreurs importantes. Cela permet donc de détecter si en moyenne, les prédictions sont plus proches de la réalité.

$$MSE = \frac{\sum_{i=1}^n (\hat{\text{âge dentaire}}(i) - \text{âge chronologique}(i))^2}{n}$$

La racine carrée de l'erreur quadratique moyenne (RMSE) :

Si l'erreur quadratique moyenne est intéressante pour pénaliser les écarts, sa signification n'est pas représentable. Par exemple, l'âge au carré n'a pas de signification. Prendre la racine carrée de l'erreur quadratique moyenne permet de revenir à des données dont la valeur est interprétable comme une erreur moyenne. Ce n'est pourtant pas la métrique la plus utilisée dans les publications sur l'estimation de l'âge dentaire.

$$RMSE = \frac{\sum_{i=1}^n \sqrt{(\hat{\text{âge dentaire}}(i) - \text{âge chronologique}(i))^2}}{n}$$

3.1.1.4. Généralisation de l'algorithme

La dernière étape du processus d'apprentissage machine supervisé consiste à déployer l'algorithme entraîné sélectionné. Lors de cette étape, l'algorithme peut prédire des valeurs continues pour une tâche de régression, peut réaliser une tâche de classification ou peut être testé sur un nouvel ensemble de données annotées pour le comparer avec l'ancien set de données annotées ou encore pour continuer la phase d'apprentissage. Dans le cadre de ce travail cette étape n'a pas été réalisée, mais elle pourrait servir de base à un travail collaboratif avec d'autres équipes travaillant sur l'estimation de l'âge sur différentes populations. L'avantage de ces méthodes de machine learning sur les méthodes classiques est qu'elles peuvent être transposées et testées directement sans l'intermédiaire de tables de correspondance ou de lecture de graphiques de percentiles.^{4,85}

3.1.2. Méthodes de régression utilisées

3.1.2.1. Support Vector Machine

Les machines à vecteurs de support (support vector machine SVM) sont un ensemble de méthodes d'apprentissage supervisé utilisées pour la classification mais qui peuvent également être utilisées pour des tâches de régression ou de détection de valeurs aberrantes. Les SVM classifient des données en identifiant une frontière de décision linéaire, appelée hyperplan, qui sépare une classe de données de l'autre. Il faut donc que les données soient linéairement séparables. Si ce n'est pas le cas, une transformation par noyau (kernel) permet de transformer les données dans une dimension supérieure de manière à les séparer linéairement.⁸⁶

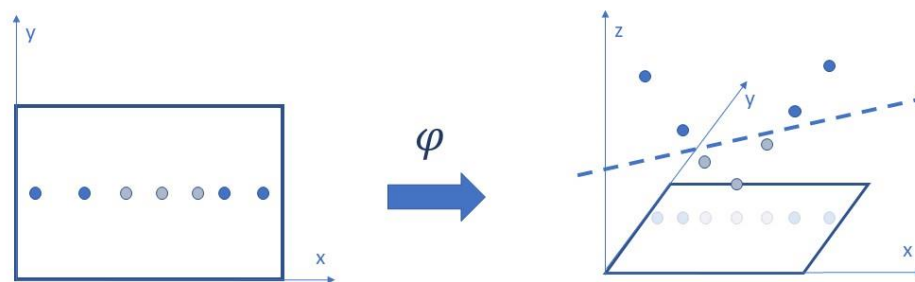


Figure 10: Données non linéairement dissociables dans un espace 2D. Après transformation des données par une fonction Kernel φ dans un hyperespace 3D, ces données sont séparables par un hyperplan.

Les hyperparamètres des SVM sont :

Kernel : un kernel est un noyau, une fonction, qui permet de trouver un espace de dimension supérieure pour les données afin de les séparer linéairement.

Hyperplan : c'est donc le plan ou la ligne qui sépare les classes. Mais pour son utilisation dans une tâche de régression, c'est la ligne qui sera utilisée pour prédire la valeur continue.

Decision boundary : C'est une ligne de démarcation autour de l'hyperplan pour classifier ou pour réaliser la régression. Dans le cas de la régression, les decision boundaries encapsulent les données d'apprentissage.

C'est un modèle de ML peu interprétable.⁸⁷

3.1.2.2. Decision tree

Les arbres décisionnels ou decision tree sont des algorithmes dédiés à l'apprentissage supervisé pour des tâches de classification ou de régression. Il est constitué d'une racine, d'une série de nœuds, de branches et se termine par des feuilles. Les branches sont donc des unités d'un arbre et peuvent contenir des nœuds, des branches et des feuilles. Chaque nœud est un lieu de décision binaire menant à un autre nœud ou une feuille. La création de l'arbre suit des critères qui permettent de décider du choix pour progresser dans l'arbre. Ce critère est un outil de mesure qui peut porter par exemple sur l'entropie d'une solution, le gain d'information après la création d'un nœud, l'erreur quadratique, la racine carrée de l'erreur quadratique ou l'erreur absolue.⁸⁸

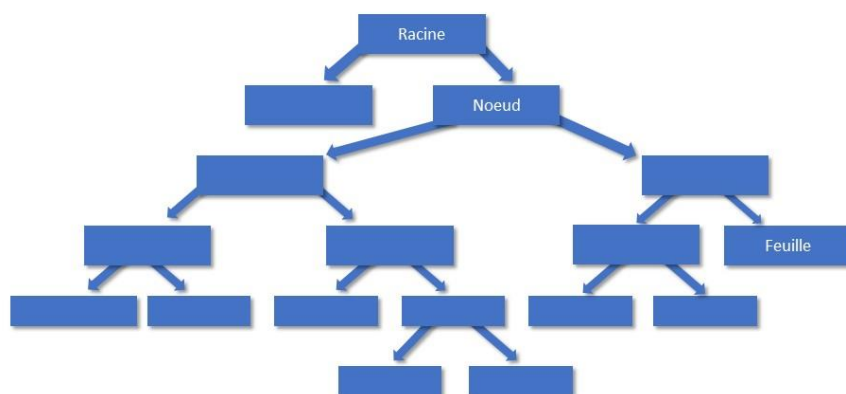


Figure 11: Représentation graphique d'un Decision Tree avec ses racine, nœuds et feuilles.

Les hyperparamètres des decision trees sont :

Le critère de partage d'un nœud : c'est la fonction qui permet de mesurer la qualité d'un partage de nœud.

Le splitter : correspond à la stratégie de partage après un nœud. Dans notre cas, cet hyper paramètre a été choisi comme « random » pour le set U16 et « best » pour le set U24. Cela illustre bien l'importance du choix de l'hyperparamètre pour le modèle de ML. Il montre aussi, qu'en fonction du nombre de prédicteurs ou pour un nouveau set d'apprentissage, toute la procédure d'apprentissage machine doit être reprise pour trouver les meilleurs hyperparamètres.

Max_depth, min_samples_split, min_samples_leaf : sont des hyperparamètres directement liés à la forme de l'arbre. Ils ont un effet direct sur le sur-apprentissage du modèle.

Max_features : est l'hyperparamètre qui considère le nombre de prédicteurs à prendre en compte pour réaliser le meilleur partage en sortie de nœud.

C'est un modèle de ML complètement interprétable.

3.1.2.3. *Bayesian ridge regression*

Ce modèle est un cas particulier de la régression Bayésienne linéaire et de la regression ridge. La régression bayésienne est une méthode de régression linéaire réalisée selon le principe de l'inférence bayésienne.⁸⁹ La regression ridge est une technique de régularisation qui a pour but de minimiser l'erreur entre la valeur prédite et la valeur réelle. Comme expliquée plus haut, il existe plusieurs méthodes pour analyser cet écart et une des méthodes les plus utilisées consiste à réaliser le carré de la différence de ces valeurs. La ridge régression utilise cette méthode.

Les hyperparamètres de la bayesian ridge regression sont :

n_iter : correspond aux nombres d'itérations

Alpha_init et alpha : correspondent respectivement à la valeur initiale d'alpha et de l'hyper- paramètre de forme pour la distribution Gamma antérieure au paramètre alpha.

Lambda_init et lambda : correspondent respectivement à la valeur initiale de lambda et de l'hyper- paramètre de forme pour la distribution Gamma antérieure au paramètre lambda.

C'est un modèle de ML complètement interprétable.

3.1.2.4. *K Nearest Neighbors*

Le principe de cette méthode est de comparer la distance qui existe entre ses « k » plus proches voisins. K est un nombre entier, c'est un hyper-paramètre de la méthode. La distance est un terme qui correspond à une mesure métrique quelconque comme la distance euclidienne par exemple. C'est une technique qui utilise la totalité des données d'apprentissage pour fournir une prédiction. Ce n'est donc pas un modèle de prédiction, il ne généralise pas à partir des données d'apprentissage. Il est question d'apprentissage basé sur la mémoire ou sur les instances.

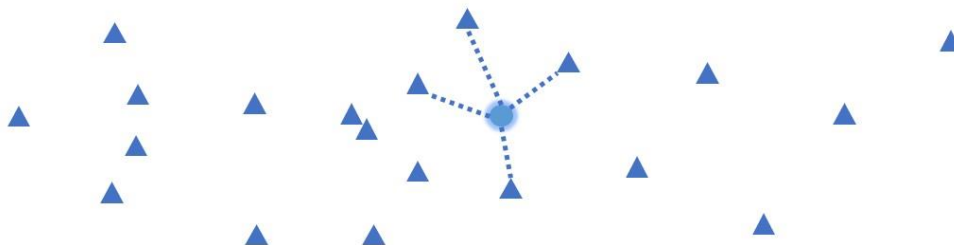


Figure 12: Représentation d'un K Nearest Neighbor avec un k à 4. Le nouvel élément (point) est évalué par rapport aux 4 plus proches données connues (triangles).

Les hyper-paramètres du K Nearest Neighbors sont :

N_neighbors : correspond au k voisins qui sont analysés.

Algorithm : correspond à l'algorithme qui choisira les k plus proches voisins.

Metric : correspond à la métrique utilisée pour le calcul de la distance avec les k plus proches voisins.

C'est un modèle de ML complètement interprétable.

3.1.2.5. *Régression Polynomiale*

Une analyse de régression en machine learning sert à prévoir la valeur d'une variable dépendante connaissant des variables indépendantes. Les variables dépendantes correspondent à la valeur de sortie de l'algorithme, l'output ou la valeur du label. Les variables indépendantes correspondent aux prédicteurs, aux variables d'entrée de l'algorithme. Dans le cadre de cette étude le

set d'apprentissage comprend plusieurs prédicteurs (entre 8 et 12 respectivement pour U16 et U24) et une seule valeur prédite (âge). Dans ce cas la régression est appelée régression linéaire multiple. La régression polynomiale en est un sous-ensemble où les variables indépendantes sont sous la forme d'une équation polynomiale pour lui permettre des relations curvilignes avec les variables dépendantes.

Pour une variable dépendante y et des variables indépendantes x_n où n représente le nombre de prédicteurs, les équations de régression sont les suivantes :

Régression linéaire simple : $y = b_0 + b_1x_1$

Régression linéaire multiple : $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$

Régression linéaire polynomiale : $y = b_0 + b_1x_1 + b_2x_2^2 + \dots + b_nx_n^n$

Les régressions linéaires polynomiales sont simples mais peuvent décrire toutes les données d'apprentissage possible. Cependant elles ont un inconvénient de surapprentissage. S'il existe une droite décrite par une fonction polynomiale de degré 1 pour rejoindre 2 points, alors il existe une fonction polynomiale de degré $n-1$ qui passera par n points. Mais cette dernière ne permet pas de décrire ou de prédire un point supplémentaire. Par exemple, le degré de la régression polynomiale utilisée pour prédire l'âge du set U24 était seulement de degré 2.

Les hyper-paramètres d'une régression polynomiale sont :

Degree : correspond au degré du polynôme

C'est un modèle de ML complètement interprétable.

3.1.2.6. Apprentissages d'ensemble :

L'avantage des méthodes d'apprentissage d'ensemble est basé sur le paradigme que l'association de modèles de ML de base permet de réaliser une tâche de régression ou de classification de manière plus efficace.⁹⁰ Un ensemble est constitué d'un ensemble de modèles de ML formés individuellement dont les prédictions sont combinées pour répondre à une nouvelle instance.⁹¹ C'est un méta-algorithme qui utilise des algorithmes. Les méthodes d'ensemble peuvent être le bagging, le boosting ou la forêt aléatoire.

3.1.2.6.1. Random forest

Un des inconvénients du decision tree est qu'il peut avoir tendance à surapprendre les données d'entraînement, ce qui aboutit à des prédictions fausses et à une augmentation de la variance sur des données test. Une méthode de random forest correspond à l'utilisation de decision trees sur des échantillons aléatoires du set de données d'entraînement. L'échantillonnage est appelé **bootstrapping**. Chaque modèle entraîné sur un échantillon (ou bootstrap) est ensuite combiné aux autres pour réaliser la prédiction. La combinaison des modèles peut être réalisée de plusieurs manières par agrégation, par collage, par sous-groupes de prédicteurs (random subspaces) ou par sous-groupes de prédicteurs et d'instances (random patches). L'agrégation concernant le bootstrapping permet d'échantillonner avec duplication. Le collage utilise des échantillons sans doublons. Dans le cas d'un bootstrapping avec agrégation, l'apprentissage d'ensemble utilise la méthode de Bagging (Bootstrapping + AGGREGatING).

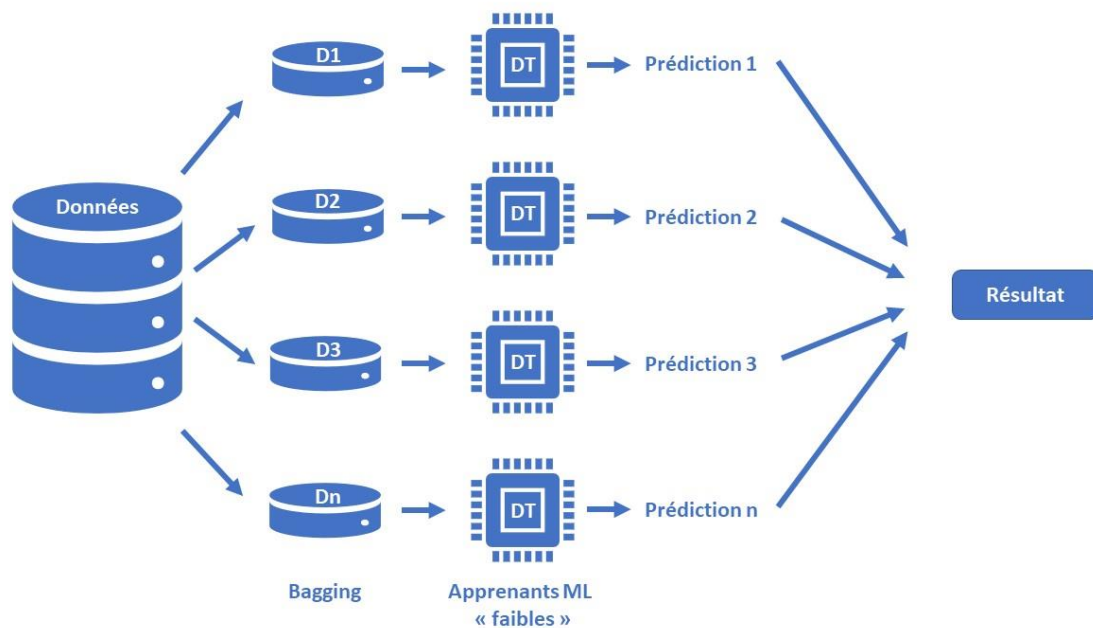


Figure 13: Schéma de fonctionnement d'un Random Forest. Les données sont bootstrappées et passent dans par des Decision Trees (DT) de faible profondeur avant de réaliser la moyenne de chaque prédicteur pour la prédiction finale.

Les hyperparamètres du random forest sont identiques aux decision tree avec des hyper paramètres supplémentaires :

Nombre d'arbres dans la forêt : correspond au nombre de d'unités de decision tree utilisés et donc du nombre de bootstraps.

Bootstrap : l'échantillonnage peut être plus important. Il peut même ne pas avoir lieu et tous les arbres sont alors entraînés sur la totalité des données d'entraînement. Par exemple, pour le set U24 de l'étude ci-dessus, la totalité des données d'entraînement a été utilisée sans échantillonnage sur les 85 arbres paramétrés.

C'est un modèle de ML peu ou pas interprétable, car plus le nombre d'arbres augmente dans la forêt, plus l'algorithme devient difficile à interpréter.

3.1.2.6.2. AdaBoost

L'AdaBoost est également une méthode d'apprentissage par ensemble et elle ajuste une séquence d'apprenants faibles (comme de petits arbres de décision) sur des versions modifiées de façon répétée des données. Son nom dérive de Adaptive Boosting. Le principe diffère du random forest où chaque ML apprenant était indépendant des autres. Pour l'AdaBoost, l'ensemble des données d'apprentissage est présenté au premier apprenant et chaque instance mal prédite se voit attribué un poids. La totalité des données d'apprentissage passe ensuite dans un nouvel apprenant de ML qui se focalisera sur les poids les plus importants. Cette procédure est répétée selon le nombre d'apprenant ML puis fournit à la fin une réponse qui prend en compte l'ensemble des réponses pondérées des apprenants.⁹¹

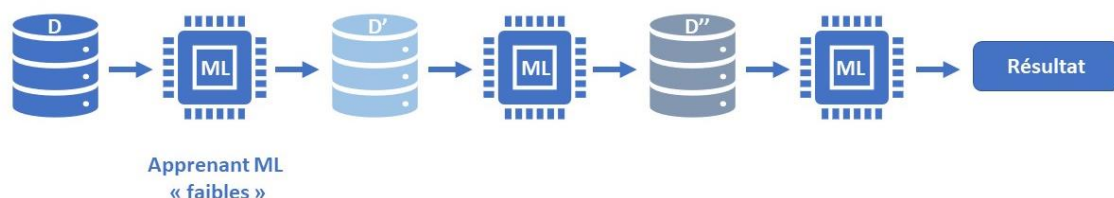


Figure 14: Schéma de fonctionnement du Boosting. Cette méthode attribue des poids aux prédictions fausses pour que les apprenants suivants se focalisent sur ces données.

Les hyper-paramètres de l'AdaBoost sont :

Base-estimator : correspond à l'apprenant ou algorithme ML de bas niveau. Dans le cadre de l'étude présentée ci-dessus, l'apprenant de base a été decision tree dont la profondeur n'excédait pas 3. Par comparaison le random forest utilisait des decision tree dont la profondeur était de 8.

n_estimators : correspond au nombre d'apprenants utilisés. Pour U16 et U24, 60 et 80 estimateurs ont été utilisés respectivement.

Learning_rate : correspond aux poids appliqués à chaque apprenant après chaque itération de boosting.

C'est un modèle de ML plus ou moins interprétable selon le type d'apprenant.

3.1.2.6.3. Stacking

Les deux méthodes d'apprentissage d'ensemble présentée ci-dessus utilisent chacune des apprenants de bas niveau de même type. Par exemple, les apprenants ne seront utilisés que des decision trees. Le stacking utilise lui une combinaison d'apprenants de différents types en tant qu'apprenant de base. Par exemple, un decision tree, un random forest, un K Neighbors regressor. Cette pile d'apprenant fournit ensuite un signal d'entrée à un meta apprenant de base, par exemple une régression logistique. Le stacking traite de la tâche d'apprentissage d'un apprenant de base de méta-niveau pour combiner la prédiction de plusieurs apprenants de base de niveau de base.⁹² Le dernier estimateur final est entraîné par cross-validation.

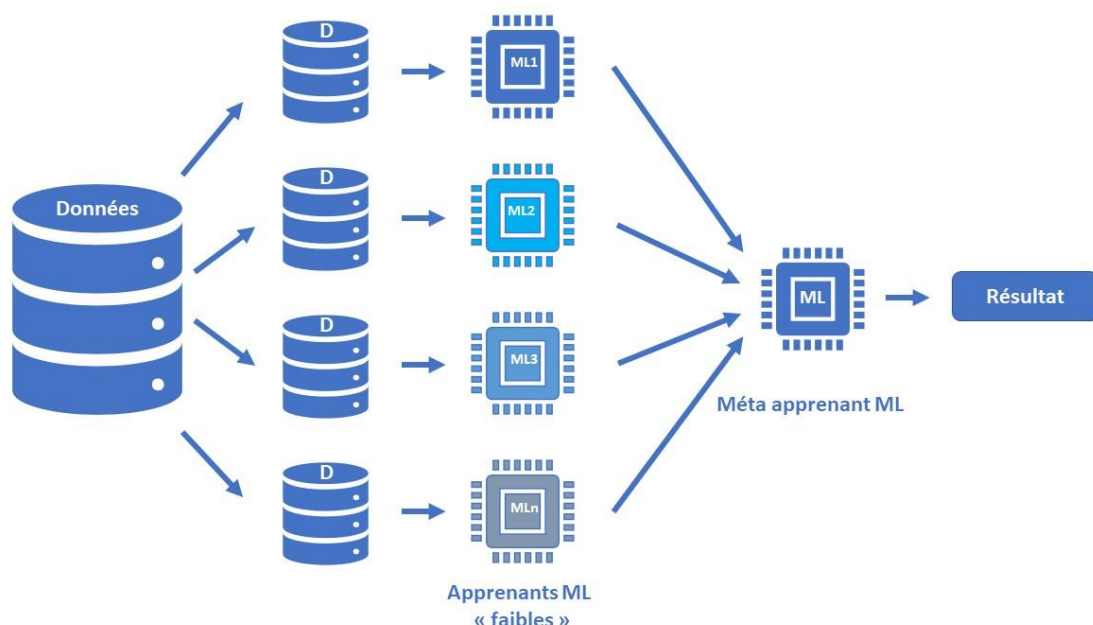


Figure 15: Schéma de fonctionnement de la méthode de Stacking. Les données passent dans différents apprenants faibles utilisant des méthodes de ML différentes. Les sorties sont ensuite dirigées vers un nouvel apprenant faible dit méta apprenant qui fournit un résultat.

Les hyper-paramètres du Stacking sont :

Estimators : correspond à la liste des apprenants de bas niveau utilisés.

Final_estimator : correspond au méta-apprenant dont le rôle est de combiner les sorties des estimators.

Cross-validation : voir plus haut.

C'est un modèle de ML plus ou moins interprétable selon le type d'apprenant.

3.1.2.6.4. Voting

Le ML Voting utilise des apprenants de base de différentes natures. Mais à la différence du Stacking, il n'y pas de méta-apprenant en sortie. Pour la prédiction, un vote est réalisé à partir des sorties de l'ensemble des apprenants de base. Pour la régression le Voting est donc un méta-estimateur d'ensemble qui calcule la moyenne des prédictions individuelles pour obtenir une prédiction finale.

Les hyper-paramètres du Voting sont :

Estimators : correspond à la liste des apprenants de bas niveau utilisés.

Weights : correspond au poids de chaque apprenant régresseur de base

C'est un modèle de ML plus ou moins interprétable selon le type d'apprenant.

3.1.2.7. Multi Layer Perceptron

Le Multi Layer Perceptron est un **réseau de neurones artificiels**. C'est un concept qui a vu le jour l'année suivant la conférence de Dartmouth et l'invention du terme d'intelligence artificielle en 1957. Le courant majoritaire à cette époque portait sur une approche symbolique mais quelques chercheurs en neuroscience postulaient de s'inspirer du fonctionnement des neurones pour que les machines puissent apprendre et agir. Franck Rosenblatt mit au point un neurone artificiel pour établir une prédiction binaire de classification.⁷⁷ Ce neurone possédait des entrées, et une sortie qui correspondaient respectivement à des dendrites et un axone. Les entrées étaient des capteurs de lumière faisant face à un écran sur lequel était projeté une lettre. L'ensemble de l'écran était observé par une multitude de ces capteurs. Chaque capteur se voyait ensuite adjoindre un potentiomètre réglable qui faisait office d'ajouter un poids à l'information reçue par le neurone. Le corps du neurone était composé d'une fonction de sommation des informations pondérées reçues et une fonction d'activation permettait de classer la prédiction. L'ordinateur n'existait pas encore et la « machine »

tenait dans une grande pièce. Si le procédé était capable d'apprendre à discriminer des lettres affichées sur un écran, les limites de ce modèle ont poussé l'approche neuronale dans un état d'hibernation pendant plusieurs années appelé «premier hiver de l'apprentissage machine ».⁹³

En s'inspirant du fonctionnement du cerveau humain, c'est l'association des structures neuronales artificielles simples qui a permis la possibilité d'apprentissage à l'algorithme. Le Multi Layer Perceptron comme son nom l'indique est un réseau neuronal disposant de plusieurs de couches de neurones. L'information circule de la couche d'entrée à la couche de sortie sans boucle. C'est un système à propagation avant ou feedforward. Tous les neurones d'une couche sont connectés à tous les neurones de la couche en aval. Les couches situées entre celles d'entrée et de sortie sont dites cachées car leur fonctionnement n'est pas connu. Chaque neurone de la première couche correspond à un prédicteur. Dans notre cas, la première couche de neurone comprend 8 et 12 neurones respectivement pour les sets U16 et U24. Le réseau est ensuite formé de manière itérative en modifiant les poids de connexion de chaque neurone pour se rapprocher de la sortie souhaitée. La figure ci-dessous représente un schéma de Multi layer Perceptron.

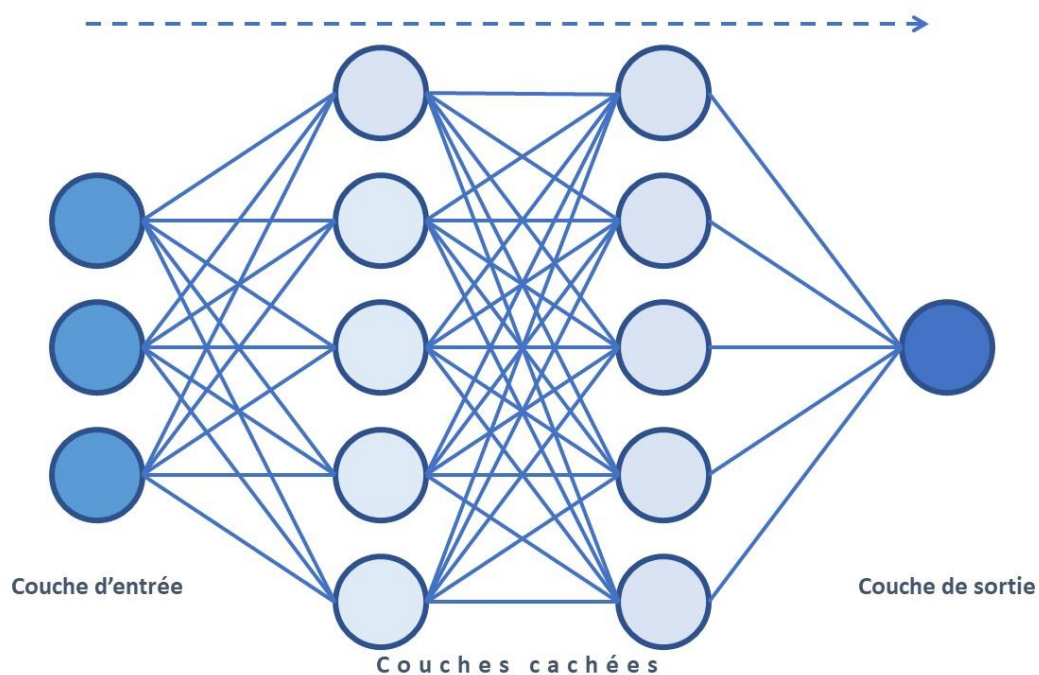


Figure 16: Schéma de fonctionnement du Multi Layer Perceptron. L'information circule dans un sens. Dans cet exemple, les couche d'entrée comprend 3 neurones correspondant chacun à 3 prédicteurs. Chaque de la couche en amont à tous les neurones de la couche en aval jusqu'au neurone de sortie qui fournit la prédiction finale. Il peut y avoir plusieurs neurones de sortie dans la dernière couche. Les couches situées entre les extrémités sont dites cachées.

Les Hyperparamètres du Multi Layer Perceptron sont :

Hidden_layers_sizes : correspond à un couple de valeurs dont la première est le nombre de neurones par couche cachée, et le deuxième correspond au nombre de couches cachées.

Activation : correspond à la fonction d'activation du neurone (par exemple : $f(x)=x$; $f(x)=1/(1+\exp(-x))$; $f(x)=\tanh(x)$; $f(x)=\max(0,x)$)

Solver : correspond au mécanisme d'optimisation des poids des neurones (par exemple la descente de gradient stochastique)

Learning rate : correspond au programme de mise à jour des poids

C'est un modèle de ML non interprétable en raison des couches de neurones cachées.

3.1.3. L'apprentissage profond

Le développement de l'intelligence est passé par plusieurs étapes. Dans son article de 1950, Turing pose la question de savoir si les machines peuvent penser.⁹⁴ Il y présente l'architecture que devrait avoir des machines pour pouvoir être programmées et résoudre des problèmes. Cette architecture est toujours la même sur les ordinateurs actuels. A cette époque deux visions s'opposaient. La première, la plus répandue, visait à développer une approche symbolique basée sur la logique. Les systèmes experts basés sur l'arborescence des données en sont des exemples. La deuxième vision partait d'un principe de biomimétisme copiant le fonctionnement biologique du neurone pour approcher la notion d'apprentissage. Cette approche part du postulat que l'intelligence émerge d'unités simples mises en réseau en paramétrant les connexions elles.⁷⁹

Le premier à obtenir des résultats à partir de cette **approche connexionniste** fut Franck Rosenblatt. A partir d'une rétine artificielle constituée d'une grille de capteurs photo-électriques, le perceptron enregistrerait des images de chiffres ou de lettre. En sortie le perceptron allumait une lampe ou non si l'image était reconnue comme juste. Par exemple, est-ce une lettre « D » ? Il visait également à reconnaître un visage de femme, d'un visage d'homme. L'engouement pour cette technologie fut immense. Dans un article du New-York Times du 13 juillet 1958 intitulé : Electronic « brain » teaches itself, le perceptron y est décrit comme « l'embryon d'un ordinateur électronique dont la Marine espère qu'il marche, parle, voie, écrive, se reproduise lui-même et soit conscient de son existence ». Cet engouement, parfois déraisonnable, s'est poursuivi pendant les années 1950-1960, jusqu'à la sortie du livre Perceptrons en 1969 de Marvin Minsky, défenseur de l'approche logique.⁹⁵ La simplicité de la démonstration de Minsky, sur l'impossibilité des perceptrons à résoudre des problèmes non linéaires à mis à l'arrêt les recherches dans le domaine connexionniste. Cette période a duré 10 ans et est décrite comme le premier hiver de l'intelligence artificielle. Au début des années 1980, des chercheurs comme Yann Le Cun ont développé des améliorations pour les réseaux de neurones avec un algorithme de rétropropagation du gradient.⁹⁶ Mais les puissances des ordinateurs de l'époque n'a pas permis de voir ces avancées développées. Un deuxième hiver de l'intelligence s'ensuit de 1987 à 1993.

Si des résultats intéressants ont vu le jour à partir des années 1993, la date de 2012 a révolutionné le domaine de l'intelligence artificielle. Geoffrey Hinton a amélioré de 10 % les performances de vision par ordinateur sur ImageNet à l'aide d'un réseau neuronal profond de 100 millions de paramètres et utilisant la technologie des cartes graphiques pour faire fonctionner l'algorithme.⁹⁷ Depuis, toutes les technologies basées sur le traitement du signal, de la voix, de la parole ou du texte utilisent cette

catégorie d'intelligence artificielle. C'est un véritable changement de paradigme qui utilise les entrées les plus « brutes » possibles, en s'affranchissant des modélisations explicites des caractéristiques des données mais en utilisant une volumineuse base de données d'apprentissage.⁹⁷ Cardon et al. évoquent le passage d'une machine hypothético-déductive à une machine inductive : « Ce qui était conçu comme la partie « humaine » de la fabrication des calculateurs, le programme, les règles ou le modèle, n'est plus ce qui est introduit dans le système, mais ce qui en résulte » . (Figure ci-dessus)

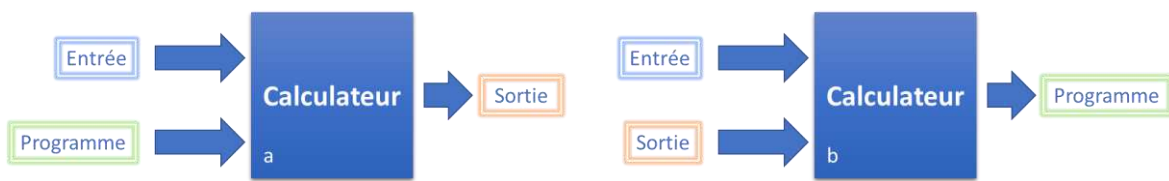


Figure 17: Figures illustrant le changement de paradigme introduite avec l'utilisation des réseaux neuronaux profonds. A) modèle classique de conception d'un algorithme B) calculateur fonctionnant selon le modèle inductif, d'après Cardon et al.⁹⁷

Une des limites de ces modèles qui sera discutée plus loin est leur fonctionnement sous forme de boîte noire. Les motifs de la prédiction sont inaccessibles à la raison humaine. Cette affirmation est remise en cause depuis quelques années avec le développement d'outils pour comprendre les choix d'un réseau neuronal profond.⁹⁸ Il existe aujourd'hui de nombreuses architectures neuronales qui permettent de répondre au mieux à une tâche de vision par ordinateur.

3.1.3.1. Principe de fonctionnement des réseaux neuronaux profonds

Le principe des réseaux neuronaux est basé sur le postulat de l'émergence de l'intelligence à partir de structures de base appelées neurones artificiels. Leur architecture a la particularité de comprendre plusieurs couches de neurones. Les réseaux neuronaux profonds « apprennent » en ajustant les paramètres du réseau de manière à minimiser l'erreur de prédiction. Yann LeCun définit l'avantage essentiel des réseaux neuronaux profonds par leur capacité à apprendre « automatiquement à représenter le signal de manière appropriée ».⁹⁹ C'est-à-dire que le réseau n'a à priori pas besoin de « connaître » les caractéristiques de la cible à prédire pour pouvoir la prédire étant donné qu'il est capable d'extraire ces caractéristiques lui-même. Un autre avantage des réseaux neuronaux est basé sur le principe mathématique selon lequel un empilement de couches « linéaires et non linéaires » est un « approximateur universel » dans le sens où toute fonction peut être approximée à condition que les couches intermédiaires possèdent un nombre suffisant d'unités.⁹⁶

Pour pouvoir réaliser ces tâches complexes, la fonction d'activation binaire « tout ou rien » 0 à 1 ou -1 à 1 a été modifiée pour obtenir une sortie sous forme de sortie sigmoïde, de tangente hyperbolique ou ReLU pour Rectified Linear Unit. Ces changements ont été possible avec l'augmentation des capacités de calcul des ordinateurs.⁹⁶ (Figure 18)

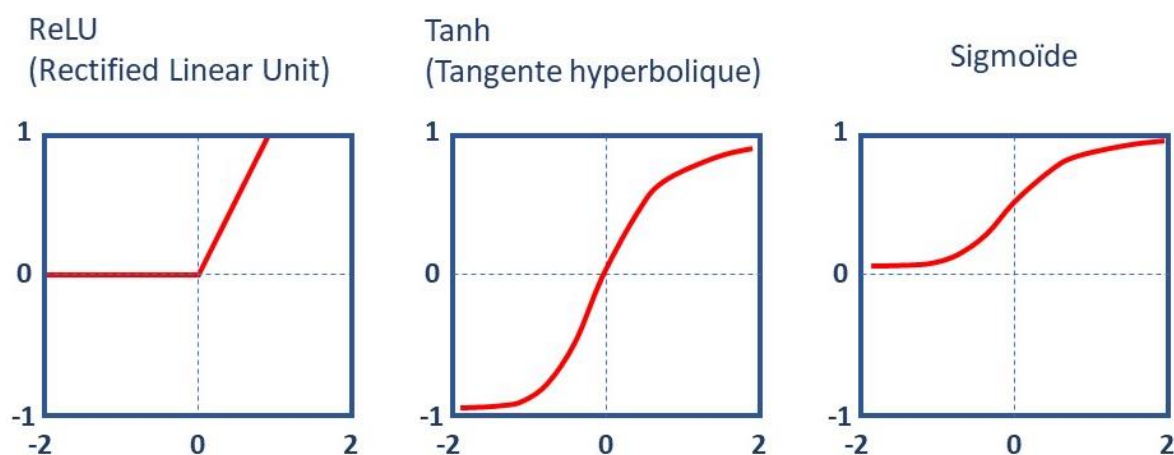


Figure 18 : Fonctions d'activation d'un neurone de type ReLU, Tanh et Sigmoïde.

L'architecture du réseau de neurones, c'est-à-dire des couches et des connexions est déterminée avant le lancement de la phase d'apprentissage. Ce sont les hyperparamètres de ces algorithmes d'apprentissage machine. Les paramètres, correspondent aux poids des connexions des différentes couches de neurones. Dans un réseau convolutif, la rétropropagation de gradient ajuste les poids pour que la fonction de coût diminue et que la précision augmente.⁹⁶ Le choix des réseaux se fait selon la précision recherchée, le temps de calcul, la lourdeur du réseau. Les couches élémentaires de neurones sont de 3 types pour un réseau neuronal profond destiné à la vision par ordinateur. On distingue les couches de convolution, les couches ReLU, les couches de pooling et les couches entièrement connectées. La figure ci-dessous illustre l'exemple du réseau AlexNet pour l'analyse d'image 64x64 pixels en niveau de gris pour estimer l'âge dentaire comme variable continue.¹⁰⁰ (Figure 19)

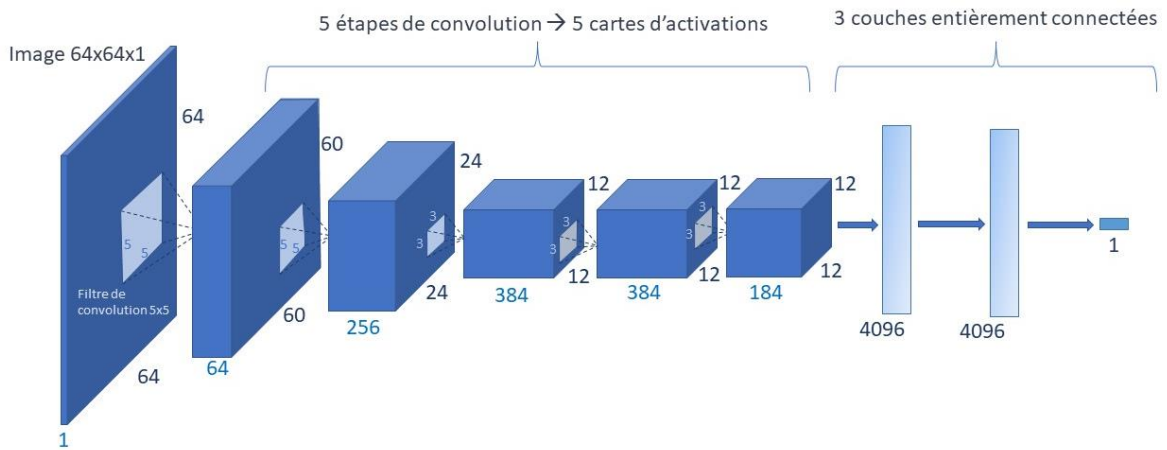


Figure 19: Exemple d'architecture du réseau neuronal profond AlexNet dont les hyperparamètres ont été réglés pour une image de 64x64 pixels à niveaux de gris avec 5 couches de convolution et 3 couches entièrement connectées dont la sortie est un neurone qui réalise une régression linéaire pour estimer l'âge dentaire.

Pour simplifier le propos, prenons comme donnée d'entrée une image sous forme de pixels pouvant prendre une valeur de gris de 0 à 255. La taille de l'image contient 100 pixels soit un carré de 10 par 10. L'image est composée de pixels qui ont une certaine valeur et la première étape consiste à passer un filtre de petite taille sur la totalité de l'image. La Figure 20 illustre la représentation matricielle d'une image et des filtres représentant des formes de bas niveau.

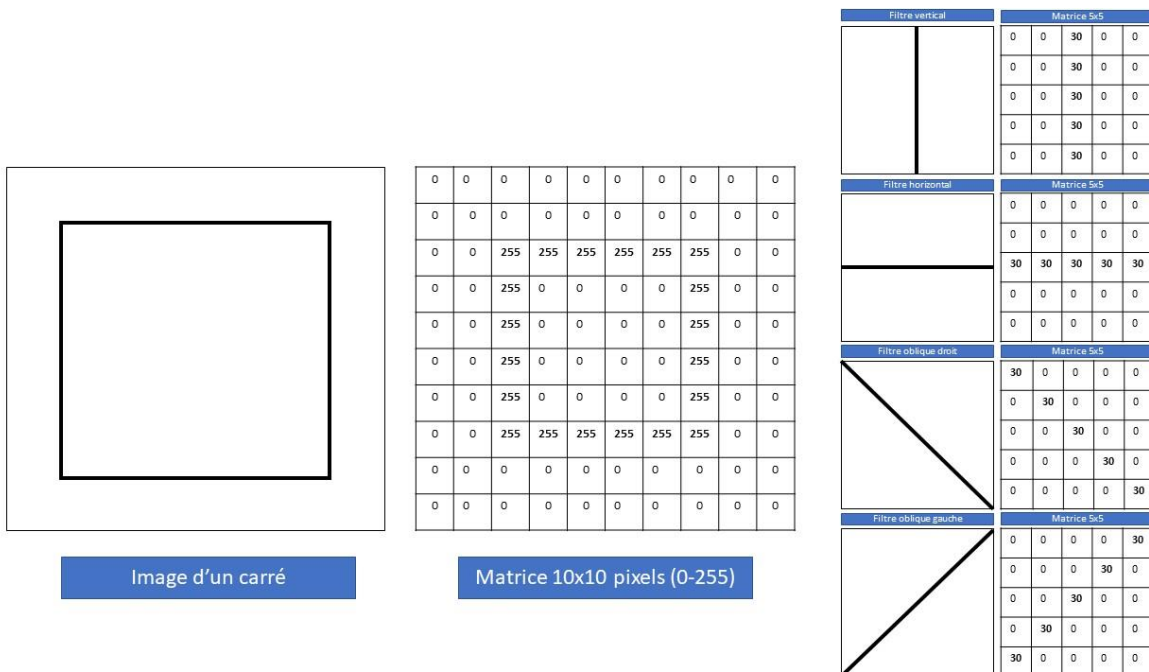


Figure 20 : Illustration de la représentation d'un pixel sur une image de 100 pixels en niveaux de gris et de 4 filtres de 25 pixels représentant des formes de base : trait vertical, horizontal, oblique droit et gauche

Dans notre exemple, ce filtre est composé d'une matrice de dimension 5 par 5 qui contient des valeurs. Ces valeurs sont attribuées de manière à représenter une forme de base comme un trait vertical, un trait horizontal, une courbe ou un point isolé. Cette petite matrice filtre (ou neurone) se déplace pas à pas sur la totalité de l'image d'entrée pour convoluer les données des pixels de l'image. La **convolution** est une opération mathématique sur 2 matrices qui réalise la multiplication et la sommation des cases. Le résultat est une matrice de taille identique ou inférieure à l'image d'entrée dont les pixels appartenant à des formes de base ont été mis en évidence. Dans notre exemple, pour un pas de déplacement de 1 pixel de notre filtre de 5x5 sur l'image de 10x10 pixels, la matrice résultante est appelée **carte d'activation** ou feature map et correspond à un tableau de 6x6 cases.

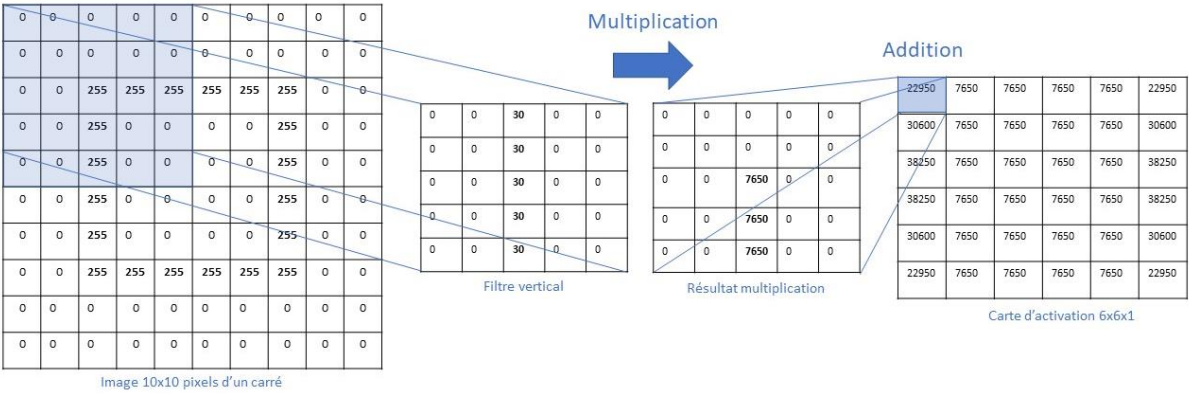


Figure 21: Illustration d'une étape de convolution d'un filtre représentant la forme de base d'un trait vertical sur l'image de 100 pixels représentant un carré. La matrice résultante de l'étape de multiplication puis sommation s'appelle carte d'activation ou feature map.

Pour chaque type de forme identifiée, une petite matrice filtre réalise une convolution. Le résultat global est donc une matrice de profondeur équivalente aux nombres de filtres appliqués. Pour identifier les traits verticaux, horizontaux et les diagonales gauche et droite, 4 filtres ou neurones sont nécessaires dans notre exemple. Donc la carte d'activation résultante aura une profondeur de 4, ce qui correspond à une matrice 6x6x4. (Figure 22)

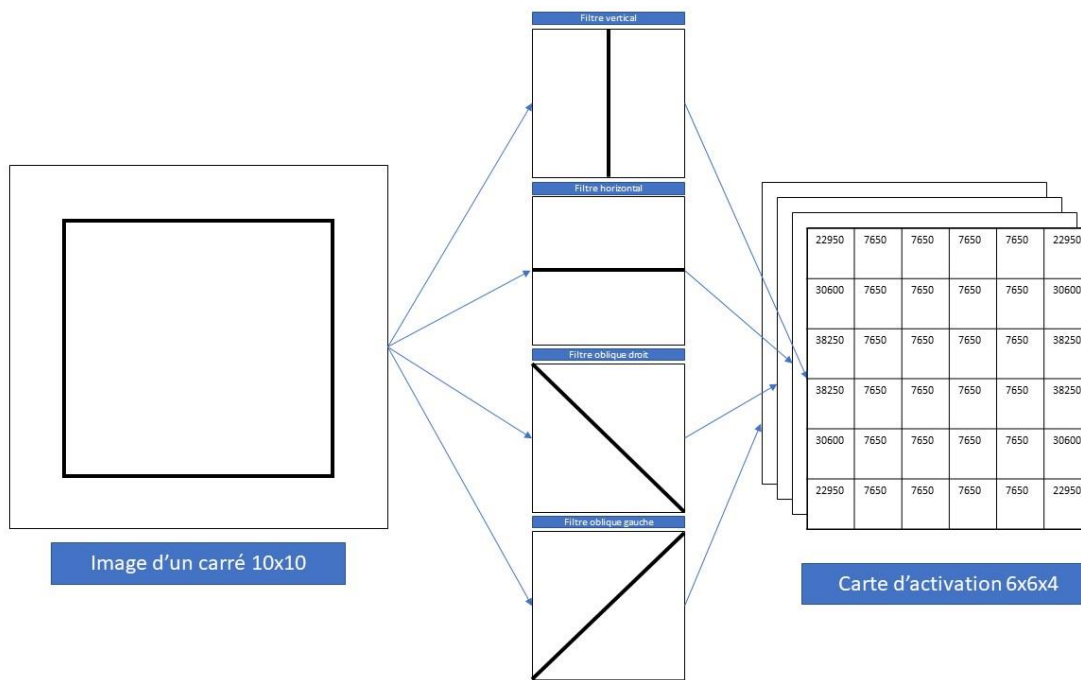


Figure 22: Illustration de la carte d'activation de profondeur 4 après la convolution par 4 filtres, noyaux de convolutions ou neurones dont les formes de base sont des traits verticaux, horizontaux et obliques.

Le résultat de cette première couche de convolution fournit des caractéristiques de bas niveau sur l'image. Dans notre exemple, elle indique la présence de traits simples. La deuxième couche de convolution permet d'identifier des caractéristiques de niveaux supérieurs à partir des cartes d'activation issues de la première convolution. Dans notre exemple, l'association de traits verticaux et horizontaux permet de reconnaître un rectangle ou un triangle. Dans le cadre de neurones ou noyaux de convolution plus nombreux et identifiant un plus grand nombre de formes de base. Les résultats successifs sont des cartes d'activation d'une taille identique ou plus petite dont la profondeur est identique ou supérieure à la carte d'activation précédente. Les formes identifiées sont de plus en plus complexes comme l'identification d'une troisième molaire pour un certain stade de maturation.¹⁰⁰ Plus la taille de la carte d'activation est petite et plus l'image dans son ensemble est analysée et plus elle est profonde et plus des caractéristiques différentes sont analysées.

Après chaque couche de convolution, une **couche d'activation** est appliquée en suivant pour introduire une non-linéarité dans une matrice issue d'opérations linéaires de multiplication et de sommes. Ces couches d'activation utilisent en général la fonction d'activation ReLU (Rectified Linear Units), elles prennent alors le nom de couches ReLU.⁹⁹ Après cette couche, une **couche de regroupement** ou pooling est utilisée. Le résultat est une matrice de taille réduite et de même profondeur que la couche précédente. A la fin du réseau, se trouve une **couche entièrement connectée** qui permet de réaliser la prédiction de classe ou de régression.⁹⁹ Par exemple pour un choix classant un individu dans 3 classes

d'âges, cette couche sera un vecteur de dimension 3 qui utilisera les caractéristiques de plus haut niveau. En revanche dans le cadre d'une régression un neurone unique aura une fonction d'activation permettant de réaliser la prédiction de l'âge comme variable continue.

3.1.3.2. Méthodologie utilisée

La phase d'apprentissage machine à l'aide d'un réseau neuronal profond pour évaluer l'âge dentaire a été débutée au cours de ce travail de thèse. Les étapes de préparation des données d'entrées ont été réalisées à l'aide d'un logiciel mis au point par le Dr Paul Monsarrat. Ce logiciel est une interface utilisateur graphique qui a permis la visualisation des radiographies anonymisées sur un écran d'ordinateur. (Figure 23) Une première fonctionnalité a consisté à définir les zones d'intérêts pour les dents à analyser pour l'estimation de l'âge. Des cadres ont été tracés autour de chacune des 7 dents mandibulaires gauche, ainsi qu'autour des troisièmes molaires quand elles étaient présentes. Cette étape a donc été une étape manuelle de détermination de la ROI. La deuxième fonctionnalité du logiciel a été d'identifier le stade de maturation de l'ensemble des dents identifiées à l'aide de la classification décrite par Demirjian. L'avantage de cette méthode a été de constituer une base de données annotées pour l'approche par apprentissage machine et l'approche par réseau neuronal profond. Dans notre travail, le parti pris de définir des zones d'intérêt autour de chaque dent a été justifié par la thématique de recherche qui visait à garder comme groupe contrôle la méthode manuelle classique décrite par Demirjian.

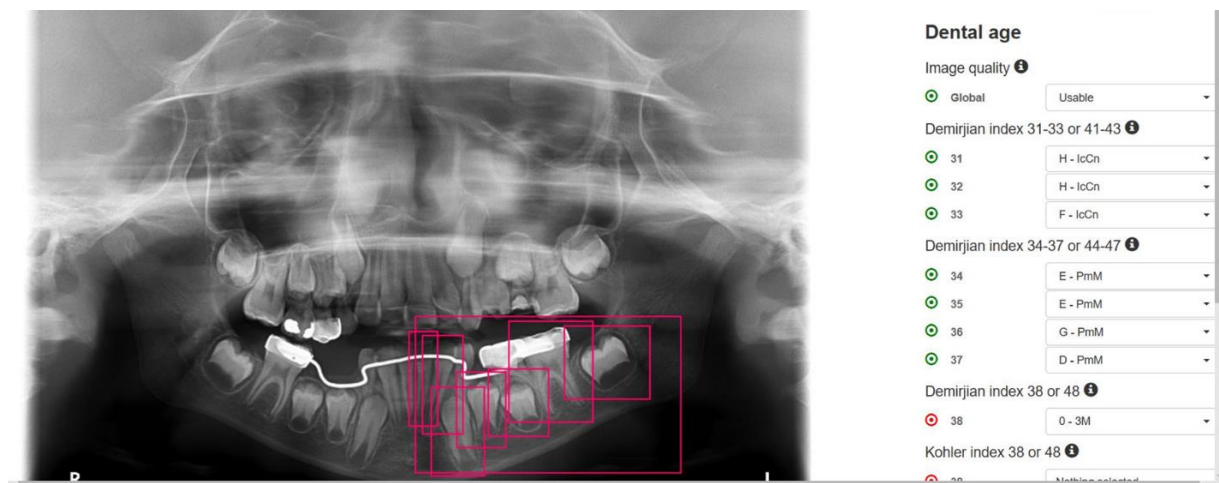


Figure 23: Capture d'écran de l'interface graphique éditée permettant de définir les zones d'intérêt sous forme de cadres et les stades de maturation de l'ensemble des dents analysées pour l'estimation de l'âge dentaire.

La base de données a été constituée par les images croppées des 7 dents mandibulaires plus des 4 troisièmes molaires. La sortie du système a été calculée par la différence entre la date de réalisation de la radiographie panoramique et la date de naissance de l'individu. Ces images annotées de l'âge et du sexe ont ensuite été utilisées directement dans un réseau neuronal profond dans un langage informatique Python.

Le choix de l'algorithme, le choix des paramètres et des hyper paramètres n'a pas été réalisé au cours de cette thèse. Les données préparées pourraient donc être utilisées pour poursuivre ce travail et obtenir des résultats issus d'un apprentissage machine par réseaux neuronaux profonds.

3.2. Mise en perspective des résultats

Pour aborder cette partie, il est intéressant de recontextualiser le sujet. L'estimation de l'âge de dentaire est utilisée pour répondre à trois types de questions : médicale, anthropologique et légale.

En **médecine**, l'âge dentaire est utilisé pour identifier la position d'un individu sur sa courbe de croissance. Il n'y a pas forcément d'évaluation de cet âge dentaire. Un score de maturation peut être attribué le cas échéant. L'intérêt pour une utilisation médicale de l'âge dentaire consiste à pouvoir prédire quelle croissance est attendue ou d'identifier des troubles de maturation des organes dentaires. L'objectif est alors de se baser sur l'évaluation de la croissance dentaire et du développement global pour réaliser un plan de traitement orthodontique ou orthopédique.

En **anthropologie**, l'âge dentaire peut être utilisé pour avoir des informations sur l'état de développement ou de croissance d'un individu et aussi pour inférer un âge chronologique. La phase de dentition, la croissance des germes, des racines et des couronnes dentaires communiquent des informations sur les traits de vie de l'espèce humaine et des hominines. Il y a aussi un intérêt à déterminer l'âge à la mort d'un individu à partir de son squelette ou de son fossile. L'âge chronologique peut alors être estimé à partir d'un âge biologique comme l'âge dentaire.

D'un point de vue **légal**, l'âge dentaire est utilisé pour estimer l'âge chronologique ou plus fréquemment pour classer un individu par rapport à un âge clé légal. Sur les 10 dernières années, à partir d'une enquête portant sur 16 pays, Sykes *et al.* ont établi que les principales raisons médico-légales d'estimation de l'âge étaient liées à l'identification de restes de squelettes (25,7%), les catastrophes de masse (17,6%), les affaires criminelles (16,2%) et les demandeurs d'asile (13,5%).¹⁰¹ Les mêmes auteurs ont noté une augmentation du nombre de demande d'estimation de l'âge pour les demandeurs d'asile et les réfugiés. Dans tous ces cas, la maturation dentaire à un instant donné est utilisée pour répondre à une question légale sur l'âge chronologique estimé d'un individu.

Notre étude portant sur l'estimation de l'âge dentaire à partir de radiographies panoramiques, le champ d'utilisation de cette technique est plutôt orienté vers la médecine légale pour les individus vivants. Les techniques d'estimation de l'âge dentaire sur des individus décédés n'ont pas été abordées ici. Les résultats publiés sur le sujet par rapport à ceux obtenus seront discutés dans un premier temps. Ensuite la thématique de l'automatisation de la méthode d'estimation de l'âge pour simplifier et améliorer la procédure sera débattue. Enfin une réflexion critique sur l'utilisation de ces méthodes d'estimation de l'âge dentaire sera menée.

3.2.1. Etat de la littérature depuis 2021

Au cours de l'année 2021, ont été publiées 62 études sur PubMed concernant l'estimation de l'âge dentaire sur des individus vivants. Parmi elles, 64% ont évalué une méthode classique d'estimation de l'âge et 21 % ont évalué l'impact d'un syndrome sur l'âge dentaire. Cinq méta-analyses ont été réalisées en 2021 sur des méthodes d'estimation de l'âge dentaire.¹⁰²⁻¹⁰⁶ En plus des études de validation d'estimation de l'âge à partir de radiographies panoramiques, d'autres méthodologies ont été testées. La tomодensitométrie à faisceau conique a été utilisée dans 6 études.¹⁰⁶⁻¹¹¹ Pour les méthodes utilisant l'intelligence artificielle, 7 études ont été publiées en plus de la nôtre.¹¹²⁻¹¹⁸ Ces données sont concordantes sur le fait que l'estimation de l'âge dentaire à partir de radiographies dentaires sur des individus vivants est un sujet scientifique toujours actif alors que les méthodes sont décrites depuis des dizaines d'années.¹¹⁹

La méthode de **Demirjian** d'estimation de l'âge fêtera bientôt ses 50 ans d'existence et elle est toujours la méthode la plus utilisée actuellement.^{4,67,120} Depuis 1973, ce sont plus de 150 études qui ont été réalisées pour calculer un score maturité dentaire de Demirjian original ou modifié en 1976.⁶⁷ La méthode de Demirjian a aussi été utilisée partiellement pour son système d'annotation des 8 stades de maturité de A à H et aussi pour évaluer la croissance de la troisième molaire pour des jeunes adultes. Dans ce cadre d'utilisation plus large, ce sont plus de 550 études référencées par PubMed avec des publications plus nombreuses depuis les années 2000. Une méta-analyse a été publiée en 2021 par Hostiuc *et al.* sur l'utilisation de la méthode de Demirjian chez les enfants de moins de 16 ans depuis 1973.⁶⁷ Les 2 méta analyses présentées dans notre article se basent sur 12 et 28 études respectivement pour Jayaraman *et al.* et Esan *et al.*^{6,68} L'intérêt de cette nouvelle étude est qu'elle s'appuie sur 89 études pour réaliser la méta-analyse. Les résultats de l'étude montrent une surestimation globale de l'âge avec 0.48 ans (p=0.95%, IC [0.32-.63]) et 0.51 ans (p=0.95%, IC [0.33-.68]) pour les filles et les garçons respectivement. Ces résultats sont de l'ordre de la demi-année de surestimation de l'âge comme pour les autres méta-analyses qui surestimaient légèrement plus l'âge dentaire des enfants. Notre étude a montré que la méthode de Demirjian surestimait l'âge de 0.7 ±0.073 ans. Ce résultat reste compatible avec ceux de la méta-analyse de Hostiuc *et al.* quand les résultats sont ventilés par zone géographique. Ainsi pour la région Europe/US, la surestimation de l'âge est de 0.45 ans (p=0.95%, IC [0.15-.75]) et 0.49 ans (p=0.95%, IC [0.14-.85]) pour les filles et les garçons respectivement.

En 2001, **Willems** *et al.* ont proposé une nouvelle méthode qui bien qu'utilisant les mêmes stades de maturation de l'organe dentaire permet d'obtenir non pas un score de maturation mais un âge exprimé en années à l'aide de tables de correspondances pour les filles et les garçons.⁵⁴ L'échantillon de

population sur lequel est basé la technique est une population belge, alors que Demirjian s'est basé sur une population française canadienne. La technique est beaucoup moins utilisée que celle Demirjian, même si, comme nous l'avons confirmé dans notre étude, elle plus précise et induit une surestimation moins importante.⁶ Une nouvelle méta-analyse a été publiée en 2021, comparant plusieurs méthode d'estimation de l'âge, dont la méthode de Willems, pour la population brésilienne.¹²⁰ Franco *et al.* montrent que la méthode de Willems est la plus précise des 7 méthodes d'estimation de l'âge testées avec une erreur moyenne de surestimation de 0.05 ans ($p=0.95\%$, IC [-0.04 - 0.14]) filles et garçons confondus. Ces résultats sont plus précis que ceux réalisés pour la méta-analyse de Esan *et al.* dont la surestimation était plus importante avec 0.29 ans ($p=0.95\%$, IC [0.24 - 0.35]) et 0.26 ans ($p=0.95\%$, IC [0.20 - 0.32]) pour les filles et les garçons respectivement. Ces différences s'expliquent par le fait que la méta analyse brésilienne n'inclut qu'une seule étude utilisant la méthode de Willems. Nos résultats sont concordants avec les résultats des autres études.^{102,103,121} La méthode de Willems est plus précise que la méthode de Demirjian pour l'indicateur de l'erreur moyenne. La question à se poser concerne la performance de ces méthodes par tranche d'âge. Cette question est traitée plus loin dans la discussion.

Concernant l'**échantillon** étudié dans la population française, une étude a été menée en 2021 par Brotons *et al.* utilisant la méthode de Demirjian pour estimer l'âge dentaire chez des habitants de la région Provence-Alpes-Côte d'Azur.¹²² Leurs résultats sont similaires aux nôtres avec une surestimation de l'âge de 0.72 ± 0.41 ans et 0.78 ± 0.84 ans pour les filles et les garçons respectivement. Leur échantillon de population comportait 192 individus âgés de 9-19 ans. L'interprétation des résultats est délicate sachant qu'il y a un nombre d'individus inférieur à 10 personnes dans certaines catégories d'âge et des individus âgés de plus de 16 ans.²³ Or la méthode de Demirjian peut être utilisée sur un intervalle de 3 ans à 16 ans alors que pour des classes d'âges supérieures, l'erreur moyenne diminue en raison de l'âge dentaire maximal de 16 ans. Il est donc probable que sur des catégories d'âge centrées sur 9-16 ans, l'erreur moyenne augmente.

Depuis 2021, 8 études portant sur l'évaluation de l'âge dentaire ont été publiées en plus de notre article. Une étude a dressé un bilan sur l'utilisation globale de **l'intelligence artificielle** en odontologie légale avec une partie décrivant l'utilisation des réseaux de neurones profonds pour l'estimation de l'âge dentaire.¹¹⁸ Trois études se sont basées sur l'utilisation de radiographies panoramiques pour estimer en sortie l'âge dentaire en tant que variable continue.^{115,117,123} Deux études ont utilisé un autre outil d'imagerie médical pour estimer l'âge dentaire, à savoir une tomographie volumique à faisceau conique et un examen d'imagerie par résonance magnétique.^{114,124} Et enfin deux études ont utilisé des radiographies panoramiques pour déterminer des classes d'âges à partir de technologies d'apprentissage profond.^{112,116}

Trois études utilisent l'intelligence artificielle avec **en entrée une radiographie panoramique et en sortie un âge dentaire estimé comme variable continue.**

Zabarowicz et al. ont utilisé des mesures linéaires directes à la fois dentaires mais aussi osseuses et dento-osseuses.¹¹⁷ Les auteurs ont réalisé une méthodologie similaire à notre étude en partant d'un set de 619 panoramiques pour des enfants de 4 à 15 ans avec 296 filles et 323 garçons. L'échantillon a ensuite été réparti en training/validation/test set pour entraîner et évaluer 20 types réseaux neuronaux « non profonds ». Parmi ces algorithmes, le multi layer perceptron a été utilisé également dans notre étude. Les prédicteurs utilisés ont été des mesures linéaires réalisées manuellement. L'intérêt de l'étude porte sur deux points. Le premier concerne les 22 prédicteurs utilisés. Le deuxième point d'intérêt concerne l'utilisation de prédicteurs dentaires et osseux. Malheureusement, les résultats de l'étude ne sont pas comparables avec les résultats habituels pour l'estimation de l'âge dentaire. En effet les auteurs fournissent un taux de qualité et d'erreur sur l'ensemble set d'apprentissage, de validation et de test qui ne correspond pas aux paramètres utilisés dans notre travail.

Shen et al. ont publié une étude dont la méthodologie est similaire à la nôtre à l'exception de la méthode d'évaluation de la maturation dentaire.¹¹⁵ La méthode de Cameriere a été utilisée.⁵⁶ Les prédicteurs consistaient en des mesures linéaires manuelles sur 7 dents mandibulaires gauche présentes sur une radiographie panoramique. Les données d'apprentissage ont consisté en 748 radiographie panoramiques d'enfants de 5 à 13 ans avec 356 filles et 392 garçons. Trois modèles de machine learning ont été évalués pour estimer l'âge dentaire : Support Vector Machine (SVM), Random Forest (RF) et Linear Regression. La méthode classique de Cameriere a également été évaluée. Les résultats ont montré que toutes les méthodes de machine learning sont statistiquement plus précises que la méthode classique de Cameriere, ce qui est équivalent à nos résultats. Les ME et RMSE rapportées ont été de 0.004 et 0.625 ans respectivement pour SVM et de -0.004 et 0.623 ans respectivement pour RF. Ces résultats sont meilleurs que ceux que nous avons obtenus pour les mêmes algorithmes. Les ME et RMSE de notre étude ont été de 0.016 et 0.949 ans respectivement pour SVM et de -0.004 et 0.949 ans respectivement pour RF. Cela peut s'expliquer sur les bornes des âges de l'échantillon qui était plus large pour nous avec des enfants âgés de 4 à 16 ans.

Wallraff et al. ont utilisé un réseau neuronal profond pour estimer l'âge dentaire automatiquement à partir de radiographies panoramiques sans passer par l'intermédiaire d'une méthode d'évaluation de la maturation des dents.¹²³ L'algorithme pré-entraîné ResNet18 a été utilisé sur une base de données d'apprentissage la plus importante avec 14 000 radiographies panoramiques d'individus âgés de 11 à

20 ans. En entrée, les radiographies numériques sont fournies sans pré-traitement avec ou sans la connaissance du sexe et en sortie l'âge dentaire est prédit. Une méthode de référence manuelle a été utilisée sur un set de test de 244 images, mais cette dernière n'est pas précisée. Les résultats de ME et MAE pour la méthode automatique non-sexe spécifique et sexe spécifique sont respectivement de -0.3 ± 1.41 et 1.08 ans et de -0.24 ± 1.41 et 1.08 ans. Il est à noter que contrairement aux autres études, la méthode d'estimation de l'âge classique a surpassé le réseau neuronal profond. Ces résultats sont similaires avec ceux obtenus par notre technique MLP de machine learning basée sur les réseaux neuronaux sur l'intervalle d'âge 4 – 24 ans. La ME et la MAE mesurées étaient de -0.009 ± 0.148 et 1.187 ans.

Deux études utilisent l'intelligence artificielle avec **en entrée une source différente de la radiographie 2D et en sortie un âge dentaire estimé comme variable continue.**

Zheng et al. ont proposé une méthode automatique de segmentation de la chambre pulpaire des premières molaires à partir d'une tomographie volumique à faisceau conique.¹¹⁴ La segmentation automatique a été réalisée à partir d'un réseau neuronal profond. Le volume obtenu a permis de calculer le volume de la chambre pulpaire qui a été utilisée en variable prédictive d'un modèle de régression linéaire pour calculer l'âge dentaire. L'échantillon était constitué de 180 patients âgés de 10 à 60 ans, dont 202 dents ont été utilisées pour établir la régression linéaire. L'intérêt de l'étude est de trouver un marqueur biologique avec le volume de la chambre pulpaire qui pourrait être utilisé sur des plages d'âges plus importantes, au-delà des jeunes adultes. Les résultats de l'étude sont un coefficient de corrélation entre l'âge estimé et l'âge chronologique $r = 0.74$, un $p = 0.57$, une MAE de 7.26 ans et une RMSE de 7.26 ans. Ces résultats sont peu précis et ne sont pas utilisables pour des questions médico-légales sur des individus vivants pour des questions d'âge légal. Une utilisation anthropologique sur des questions de traits de vie pourrait être envisagée.

Štern et al. ont utilisé une méthode d'imagerie par résonance magnétique associée à des réseaux neuronaux convolutionnels profonds pour une tâche d'estimation ou de classification d'un âge biologique à partir de 3 prédicteurs : la main, la clavicule et la troisième molaire mandibulaire.¹²⁴ L'échantillon comprend 322 individus âgés de 13 à 25 ans. Trois architectures de fusion des réseaux ont été testées. Le modèle de fusion tardif des réseaux neuronaux est le plus précis avec une MAE de 1.01 ± 0.79 ans. Ce résultat est comparable avec les méthodes de machine learning de notre étude sur le groupe d'âge 4 – 24 ans. La méthode de Štern est intéressante car elle utilise des prédicteurs de nature différente avec l'estimation de l'âge par les dents, la clavicule et la main. Cela a permis de vérifier que les informations portées par les dents et les os sont complémentaires et répondent aux recommandations de l'AGFAD d'utiliser une analyse multi-factorielle dans le cadre d'une estimation

médico légale de l'âge. Les auteurs soulignent que l'importance de la troisième molaire est plus faible que la clavicule au-delà de 18 ans et la main en-dessous de 18 ans.

Deux études utilisent l'intelligence artificielle avec **en entrée une radiographie panoramique et en sortie un classement dans plusieurs catégories d'âges**.

Kim et al. ont utilisé des réseaux neuronaux profonds pour une classification en classes multiples d'âges.¹¹² Ce procédé est un intermédiaire entre une régression pour une variable continue et une classification binaire. Dans le cadre de l'estimation de l'âge, les classes d'âges peuvent être réparties avec des intervalles plus ou moins grand et donc se rapprocher plus ou moins d'une régression. Les auteurs ont entraîné le réseau neuronal profond pré entraîné ResNet-152 à partir des 4 premières molaires. La base de données est constituée 1 586 radiographies panoramiques d'individus âgés de 0 à plus de 60 ans. Deux options de classes multiples ont été testées : la première avec 5 classes (0 – 19, 20 – 29, 30 – 39, 40 – 49 et plus de 50 ans) et la deuxième avec 3 classes (0 - 19, 20 – 49 et plus de 50 ans). Les résultats montrent que la combinaison des prédicteurs permet une amélioration de la précision. Dans cette étude, chaque molaire correspond à un prédicteur et donc la combinaison des 4 molaires permet une première amélioration de la précision. La deuxième a consisté à combiner les précisions issues des modèles à 3 et à 5 classes d'âges. Les résultats sur la précision des tâches de classification multi-classes ne sont pas comparables avec nos résultats de régression. Si les auteurs mesurent une précision de 89.05 à 90.27%, les matrices de confusion montrent des erreurs de classification dans toutes les classes de plus d'une classe d'écart dans le modèle à 5 classes. Ces erreurs sont les plus marquées pour la classe 30 – 39 ans, mais la classe d'âge la plus critique de 0 – 19 ans est également touchée. Même avec un regroupement des classes de 5 à 3, des individus de moins de 19 ans sont classés dans les patients de plus de 60 ans. Pour une utilisation en médecine légale autour d'âges clés, cette méthode ne peut donc pas être préconisée.

Guo et al. ont comparé plusieurs réseaux neuronaux profonds entre eux et avec une méthode de classification manuelle binaire pour les âges clés de 14, 16 et 18 ans.¹¹⁶ L'intérêt de leur méthode est l'utilisation d'une base de données comprenant 10 257 radiographies panoramiques d'individus âgés de 5 à 25 ans. La méthode manuelle a utilisé les stades de Demirjian pour annoter les degrés de maturation des 8 dents mandibulaires gauche. Des équations de régressions linéaires ont été obtenues ensuite pour chaque âge clé étudié. Les radiographies croppées sur la région mandibulaire gauche ont été utilisées directement par le réseau EfficientNet ou le réseau SE-ResNet. Les auteurs ont étudié deux cas de figure avec les 8 dents mandibulaires gauche ou seulement la troisième molaire gauche en tant que prédicteurs des algorithmes. Si les résultats de classification ne sont pas comparables avec nos résultats de régression, ils ont un intérêt d'évoquer la méthode la plus efficace pour répondre à

une question médico-légale. Est-ce qu'un individu a atteint un âge clé ? Les résultats montrent que toutes les techniques d'intelligence artificielles sont plus précises, sensibles et spécifiques que la méthode manuelle. Les résultats montrent également qu'à partir de l'âge clé de 16 ans, la seule troisième molaire gauche est plus informative que l'utilisation des 8 dents mandibulaires gauche. Cela se traduit par l'utilisation d'une image croppée autour de la troisième molaire seule comme donnée d'entrée au réseau neuronal. Parmi les trois âges clés testés, 18 ans, est celui pour lequel la précision, la sensibilité et la spécificité sont les plus faibles avec respectivement 92,6%, 91,1% et 94,4%. La signification de cette sensibilité implique que sur la base de données environ un individu sur 10 a été classé majeur à tort. Cela correspond à la répartition des stades de maturation de la troisième molaire, dont le stade G est toujours présent dans cette base de données jusqu'à 24 ans. La question se pose alors de savoir si l'intelligence artificielle basée sur l'évaluation de la maturation dentaire de la troisième molaire est un bon indicateur pour classer les individus par rapport à l'âge clé de 18 ans ?

L'état de la littérature pour l'année 2021 au sujet de l'intelligence artificielle utilisée pour l'estimation de l'âge dentaire montre tout d'abord l'intérêt croissant des chercheurs pour l'utilisation des méthodes d'apprentissage machine. Les méthodes classiques d'estimation de l'âge sur les vivants à partir radiographies panoramiques ne sont pas abandonnées pour autant. Elles permettent de comparer de nouvelles méthodes, de valider les méthodes pour des populations différentes et leur application bien normée permet la réalisation de méta-analyses.¹⁰² Les méthodes d'apprentissage machine sont disparates. Elles n'utilisent pas toutes la même méthodologie, la même présentation des résultats et les mêmes objectifs. Mais elles ont toutes pour point commun une tendance à automatiser ou numériser la procédure d'estimation de l'âge dentaire.

3.2.2. Automatisation de l'estimation de l'âge

L'intelligence artificielle est utilisée en médecine légale pour l'étude des morsures, la morphologie mandibulaire, la détermination du genre et de l'âge.¹¹⁸ En ce qui concerne l'estimation de l'âge osseux, l'intelligence artificielle a apporté un nouvel élan, notamment sous le format de concours comme le RSNA Pediatric Bone Age Machine Learning Challenge.^{125,126} L'apprentissage machine permet de résoudre pour l'estimation de l'âge des problèmes de régression ou de classification, que ce soit sur des données tabulaires ou de l'analyse.^{112,118} Le domaine de l'odontologie a également développé des solutions à base d'intelligence artificielle pour la détection et la classification des dents, la détection des lésions apicales, l'aide au diagnostic de la maladie parodontale et le diagnostic de l'ostéoporose.¹¹⁶

Pour l'estimation de l'âge dentaire, la raison la plus souvent avancée est celle qui permettrait de s'affranchir du biais humain et de la subjectivité de l'observateur tant au niveau inter qu'intra observateur.^{112,113,118,124,126} Une autre avancée est le gain de temps fourni par ces technologies. D'une part les procédures demandent du temps pour l'estimation de l'âge individuel mais ce critère est d'autant plus important lorsque le but de l'estimation de l'âge vise à vérifier ou créer un modèle pour un échantillon contenant plusieurs centaines jusqu'à des dizaines de milliers de données. Banar *et al.* ont mesuré le temps nécessaire pour estimer l'âge à 2.72 secondes pour leur méthode automatique contre 15 minutes pour la méthode manuelle.¹¹³ Dans notre étude la méthode de Demirjian prenait moins de 3 minutes. Selon Dallora *et al.*, le temps d'exécution de la procédure est important en raison d'une demande accrue d'estimation de l'âge en lien avec le nombre de réfugiés ou de demandeurs d'asile.¹²⁶ Enfin Shen et al. estiment que les méthodes d'intelligence artificielle sont plus robustes et plus précises que les méthodes classiques d'estimation de l'âge dentaire à partir de radiographies.¹¹⁵

L'automatisation de l'estimation de l'âge dentaire peut être analysée selon les trois phases du processus. La première étape consiste à isoler la région d'intérêt, la deuxième étape consiste à évaluer la maturité et la dernière étape consiste à prédire l'âge dentaire. Pour chacune de ces phases de manière globale ou isolée, l'intelligence artificielle peut être utilisée.

3.2.2.1. Détermination automatique de la région d'intérêt

L'identification de la région d'intérêt est la première étape dans le processus d'automatisation d'un processus de traitement d'image. Cette étape est relativement instinctive pour l'homme. L'apprentissage et l'habitude de lecture des examens radiologiques lui permettent de visualiser la structure d'intérêt. Par exemple sur une image radiologique 2D les structures radiologiques se superposent et la visualisation d'un examen radiologique en 3D nécessite un effort de reconstruction mental du volume en 3 dimensions. Pour un algorithme, une image 2D ou 3D est constituée d'un ensemble de pixels ou de voxels en niveau de gris (pour chaque canal s'il s'agit d'une image couleur). Pour isoler un ensemble de pixels ou de voxels appartenant à une région d'intérêt la première étape consiste à la segmenter ou à l'isoler du reste de l'image.

La segmentation d'une image peut être plus ou moins automatisée. La segmentation manuelle constitue la référence notamment pour contrôler la segmentation semi-automatique ou automatique.¹²⁷ La segmentation est dite automatique quand il n'y a pas d'intervention de l'homme entre le début et la fin du processus. Galibourg *et al.* ont proposé en 2018, une méthode automatique de segmentation de germes dentaires sur des tomographies à rayons X en se basant sur une méthode de

Watershed. Les résultats ont montré que la segmentation automatique permettait de gagner du temps par rapport à une segmentation manuelle. Si la forme générale permet d'être appréhendée à partir d'un CBCT, seul un examen micro-CT permettait selon l'étude d'évaluer le volume des germes dentaires précisément. La limite de cette étude est qu'elle nécessite une délimitation manuelle de la région d'intérêt globale pour chaque examen radiologique avant de d'initier le processus de segmentation.

Ce constat de définition variable de la détermination automatique de la région d'intérêt (ROI) est présent dans d'autres études. Par exemple, l'étude de Štern *et al.* vise à déterminer de manière automatique les ROI pour la main, la clavicule et la troisième molaire à partir d'un IRM.¹²⁴ La ROI de la troisième molaire est identifiée à partir du positionnement automatique de landmarks au niveau des centres des deuxièmes et troisièmes molaires. L'identification des 13 os de la main est obtenu à partir de la détermination de 2 landmarks manuellement. Et la clavicule est identifiée à partir de 2 landmarks placés manuellement. Donc au sein de cette étude, le degré d'automatisation est variable et les auteurs l'ont justifié par l'existence de méthodes existantes permettant d'isoler les ROI dans chaque cas et dans un souci de simplifier la procédure. Zheng *et al.* dans leur étude sur l'étude du volume de la chambre pulpaire ont également utilisé une étape de détermination globale de la ROI avant l'initiation de la segmentation et du reste de la procédure totalement automatisée.¹¹⁴

La segmentation automatique des ROI est aussi utilisée sur des radiographies panoramiques. Certains auteurs ont fait le choix de segmenter les dents.^{113,128} Dans l'étude d'Hemalatha *et al.* la segmentation aboutit à l'obtention des surfaces et des contours de toutes les dents d'une radiographie panoramique.¹²⁸ Une limite à cette étude concerne l'absence de concordance de la segmentation automatique avec une méthode manuelle. Les figures présentées montrent clairement une segmentation partielle des dents qui serviront pourtant à l'estimation de l'âge. Banar *et al.* segmentent automatiquement la troisième molaire et comparent ce résultat avec une segmentation manuelle avec un Dice score de 93 %.¹¹³ Ces deux exemples soulèvent la problématique de la segmentation précise de la dent dans le cadre de l'estimation de l'âge. En effet, la perte d'information liée à la segmentation peut potentiellement entraîner des répercussions importantes sur l'estimation du degré de développement. Banar *et al.* ont montré à ce sujet une baisse de la performance globale de l'évaluation des stades par rapport à l'évaluation manuelle.

Que choisir ? Utiliser une ROI ou segmenter les dents ? Certains auteurs s'en affranchissent en partant du principe que l'algorithme d'apprentissage machine peut identifier lui-même les zones d'intérêt pour les prédictions.^{123,129} D'autres auteurs ont une position intermédiaire avec la création d'une ROI contenant une dent et dans une certaine mesure son environnement immédiat.^{112,116,124} Si une certaine

liberté est laissée pour éviter une sur-segmentation de la ROI, la question peut se poser de l'impact de cet environnement immédiat sur la prédiction de l'âge. Par exemple, pour la troisième molaire, cette dernière étant parfois inclinée sur une radiographie panoramique, la ROI en cadre peut être plus importante que pour une dent verticale. Ce problème a été rencontré dans notre travail de cropping. Pour pallier ce problème, De Tobel et al. ont introduit une étape supplémentaire de redressement numérique de l'axe d'une dent versée avant l'étape de cropping.¹³⁰

En synthèse, l'étape de détermination de la zone d'intérêt est plus ou moins automatisable. Mais le fait de la déterminer, de la segmenter du reste de l'image peut induire soit une sur-segmentation ou une sous-segmentation pouvant engendrer des erreurs sur les prédictions de l'âge. Il semblerait qu'une ROI laissant voir suffisamment d'environnement soit à privilégier. Mais quelle dimension donner à cette ROI ? Cette question ne semble pas avoir été évaluée dans la littérature. En ce qui concerne les méthodes d'apprentissage profond par réseaux de neurones, elles ne nécessitent pas forcément de déterminer la ROI si l'algorithme évalue directement l'âge dentaire sans passer par la phase intermédiaire d'estimation du stade de croissance.

3.2.2.2. Détermination du degré de développement

Dans les méthodes classiques d'évaluation de l'âge dentaire à l'aide de radiographies panoramiques, la deuxième étape consiste à évaluer un paramètre pour déterminer le degré de développement de la dent. L'évaluation passe par une mesure directe ou indirecte sur une ou plusieurs dents. Pour l'apprentissage machine, cette phase n'est pas forcément utilisée pour l'apprentissage profond opérant sur l'analyse d'image.^{112,129} Dans les autres cas, l'apprentissage machine automatise plus ou moins cette étape.

Certaines études visent à établir si les stades ou les informations métriques de maturation peuvent être utilisés de manière plus performante à l'aide d'algorithme d'apprentissage machine. Dans cette optique, l'évaluation de la maturation est réalisée de manière manuelle et ces informations sont utilisées comme prédicteurs ou entrées pour l'algorithme. Ces études ne visent donc pas à automatiser l'évaluation de la maturation des dents.

Enfin des équipes travaillent sur l'évaluation automatique des stades. Deux possibilités sont évaluées et utilisent obligatoirement une étape de segmentation préalable. La première concerne les méthodes métriques directes et la deuxième l'évaluation d'un stade de maturation. Pour les méthodes métriques, l'évaluation d'une distance, d'une surface ou d'un volume, est réalisée à partir de la

segmentation d'une image radiologique 2D ou 3D. Zheng *et al.* en évaluant le volume de la chambre pulpaire des premières molaires, ceci permettant d'obtenir une mesure de cet indicateur biologique dont le volume est corrélé à l'âge chronologique.¹¹⁴ Pour les méthodes indirectes, la ROI sera utilisée par un algorithme d'apprentissage machine d'analyse d'image. Ce type de traitement est plutôt dédié aux réseaux neuronaux profonds. Banar *et al.* et De Tobel *et al.* ont publié des articles pour évaluer de manière automatique les stades de Demirjian pour la troisième molaire.^{113,130} Les résultats moyens de précision et d'erreur de classification des stades sont de 54 % et 51 % respectivement pour Banar et De Tobel. L'écart de stade moyen est de 0.69 et 0.6 respectivement pour Banar et De Tobel. A l'échelle des stades, les matrices de confusion montrent des écarts allant jusqu'à 7 et 2 stades de différence respectivement pour Banar et De Tobel. Ces résultats ne sont pas en faveur d'une étape de détermination des stades de maturation de Demirjian à l'aide d'un réseau neuronal profond. De Tobel *et al.* ont utilisé une étape de détermination de la ROI large et Banar *et al.* ont utilisé une segmentation stricte de la dent. Donc l'étape préalable ne semble pas influencer le degré de détermination des stades de maturation.

Ces résultats, sont intéressants car ils montrent la supériorité de l'analyse humaine dans l'identification des stades. Cela diffère des résultats d'identification dans d'autres domaines médicaux où la machine tend à égaler ou surpasser l'homme. L'identification des images évoquant des lésions tumorales sur des radiographies pulmonaires ou des mammographies en sont des exemples.¹³¹ En 2018, l'Agence américaine du médicament a autorisé pour la première fois une intelligence artificielle à poser un diagnostic sans être supervisée par un médecin dans le cadre de l'identification des rétinopathies diabétiques.¹³² Deux explications peuvent être avancées ; la première concerne les données et l'autre la tâche à réaliser.

Etant donné que la technique de Demirjian est une des techniques proposant le moins de stades à savoir 8 contre 15 pour certaines techniques, il est raisonnable de penser que l'identification des stades sera moins performante pour l'identification de plus de stades. La question à se poser est de savoir si le stade est réellement informatif pour l'évaluation de l'âge ? Dans notre travail préparatoire à l'utilisation d'un réseau neuronal profond, nous avons fait le choix de se passer de cette étape intermédiaire pour laisser l'algorithme identifier lui-même les zones d'intérêt pour déterminer l'âge dentaire.

En synthèse, l'évaluation du degré de maturation d'une dent est possible pour les méthodes métriques directes. La méthodologie visant à estimer les stades de maturation dentaire ne semble pas nécessaire pour l'estimation de l'âge. Elle induit un biais de classification et alourdit le protocole d'apprentissage machine.

3.2.2.3. Détermination automatique de l'âge chronologique

La détermination de l'âge est l'étape ultime de la procédure d'intelligence artificielle. Deux sorties sont possibles pour un algorithme d'intelligence artificielle qui a été entraîné de manière supervisée. La première manière permet une estimation de l'âge dentaire comme une variable continue. La deuxième manière vise à prédire une classification binaire. Il est aussi possible de réaliser une classification multi-classes dont le résultat est une solution intermédiaire entre une tâche de régression et de classification.

La première difficulté à laquelle se retrouve confronté un chercheur en intelligence en artificielle est de choisir l'algorithme pour la tâche à réaliser. Fjodor Van Veen et Stefan Leijnen ont répertorié pas moins de 29 architectures de réseaux neuronaux différentes en 2019. (asimovinstitute.org) Chaque architecture est elle-même subdivisée en des sous architectures, par exemple en fonction du nombre de couches cachées ou de la base d'entraînement utilisée. Concernant les algorithmes d'apprentissage machine sans utilisation de la vision par ordinateur, 29 algorithmes sont disponibles pour des tâches de régression en apprentissage machine supervisé avec la bibliothèque *SciKit Learn* pour le langage de programmation Python. Dans notre cas, 10 algorithmes ont été testés. Les résultats ont montré d'une part que les algorithmes n'avaient pas le même niveau de performance et d'autre part que les algorithmes les plus performants pour une base de données n'étaient pas les mêmes pour une autre base de données.

Ces résultats mettent en évidence l'absence d'algorithme universel. Les termes d'intelligence artificielle, d'apprentissage machine recouvrent une réalité beaucoup plus dense que le laissent supposer ces termes génériques. La comparaison de ces algorithmes est alors compliquée par la multitude d'outils possibles. D'ailleurs le choix d'un algorithme plutôt qu'un autre n'est pas toujours spécifié dans les études. Pourtant la méthodologie recommandée est de tester plusieurs modèles et de choisir le plus performant. Cette étape est généralement automatisée et permet d'isoler un ou plusieurs algorithmes pour poursuivre l'expérimentation. Dans notre cas, nous avons opté pour une stratégie de choix des algorithmes à l'aide d'une revue de la littérature recensant différentes méthodes d'apprentissage machine pour évaluer l'âge osseux.¹²⁶

Selon les algorithmes choisis, deux options d'estimation de l'âge en tant que variable continue sont possibles en fonction du type de prédicteur utilisé. Il peut prendre soit la forme d'une variable unique discrète ou continue ou soit la forme d'une collection de pixels issus d'une image. Dans le premier cas, des algorithmes de machine learning basés sur des statistiques ou des réseaux neuronaux peuvent

être utilisés. Dans le second cas, les réseaux neuronaux profonds dédiés à la vision par ordinateur sont utilisés et appartiennent aux méthodes d'apprentissage profond.

3.2.2.3.1. Prédicteurs issus d'une variable de maturation continue ou discrète

Dans le cadre des études évaluant un stade de maturation de manière manuelle ou automatique, ces informations sont utilisées comme prédicteurs ou entrées pour l'algorithme. En ne modifiant qu'un seul paramètre dans la méthodologie classique, à savoir l'analyse des stades par apprentissage machine, notre étude a évalué quelle était la performance des algorithmes par rapport à la régression originale de Demirjian.¹³³ Les résultats ont montré que tous algorithmes d'intelligence artificielle étaient plus performants que la méthode de Demirjian et de Willems. Shen *et al.* ont utilisé la même méthodologie mais avec la méthode métrique directe de Cameriere.¹¹⁵ Les auteurs ont mesuré sur des radiographies panoramiques les ouvertures d'apex et les longueurs de dents qui ont été utilisées comme prédicteurs. Les résultats ont également montré la supériorité des algorithmes d'intelligence artificielle sur la formule de régression originale de Cameriere.

3.2.2.3.2. Prédicteurs issus de pixels ou voxels d'une image

Ces méthodes s'appuient sur la vision par ordinateur. Elles sont constituées de couches de neurones qui analysent des portions d'images de plus en plus réduites. Les poids et les fonctions d'activation des neurones sont paramétrés pendant la phase d'apprentissage grâce une fonction de coût qui vise à améliorer la prédiction finale en réduisant l'écart entre la prédiction et la valeur réelle. Elles peuvent utiliser selon les études des images de radiographies complètes^{123,129} ou des images dont la ROI a été réduite automatiquement ou manuellement pour focaliser l'analyse du réseau neuronal.^{112,124} Ces études sont difficilement comparables avec les méthodes classiques car elles s'affranchissent des limites d'âge des méthodes classiques. Vila-Blanco *et al.* ont proposé une subdivision de leurs résultats par rapport aux classes d'âge de méthodes classiques.¹²⁹ Pour les groupes d'âges inférieur à 15 et à 25 ans, l'erreur moyenne et l'erreur absolue sont respectivement de -0.07 et 0.75 ± 0.57 ans et 0.02 et 1.17 ± 1.11 ans. Les résultats sont donc comparables pour les deux méthodologies. Štern *et al.* sur un groupe d'âge de 13 à 25 ans ont obtenu une erreur moyenne absolue de 1.01 ± 0.74 ans. Wallraff *et al.* ont obtenu sur un groupe d'âge de 11 à 20 ans une erreur moyenne absolue de 1.08 ans.¹²³ Nos résultats d'erreur moyenne et d'erreur absolue de la méthode

random forest pour le groupe U16 et U24 ont été respectivement de -0.07 et 0.731 ± 0.025 ans et 0.012 et 1.79 ± 0.038 ans. Il est intéressant de noter que les réseaux neuronaux profonds de vision par ordinateur sont capables d'identifier dans un ensemble de pixels ou de voxels des informations permettant d'inférer une estimation de l'âge dentaire. Comparativement aux méthodes classiques, elles sont aussi voire plus performantes. Comparativement aux méthodes d'apprentissage machine utilisant des prédicteurs unitaires les résultats sont du même ordre de grandeur.

L'avantage de ces méthodes de vision par ordinateur est de s'affranchir de l'utilisation d'un marqueur biologique particulier. L'idée sous-jacente suppose que le réseau neuronal profond est capable d'identifier lui-même les marqueurs biologiques sur l'ensemble de la radiographie panoramique. De cette manière, la limite de croissance des dents ne constitue plus un plafond pour l'estimation de l'âge. Il conviendrait peut-être alors de parler, non plus d'estimation de l'âge dentaire mais de l'âge radiographique. Les plages d'âge de ces techniques peuvent s'étendre de l'enfance à l'âge adulte avancé. Ainsi l'échantillon de radiographies panoramiques inclut des individus de 0 à plus de 60 ans pour Kim *et al.* et de 4.5 à 89.2 ans pour Vila-Blanco *et al.*^{112,129} Ces derniers ont ainsi obtenu une erreur moyenne et une erreur moyenne absolue de 0.12 et 2.84 ± 3.75 ans.

3.2.2.3.3. Explicabilité des méthodes d'apprentissage machine

Si les méthodes d'apprentissage machine peuvent s'affranchir des marqueurs biologiques conventionnels, la question se pose de savoir quels sont les indicateurs permettant de prédire une classe d'âge ou d'inférer un âge dentaire ? Une des caractéristiques des algorithmes est leur explicabilité plus ou moins importante. Les réseaux de neurones sont d'autant moins explicables que le nombre de couches cachées de neurones est grand. Ainsi dans notre étude, les modèles explicables sont : le Support Vector Machine, le K Nearest Neighbors, la Bayesian Ridge Regression, le DecisionTree. Les méthodes d'ensemble utilisant des apprenants faibles comme le Stacking, le Random Forest, l'Adaboost et le Voting ne peuvent pas être considérées comme explicables facilement. Le modèle le moins explicable utilisé dans ce travail correspond au Multi Layer Perceptron qui est un exemple de modèle connexionniste. Dans le cadre d'une utilisation en médecine, l'utilisation d'un réseau neuronal profond pour diagnostiquer la rétinopathie diabétique a été autorisée alors que ce type d'algorithme est considéré comme une « boîte noire ».¹³⁴ Pour une utilisation en médecine-légale l'explicabilité du modèle est essentielle. Nous proposerions de donner suite à notre travail en associant aux techniques de ML un framework d'explicabilité tel que SHAP (SHapley Additive exPlanations).¹³⁵

Cette explicabilité est fondamentale pour comprendre la prise de décision de l'algorithme et pour pouvoir l'améliorer. Les réseaux neuronaux profonds utilisés pour des résolutions de problèmes de vision sont considérés comme des « boîtes noires » dont il n'est pas possible de comprendre le fonctionnement interne. Quand ces réseaux ne fonctionnent pas, quand ils fournissent une prédiction incohérente, il est difficile voire impossible d'en expliquer la raison.¹³⁶ Une nouvelle discipline dans l'intelligence artificielle a donc vu le jour avec pour objectif d'expliquer le fonctionnement et l'efficacité d'un réseau neuronal à des utilisateurs humains.¹³⁷ Les auteurs évoquent l'objectif de surmonter « le dilemme de l'utilisation aveugle » des modèles d'apprentissage profond. Parmi les techniques disponibles, une possibilité est de créer un autre réseau neuronal profond dont le but est d'expliquer les choix du premier et une autre consiste à sonder le réseau neuronal en utilisant d'autres variables d'entrées et en analysant les sorties.¹³⁷ Une autre technique, appelée Grad-CAM a vu le jour en 2016 et est améliorée régulièrement.¹³⁶⁻¹³⁸ Grad-CAM est l'acronyme de Gradient-weighted Class Activation Mapping et elle s'applique aux réseaux neuronaux réalisant des tâches de vision par ordinateur. Cette méthode permet de mettre en évidence sur l'image d'entrée du système, à l'aide d'une carte de fréquentation (ou heatmap), les zones qui ont été utilisées par les neurones du réseau pour prédire la classe ou la variable de sortie.

La méthode de Grad-CAM est utilisée dans les études estimant l'âge à partir d'une image radiographique en entrée pour expliquer et vérifier le comportement de l'algorithme. L'algorithme peut être utilisé pour identifier et analyser un cas particulier dont la prédiction est fautive.¹²³ Il peut aussi servir sur une image représentant la totalité de la base de données sur laquelle une moyenne des zones utilisées par les neurones est mise en évidence avec la heatmap.^{112,129}

Kim *et al.* mettent en évidence qu'en fonction du nombre de classes d'âges en sortie de l'algorithme, le réseau neuronal ne se base pas sur les mêmes zones de l'image.¹¹² Par exemple, pour la classe d'âge 20 – 29 ans, le Grad-CAM a mis en évidence, pour la même architecture de réseau la zone de la chambre pulpaire et la zone inter-radulaire respectivement pour une sortie 3 et 5 classes d'âges. Le Grad-CAM a été développé pour mettre en évidence des objets ou des notions simples à appréhender ou vérifier. Par exemple, dans l'image présentant un chat, si le Grad-CAM ne met pas en valeur le chat, il est évident que le réseau neuronal a mal interprété l'image quel que soit le résultat.

Dans le cadre de l'estimation de l'âge à partir d'images issues de radiographies, il n'y a pas de zone ou d'organe ciblé comme étant juste. Quel que soit le résultat, la zone utilisée par le Grad-CAM est justifiée et non critiquée. Ceci est visible dans l'étude de Wallraff *et al.* où les angles supérieurs de la radiographie panoramique aident au même titre que la racine d'une molaire maxillaire pour estimer

précisément l'âge d'un individu.¹²³ Dans l'étude de Vila-Blanco, un individu de 9 ans et 8 mois a été identifié par le réseau de neurone à l'aide de l'espace d'inocclusion entre 26 et 37 avec le même taux d'activation que la 37.¹²⁹ Toujours dans cette étude, un enfant de 4 ans a été identifié précisément grâce au septum osseux entre 46 et 85 au même titre que la zone de furcation en développement de 36. L'étude de Kim et al. détaille précisément les résultats du Grad-CAM.¹¹² Mais ils justifient ou surinterprètent ces résultats. Par exemple, pour le groupe d'âge 10 – 19 ans, les auteurs interprètent un Grad-CAM qui met en évidence la quasi-totalité de la ROI, comme permettant à l'algorithme de prédire un résultat « en fonction de l'état de l'os alvéolaire et du sinus maxillaire ». Pourtant la partie radio-claire d'une partie du sinus croppé, puisque c'est cette zone qui est mise en évidence, devrait au minimum être questionnée. Une voie de recherche dans l'estimation de l'âge consisterait donc à investiguer plus précisément les zones d'activation sur une radiographie panoramique. Elle permettrait d'identifier des marqueurs « biologiques radiologiques » nouveaux ou déjà validés. Elle permettrait ensuite de déterminer des zones « acceptables » ou « non acceptables » pour juger de l'interprétabilité d'une radiographie panoramique.

En synthèse, les algorithmes d'apprentissage machine apportent une vraie plus-value sur le traitement statistique des prédicteurs pour évaluer l'âge dentaire par rapport aux tables de conversion ou aux équations de régression des méthodes classiques. L'utilisation de la vision par ordinateur pour analyser une partie ou la totalité d'une radiographie panoramique donne des résultats similaires aux méthodes classiques sur la plage d'âge d'utilisation de ces dernières. L'avantage de la vision par ordinateur est de s'affranchir du prédicteur concernant l'organe dentaire et d'envisager des zones ou des rapports non étudiés jusqu'ici par les méthodes classiques. Les plages d'âges investiguées peuvent être étendues au-delà de 60 ans. Les résultats ne montrent pas d'amélioration dans ces plages d'âges par rapport aux méthodes classiques d'estimation de l'âge chez l'adulte. Le questionnement sur l'interprétabilité des résultats obtenus avec un algorithme sous forme de réseau neuronal profond devrait être investigué dans de futures études.

3.2.3. Problématique de la nature des données

En 2017, The Economist titrait pour son édition du 6 mai que la ressource la plus précieuse sur terre n'était plus le pétrole mais les données. Est-ce que cette assertion s'est vérifiée depuis ? Le développement des NFT, non-fungible tokens, qui sont des données numériques représentant des photos, des vidéos ou un contenu audio numérique, a été à l'origine d'investissements massifs ces

derniers mois. Sur les 3 premiers mois de 2021, Time a révélé que 200 millions de dollars avaient été investis dans des données virtuelles en NFT.¹³⁹ L'artiste Pak a vendu une image numérique au format NFT vendue pour 91.8 millions d'euros en 2020. Cette image numérique, appelée Merge, représente une sphère blanche sur fond noir. La prolifération des données numérique à tous les niveaux et évidemment dans le domaine médical apporte un flux continu de données à traiter. Les plus gros pourvoyeurs de données sont les objets connectés qui fournissent des informations en continue. En 2019, le World Economic Forum relatait l'analyse de Statista qui montrait que la totalité des données numériques créées depuis l'origine de l'informatique jusqu'en 2018 se verraient doublées au cours de la seule année 2020.⁷ Les projections montrent également que la totalité des données numériques créées en 2020, seront créées en seulement 8 jours en 2035. L'enjeu pour le stockage et le traitement de ces données est donc un enjeu à l'échelle des états.

Le domaine médical est un créateur important de données. Elles peuvent être utilisées dans le traitement des données massives (big data) et par l'intelligence artificielle pour la réalisation d'une médecine prédictive, une médecine de précision, une aide à la décision, la chirurgie assistée par ordinateur ou la prévention à l'échelle d'une population.¹⁴⁰ La notion de données numériques massives et plus particulièrement de big data désigne les données numériques si volumineuses qu'elles nécessitent des analyses particulières pour en extraire de l'information. La règle des 3V, définit cette notion de big data. Le premier V correspond au volume dont la croissance est actuellement exponentielle avec les données issues de l'internet des objets connectés. Le deuxième V correspond à la vitesse, c'est-à-dire à la fréquence de génération, de capture et de partage des données. Le troisième V correspond à la variété des données qui sont en règle générale non structurées.

Les algorithmes d'intelligence artificielle peuvent être utilisés pour traiter ces données nombreuses et non structurées. Mais la qualité et la quantité des données est essentielle à la bonne résolution d'un problème. Un article de Halevy *et al.* a démontré que la quantité de données importait plus que le type d'algorithme utilisé.¹⁴¹ En testant 4 algorithmes d'apprentissage machine différents, ils ont prouvé que ces derniers obtenaient un plafonnement des performances plus ou moins rapidement. Au même titre que le travail sur l'explicabilité d'un algorithme, la qualité des données utilisées est un paramètre essentiel.¹⁴² L'imagerie médicale numérique est donc un domaine parfait pour l'application de l'intelligence artificielle par la génération de ses données et par la possibilité d'utiliser la vision par ordinateur. Dans les données utilisées pour l'estimation de l'âge quels sont les paramètres influençant les algorithmes d'intelligence artificielle ?

3.2.3.1. Origine des prédicteurs

Dans le cadre de l'estimation de l'âge dentaire chez les vivants les prédicteurs utilisés sont généralement issus d'examen de radiographies panoramiques.^{10,19,30,35} Il existe aussi des techniques radiologiques basées sur des examens radiologiques tridimensionnels comme la tomographie volumique à faisceau conique ou le scanner.^{62,106,107,110} Les recommandations de l'AGFAD, un groupe d'étude sur le diagnostic de l'âge médico-légal de l'Association allemande de médecine légale, recommandent l'utilisation des rayons ionisants et notamment de la radiographie panoramique pour l'estimation de l'âge dentaire.¹⁴³

Une des limites de l'utilisation de ces examens est l'utilisation des radiations ionisantes sur des individus sains en l'absence de justification diagnostique. C'est une des raisons pour laquelle l'estimation de l'âge dentaire n'est pas privilégiée en première intention dans l'estimation de l'âge.²⁶ Pour pallier ce problème, Štern *et al.* ont proposé l'utilisation d'examen d'imagerie par résonance magnétique.¹²⁴ Les données issues de ces examens sont sous la forme de voxels et peuvent être utilisées de la même manière que les examens utilisant des rayons ionisants.

3.2.3.2. Multi prédicteurs

L'origine des données utilisées n'est donc pas une contrainte pour les algorithmes d'apprentissage machine. C'est un véritable avantage pour ces techniques qui peuvent alors s'affranchir ou au moins simplifier l'utilisation de plusieurs types de prédicteurs. Štern *et al.* ont utilisé des prédicteurs issus de l'analyse des os du poignet, de la clavicule et de la troisième molaire.¹²⁴ La fusion des données a néanmoins une incidence sur la prédiction. Dans l'étude, les résultats sont d'autant plus précis que la fusion est tardive. En d'autres termes, les résultats étaient plus précis quand la prédiction opérait à partir des prédictions de chaque catégorie de prédicteurs à l'image d'un système de vote pour une technique d'ensemble. Zaborowicz *et al.* ont intégré dans leurs algorithmes 22 prédicteurs correspondant aussi bien à des indicateurs dentaires qu'osseux.¹⁴⁴

Cette possibilité d'intégrer facilement des prédicteurs permet de prendre en compte le dimorphisme sexuel comme prédicteur dans les algorithmes. Cette donnée apporte une information qui permet d'améliorer la prédiction des algorithmes.^{117,123,129,133} Dans notre étude, les stades de toutes les troisièmes molaires ont été intégrés pour évaluer l'âge dentaire sur des plages plus importantes.

3.2.3.3. *Données incomplètes*

Les méthodes classiques d'estimation de l'âge chez les vivants ont été mises en place à partir de radiographies sans dents manquantes et avec des individus sans problème de santé avec une croissance normale.^{23,44,54,63} La présence de traitement orthodontique, de prothèses, d'implants, d'éléments externes tels que des boucles d'oreilles, des restaurations conservatrices, des traitements endodontiques, des caries, des restes radiculaires sont des éléments visibles sur des radiographies panoramiques. Si une ou plusieurs dents sont intéressées dans les cas précédents, l'estimation de l'âge dentaire ne peut pas être réalisée. Les méthodes d'apprentissage machine peuvent prendre en compte de manière plus souple des données incomplètes. Vila-Blanco et al. ont réalisé leur étude sur 2 289 radiographies dont seulement un quart étaient exemptes des conditions citées plus haut.¹²⁹ Les résultats montrent une erreur moyenne de 0.36 ± 4.69 ans contre 0.23 ± 2.15 ans sur le set de panoramiques sans défaut.

3.2.3.4. *Quantité de données*

La quantité de données utilisées influe également sur la performance des algorithmes. Par exemple, ImageNet dispose de 14 millions d'images pour entraîner les algorithmes à classer les images par catégories. Dans le cadre de l'estimation de l'âge, le nombre de données varie d'une centaine à plus de 14 000 images. Mais cela est à relativiser car pour identifier des races de chien par exemple, ImageNet fournit 1000 images annotées pour chacune des 90 races de chien identifiées. Dans les exemples cités plus haut, Banar et De Tobel ont utilisé une base de données comprenant 400 radiographies panoramiques pour identifier 8 ou 9 stades.^{113,130} Ces deux auteurs ont mis au point un algorithme visant à déterminer chaque stade à partir de 44 images. Pour augmenter ce faible nombre d'images, les auteurs ont utilisé des procédés d'augmentation de la base de données d'entraînement avec des pré-traitements d'images comme les outils de rotation, d'inversion, de grossissement et de luminosité. Cependant les algorithmes opérant sur ImageNet utilisent aussi ces procédés d'amplification de la base de données. Cela pourra expliquer la différence de performance entre l'identification des stades de maturation dentaire et les races de chien. La tâche à réaliser peut également être plus ou moins accessible à un algorithme d'intelligence artificielle. La compréhension d'une tâche est un paramètre inaccessible à un algorithme alors qu'elle peut avoir une importance dans la prédiction.

Les études de Wallraff et Vila-Blanco utilisent chacune la radiographie panoramique entière en entrée d'algorithme avec une base de données de 14 000 et 1381 images respectivement pour une classe d'âge inférieure à 20 ans.^{123,129} Les réseaux neuronaux utilisés ont été le ResNet18 (technique de « transfert learning » par un réseau pré-entraîné sur ImageNet) et le DASNet (un réseau construit pour la tâche d'estimation de l'âge). Les performances mesurées avec l'erreur moyenne absolue ont été de 1.08 et 0.69 ans respectivement pour Wallraff et Vila-Blanco. Ces différences pourraient s'interpréter par le design d'une architecture du réseau neuronal spécifique dans l'étude de Vila Blanco et al. Une question se pose aussi sur l'utilisation de la totalité de la radiographie panoramique qui est le résultat d'une superposition de 27 structures anatomiques osseuses en plus des dents. La tâche de l'algorithme est de trouver des corrélations entre ces structures et l'âge de l'individu.

3.2.3.5. Coworking

L'une des problématiques le plus fréquemment soulevée avec l'utilisation des méthodes classiques d'estimation de l'âge porte sur l'échantillon de population d'origine pour la réalisation des tables de correspondance ou des équations de régression. Des études réalisées à partir d'autres populations ont montré une variabilité dans la maturation dentaire. Pour cette raison de nombreuses études de validation ou des méta-analyses sont réalisées encore aujourd'hui par des équipes scientifiques dans différents pays. L'objectif est de créer des bases de données représentatives pour chaque population au sens éthique ou géographique. Chaillot *et al.* ont regroupé en 2005 une base de données de 9 577 radiographies panoramiques de 8 pays différents.¹⁴⁵ Une collaboration a donc été réalisée entre différentes équipes pour tenter de fournir une méthode représentative de chaque population et de mieux appréhender la maturation dentaire humaine. Les méthodes d'apprentissage machine pourraient faciliter les travaux de coopération entre différents pays. Cela permettrait d'éviter l'inconvénient des difficultés de standardisation des résultats pour réaliser des méta-analyses. Différents algorithmes d'estimation de l'âge peuvent être testés simultanément sur chaque échantillon. Le processus étant plus ou moins automatisé, le temps de réalisation d'une évaluation n'est plus un frein.

En synthèse, l'utilisation des données nombreuses et variées est beaucoup plus souple grâce aux algorithmes d'intelligence artificielle. Ils laissent la possibilité de fusionner plusieurs types de prédicteurs dentaires et osseux. Ils peuvent aussi intégrer d'autres notions peu investiguées jusqu'ici comme le niveau social.

3.2.4. Usage de l'estimation de l'âge dentaire à partir d'orthopantomogrammes

L'utilisation des méthodes d'apprentissage machine a été discutée précédemment en comparaison aux méthodes de référence et par rapport à d'autres études utilisant l'intelligence artificielle. Les résultats sont concordants sur le fait que les méthodes d'apprentissage machine sont plus efficaces que les méthodes classiques d'estimation de l'âge chez les vivants. La question se pose alors de savoir quel est l'impact de ces résultats sur les pratiques actuelles de l'estimation de l'âge dentaire. L'objectif de cette partie est de prendre du recul par rapport à l'aspect technique et de repositionner les résultats dans leur contexte d'utilisation.

Les méthodes radiologiques évaluant le degré de maturation de l'organe dentaire chez les vivants sont utilisées dans plusieurs disciplines.⁸ Elles sont utilisées en orthodontie, en endocrinologie, en pédiatrie, en sciences médico-légales et dans le domaine sportif. Le sujet de recherche portant sur l'estimation de l'âge est un sujet populaire en anthropologie médico-légale du vivant.¹⁴⁶ Une recherche PubMed, basée sur les mots clés « dental age estimation » sur l'année 2021 renvoie à 75 études. Parmi ces résultats, 6 portaient sur des animaux. Dans ces 69 études restantes, 41 portent sur l'estimation de l'âge dans un contexte de médecine légale avec des applications possibles pour des demandeurs d'asile ou des réfugiés.^{5,102,105,107,108,110,111,133,146-182} En marge de cette dynamique de recherche, certains auteurs remettent en cause l'utilisation de ces méthodes.^{119,183}

Dans la spécialité d'orthodontie ou d'orthopédie dento-faciale, le degré de maturité dentaire d'un patient permet de planifier le traitement en s'appuyant la croissance résiduelle osseuse et dentaire. En médecine pédiatrique ou en endocrinologie, l'objectif est similaire et vise à évaluer la position d'un individu sur sa courbe de développement global.¹⁴⁹ Dans ce cadre d'utilisation, le principe est fondé sur la discordance entre l'âge chronologique et l'âge biologique.¹¹⁹ L'âge dentaire n'est utilisé que comme indicateur de maturation chez un individu, qui peut être en avance ou en retard par rapport à son âge chronologique. Il est intéressant de noter que dans les deux études de Demirjian de 1973 et 1976 présentant son système de stades de maturation, le terme « estimation de l'âge » n'est pas mentionné.^{4,23,49} Donc une partie des études partent du postulat que l'âge chronologique ne reflète pas l'âge biologique et nécessite l'évaluation du degré de maturation ou de sénescence. Et une autre partie des études se basent sur l'âge biologique pour connaître avec précision l'âge chronologique.

L'estimation de l'âge chronologique à partir de l'estimation ou de la classification de l'âge dentaire est utilisée pour des questions médico légales et pour des problématiques sportives. Ces dernières portent sur les catégories d'âges des athlètes se confrontent lors de compétitions officielles. Si Shen *et al.*

estiment qu'il est « vital » de déterminer l'âge des athlètes pour garantir l'équité des compétitions, l'âge exact d'un individu devient parfois essentiel au niveau médico-légal.¹⁸⁰ La liste suivante présente les différentes utilisations possibles de l'estimation de l'âge au niveau légal : l'âge d'un criminel ou d'une victime, l'âge des réfugiés, demandeurs d'asile, les procédures d'adoption, l'âge d'éligibilité pour la retraite, l'âge minimum légal du travail, la prostitution et l'âge d'entrée à l'armée.¹⁹ Toutes ces utilisations médico-légales, nécessitent de positionner un individu par rapport à un âge clé. Comme cela a été présenté dans la première partie, les conséquences d'une mauvaise estimation de l'âge peuvent être dramatiques pour des enfants indûment classés comme adulte.¹⁸³

En raison de ces risques, des lignes directrices ont été proposées par des organisations non gouvernementales mondiales pour l'estimation de l'âge.¹⁰ Les états gardent néanmoins la souveraineté dans l'application des procédures d'estimation de l'âge qui restent variables selon les pays.²⁵ L'UNICEF prévoit que dans les pratiques visant à obtenir le meilleur intérêt pour l'enfant, une approche holistique et multidisciplinaire doit être employée.¹⁴⁶ Ainsi, une approche non médicale et une approche médicale peuvent être appliquées sous certaines conditions.

Tout d'abord, l'estimation de l'âge ne doit être une technique de routine, elle ne devrait être envisagée qu'en dernier ressort.¹⁰ Ensuite, l'estimation de l'âge doit suivre un gradient des différentes méthodes possibles,¹⁴⁶ les moins contraignantes et invasives seront d'abord choisies, soit les méthodes non médicales, puis les méthodes médicales invasives seront utilisées si aucune des méthodes précédentes n'aura permis de répondre à la question.

S'il y a une unanimité de la littérature scientifique sur le fait qu'il est impossible de déterminer précisément l'âge chronologique d'une personne à partir d'une radiographie dentaire, il existe des débats sur la méthodologie.^{10,64} Par exemple, les entretiens psychologiques et sociaux sont considérés comme moins objectifs que les méthodes médicales.¹⁴⁶ D'autres auteurs pointent les difficultés d'utilisation des méthodes radiologiques dans l'estimation de l'âge dentaire.^{8,64,119} L'âge chronologique n'est pas représentatif de l'âge biologique, social ou psychologique et on peut se poser la question de savoir dans quel intérêt est réalisé le test : pour l'individu ou pour la société ?

D'un point de vue statistique, la présentation des résultats est problématique pour être utilisée dans une procédure médico-légale. Les résultats donnés dans les études sur l'estimation de l'âge sous forme de moyenne et d'un intervalle de confiance ne donnent pas une information sur une estimation ponctuelle pour un individu et il n'est pas possible de déterminer la probabilité d'être au-dessus ou en-dessous de l'âge clé.²⁸ Les troisièmes molaires sont les seules dents à présenter une maturation des

racines après 16 ans. L'utilisation de ces dents pour l'estimation de l'âge n'est pas recommandée pour plusieurs raisons. Tout d'abord la variabilité inter-individuelle du développement des racines est très variable et s'étend de 16 à 22 ans mais peut persister jusqu'à 30 ans.^{30,119} Ensuite l'agénésie ou l'absence de troisième molaire toucherait jusqu'à 25 % de la population.¹⁰ Enfin il existe un dimorphisme sexuel et des variations de développement significatives entre différentes ethnies.¹¹⁹ Les recherches se concentrent donc non pas sur l'estimation mais sur la classification d'un individu par rapport à l'âge de 18 ans.

Les auteurs présentent des méthodes permettant de discriminer des individus mineurs d'individus majeurs. Une méta-analyse de Santiago *et al.* en 2018 présente la méthode I3M < 0.8 comme ayant une spécificité moyenne de 0.93 (allant de 0.92 à 0.94). C'est-à-dire que les mineurs évalués sont classés comme mineurs dans 93 % des cas.¹⁸⁴ La sensibilité est plus faible avec 0.86 (allant de 0.84 à 0.87) mais est moins préjudiciable pour l'individu qui sera identifié à tort comme mineur. Si pour les auteurs, l'utilisation est fiable pour une question légale, la question se pose quand même pour la proportion de faux positifs, c'est-à-dire de mineurs considérés à tort comme majeurs. Cette proportion n'étant pas nulle, est-ce qu'elle répond aux principes du meilleur intérêt de l'enfant ? Laniado questionne aussi l'utilité d'irradier des individus sans justification médicale et propose de s'aligner sur l'Académie de pédiatrie Européenne, sur le collège royal britannique de pédiatrie, sur le collège royal britannique de radiologie, sur l'académie française de médecine, sur le comité national d'éthique français et sur la société nationale néerlandaise des médecins qui n'utilisent les examens irradiants qu'en dernier ressort et comme faisant partie d'un faisceau de preuves.¹¹⁹

4. Conclusion

Les méthodes d'apprentissage machine utilisées dans ce travail ont été entraînées de manière supervisée avec des prédicteurs correspondant aux stades de maturation de Demirjian des 7 dents mandibulaires gauche et des 4 troisièmes molaires ainsi que le sexe des individus. Les dix méthodes d'apprentissage machine testées pour estimer l'âge dentaire à partir de radiographies panoramiques sont plus précises que la méthode de Demirjian et celle de Willems dans les conditions d'utilisation de ces dernières.

L'intérêt de ces algorithmes est leur souplesse d'utilisation dans l'estimation de l'âge dentaire. Une fois entraîné, l'algorithme peut réaliser des prédictions sur des nouveaux cas et peut donc être utilisé dans une pratique médico-légale en fournissant des intervalles de confiance associés. Un autre avantage est la possibilité d'ajouter simplement des prédicteurs aux données d'entraînement. Dans ce travail, le sexe des individus a été ajouté de manière à apporter une information sur les différences de maturation des dents entre les femmes et les hommes. Les stades de maturation des troisièmes molaires ont également été ajoutés pour étendre les possibilités d'estimation de l'âge dentaire aux jeunes adultes. Dans chaque cas d'ajout d'un prédicteur ou pour un ajout de données annotées, le processus de détermination des hyperparamètres et des paramètres des algorithmes reste programmable et peut donc être automatisé dans une certaine mesure. Cela est intéressant pour un travail collectif avec différentes équipes de recherche pour réaliser des analyses sur des bases de données issues de différentes populations ou pour renouveler les études dans le temps afin d'investiguer l'effet de l'origine ethnique, de l'origine géographique, des variations de maturation dans le temps. Cette souplesse d'utilisation et la possibilité d'ajout des prédicteurs permet aussi d'envisager l'estimation de l'âge biologique dans un sens plus large en combinant des prédicteurs osseux, dentaires, biologiques mais aussi des caractéristiques sur l'environnement dans lequel les individus ont évolué.

Ce travail a aussi confirmé que l'ajout de la troisième molaire comme prédicteur permettait d'étendre la plage d'utilisation de l'estimation de l'âge dentaire mais au détriment de la précision. Les résultats obtenus ont également montré une précision comparable aux résultats obtenus par des techniques d'intelligence artificielle dites d'apprentissage profond. Ces dernières ont révolutionné les domaines du traitement d'image, du traitement du son, du traitement du langage naturel. Leur utilisation dans le domaine de la santé a permis d'aider et d'automatiser des tâches auprès du patient et du praticien. La sélection d'un algorithme d'apprentissage automatique est un processus d'essais et d'erreurs. C'est aussi un compromis entre les caractéristiques de vitesse d'apprentissage, de la précision des prédictions obtenues ou souhaitées et de l'interprétabilité des algorithmes.

L'estimation de l'âge dentaire à partir de radiographies panoramiques convient parfaitement au domaine d'utilisation des réseaux neuronaux profonds. La non-supériorité de ces techniques par rapport aux méthodes d'apprentissage machine testées dans ce travail, pourrait s'expliquer par la variabilité inter-individuelle de la maturation de l'âge dentaire. Comme évoqué plus haut, l'utilisation des données numériques massives pourrait permettre d'ajouter des prédicteurs portant sur les conditions de vie des individus et tester leur influence sur la maturation dentaire.

Si le nombre des publications sur l'estimation de l'âge dentaire chez des individus vivants témoigne d'un intérêt pour le sujet, il est intéressant de noter que le fondement de l'utilisation de cette technique trouve son intérêt dans des applications diamétralement opposées. D'un côté, l'utilisation pour des raisons cliniques, orthodontiques par exemple, évalue l'âge dentaire car ce dernier ne reflète pas l'âge chronologique. D'un autre côté, l'utilisation pour des raisons médico-légales vise à estimer le plus précisément possible l'âge chronologique à partir de l'âge dentaire. Au niveau médico-légal, les demandes d'estimation de l'âge des individus ont augmenté ces dernières années avec l'augmentation des flux migratoires à partir de pays dont les certificats de naissance ne sont pas systématiques ou en raison de la fuite de réfugiés ou de demandeurs d'asile. Les organisations internationales préconisent l'estimation de l'âge chronologique à partir d'examens médicaux et notamment radiologiques en dernier recours quand toutes les autres méthodes ont été épuisées. Dans ces rares cas, l'examen devrait être réalisé par une personne formée aux techniques d'estimation de l'âge de manière à fournir une réponse circonstanciée avec les marges d'erreur inhérentes à ces techniques. La présentation de nos résultats ne permet pas de fournir l'estimation d'un âge dentaire assorti d'un intervalle de confiance dans le cas d'un individu unique.

Les futurs travaux à entreprendre dans l'estimation de l'âge dentaire pourraient viser à améliorer l'interprétabilité des algorithmes, plus précisément pour les réseaux neuronaux profonds. Une coopération avec un partage des données entre différentes équipes permettrait de travailler sur des bases de données conséquentes. Cette option pourrait être mise en place facilement avec l'utilisation des algorithmes d'apprentissage machine. Enfin l'automatisation du processus d'estimation de l'âge dentaire à partir de radiographies panoramiques pourrait être poursuivi à trois niveaux : la segmentation automatique des zones d'intérêt, l'estimation du degré de maturité et l'estimation de l'âge dentaire.

Le transfert de compétences de l'utilisation de l'intelligence artificielle utilisée pour l'estimation de l'âge dentaire existe dans le domaine de l'odontologie clinique avec le développement d'algorithmes pour la détection de carie, l'identification de lésions parodontales et la génération d'un schéma dentaire à partir d'une radiographie panoramique dentaire en utilisant des outils d'apprentissage

machine similaires à ceux présents dans ce travail. Dans un sens plus large, l'intelligence artificielle permet également repousser les limites de l'enseignement pédagogique tel qu'il est connu actuellement en santé. Les épisodes de distanciation sociale, l'augmentation du nombre d'étudiants et les normes, l'obsolescence et les limites des lieux d'enseignement ont précipité un changement de paradigme dans l'enseignement. L'enseignement en distanciel est devenu la norme mais présente des désavantages avec une déconnexion, un manque d'interactivité par rapport au présentiel. L'intelligence artificielle avec le développement du métaverse serait une solution dans le futur. Il est étonnant d'imaginer une réalité virtuelle où les attributs des avatars n'auraient aucun lien avec l'âge chronologique de son utilisateur.

Bibliographie

1. Slimak L, Zanolli C, Higham T, et al. Modern human incursion into Neanderthal territories 54,000 years ago at Mandrin, France. *Sci Adv.* 2022;8(6):eabj9496. doi:10.1126/sciadv.abj9496
2. Weyrich LS, Duchene S, Soubrier J, et al. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nat Publ Gr.* 2017;544. doi:10.1038/nature21674
3. Nuzzolese E, Di Vella G. *Legal Background of Age Estimation for the Dead and the Living.* Elsevier Inc.; 2019. doi:10.1016/B978-0-12-814491-6.00002-9
4. Demirjian A, Goldstein H, Tanner JM. A NEW SYSTEM OF DENTAL AGE ASSESSMENT. *Hum Biol.* 1973;45(2):211-227.
5. Hostiuc S, Edison S-E, Diaconescu I, Negoii I, Isaila O-M. Accuracy of the Demirjian's method for assessing the age in children, from 1973 to 2020. A meta-analysis. *Leg Med (Tokyo).* 2021;52:101901. doi:10.1016/j.legalmed.2021.101901
6. Esan TA, Yengopal V, Schepartz LA. The Demirjian versus the Willems method for dental age estimation in different populations: A meta-analysis of published studies. *PLoS One.* 2017;12(11):1-23. doi:10.1371/journal.pone.0186682
7. Gaudiaut T. La création de données numériques va exploser. World Economic Forum. <https://fr.weforum.org/agenda/2019/04/la-creation-de-donnees-numeriques-va-exploser/>. Published 2019. Accessed January 22, 2022.
8. Ellingham S, Adserias-Garriga J. *Complexities and Considerations of Human Age Estimation.* Elsevier Inc.; 2019. doi:10.1016/B978-0-12-814491-6.00001-7
9. Sayn I. Le critère de l'âge dans les dispositifs juridiques. *Retraite et société.* 2001;34:131-149.
10. Smith T, Brownlees L. *Age Assessment Practices: A Literature Review & Annotated Bibliogr.* UNICEF. 3 UN Plaza, New York, NY 10017; 2011.
11. Brumberg HL, Dozor D, Golombek SG. History of the birth certificate: From inception to the future of electronic data. *J Perinatol.* 2012;32(6):407-411. doi:10.1038/jp.2012.3
12. Kuppelwieser VG. Towards the use of chronological age in research – A cautionary comment. *J Retail Consum Serv.* 2016;33:17-22. doi:10.1016/j.jretconser.2016.05.016

13. United-nations. The Sustainable Development Goals. United Nations Development Programme. <https://www.undp.org/sustainable-development-goals#peace-justice-and-strong-institutions>. Published 2015.
14. UNICEF. *Birth Registration Based on Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), Other National Surveys, Censuses and Vital Registration Systems, 2010-2019.*; 2020. <https://data.unicef.org/topic/child-protection/birth-registration/>.
15. United Nations Human Rights office of the High Commissioner. The Convention on the Rights of the Child. In: *Social Policy & Administration.* ; 1989.
16. Bagley AM. Roman children in the early empire: a distinct epidemiological and therapeutic category? 2017;(August).
17. Schmeling A, Black S. An Introduction to the History of Age Estimation in the Living. *Age Estim Living Pract Guid.* 2010:1-18. doi:10.1002/9780470669785.ch1
18. Saunders E. The Teeth a Test of Age, Considered with Reference to the Factory Children. *Am J Dent Sci.* 1847;7(4):330-375.
<http://www.ncbi.nlm.nih.gov/pubmed/30751250><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6080001>.
19. Schmeling A, Black S. An Introduction to the History of Age Estimation in the Living. In: *Age Estimation in the Living.* John Wiley & Sons, Ltd; 2010:1-18.
doi:<https://doi.org/10.1002/9780470669785.ch1>
20. Cunha E, Baccino E, Martrille L, et al. The problem of aging human remains and living individuals: A review. *Forensic Sci Int.* 2009;193(1-3):1-13. doi:10.1016/j.forsciint.2009.09.008
21. Metsäniitty M, Waltimo-Sirén J, Ranta H, Fieuws S, Thevissen P. Dental age estimation in Somali children and sub-adults combining permanent teeth and third molar development. *Int J Legal Med.* 2019. doi:10.1007/s00414-019-02053-w
22. Jylhävä J, Pedersen NL, Hägg S. Biological Age Predictors. *EBioMedicine.* 2017;21:29-36.
doi:10.1016/j.ebiom.2017.03.046
23. Liversidge HM. The assessment and interpretation of Demirjian, Goldstein and Tanner's dental maturity. *Ann Hum Biol.* 2012;39(5):412-431. doi:10.3109/03014460.2012.716080
24. Braga J, Heuze Y, Chabadel O, Sonan NK, Gueramy A. Non-adult dental age assessment:

- Correspondence analysis and linear regression versus Bayesian predictions. *Int J Legal Med.* 2005;119(5):260-274. doi:10.1007/s00414-004-0494-8
25. EASO. *Rapport de l'EASO Sur La Détermination de l'âge En Europe.*; 2013.
 26. Justice M de la, Santé M des solidarités et de la, L'intérieur ministère de, Territoriales ministère de la C des terriitoires et des relations avec les collectivités. *Guide de Bonnes Pratiques En Matiere d'évaluation de La Minorite et de l'isolement.*; :42.
 27. Schmeling A, Grundmann C, Fuhrmann A, et al. Criteria for age estimation in living individuals. *Int J Legal Med.* 2008;122(6):457-460. doi:10.1007/s00414-008-0254-2
 28. Sironi E, Vuille J, Morling N, Taroni F. On the Bayesian approach to forensic age estimation of living individuals. *Forensic Sci Int.* 2017;281:e24-e29. doi:10.1016/j.forsciint.2017.11.007
 29. Taylor J, Blenkin M. Age Evaluation and Odontology in the Living. *Age Estim Living Pract Guid.* 2010:176-201. doi:10.1002/9780470669785.ch10
 30. Taylor J, Blenkin M. Age Evaluation and Odontology in the Living. In: *Age Estimation in the Living.* John Wiley & Sons, Ltd; 2010:176-201.
doi:https://doi.org/10.1002/9780470669785.ch10
 31. Verma M, Verma N, Sharma R, Sharma A. Dental age estimation methods in adult dentitions: An overview. *J Forensic Dent Sci.* 2019;11(2):57-63. doi:10.4103/jfo.jfds_64_19
 32. 2020 EAR. 6.3 *Age Assessment of Minors.*(2020). <https://www.easo.europa.eu/asylum-report-2020/63-age-assessment-minors#ar229>.
 33. WHO. World Health Organisation child growth standards. <https://www.who.int/toolkits/child-growth-standards/standards>. Published 2022. Accessed October 3, 2022.
 34. Rosenbloom AL, Tanner JM. Misuse of Tanner puberty stages to estimate chronologic age. *Pediatrics.* 1998;102(6):1494. doi:10.1542/peds.102.6.1494
 35. AlQahtani S. *Dental Age Estimation in Fetal and Children.* Elsevier Inc.; 2019.
doi:10.1016/B978-0-12-814491-6.00007-8
 36. Tardivo D, Sastre J, Catherine JH, Leonetti G, Adalian P, Foti B. Age determination of adult individuals by three-dimensional modelling of canines. *Int J Legal Med.* 2014;128(1):161-169.
doi:10.1007/s00414-013-0863-2

37. Baumann P, Widek T, Merkens H, et al. Dental age estimation of living persons: Comparison of MRI with OPG. *Forensic Sci Int.* 2015;253:76-80. doi:10.1016/j.forsciint.2015.06.001
38. Widek T, Genet P, Merkens H, et al. Dental age estimation: The chronology of mineralization and eruption of male third molars with 3T MRI. *Forensic Sci Int.* 2019;297:228-235. doi:10.1016/j.forsciint.2019.02.019
39. Liversidge HM, Lyons F, Hector MP. The accuracy of three methods of age estimation using radiographic measurements of developing teeth. *Forensic Sci Int.* 2003;131(1):22-29. doi:10.1016/S0379-0738(02)00373-0
40. Schour I, Massler M. Studies In Tooth Development: The Growth Pattern Of Human Teeth. *J Am Dent Assoc.* 1940;27(11):1778-1793. doi:https://doi.org/10.14219/jada.archive.1940.0340
41. Schour I, Massler M. The development of the human dentition. *J Am Dent Assoc.* 1942;28:1153-1160. doi:10.1080/08869634.2016.1258982
42. Gleiser I, Hunt EE. The permanent mandibular first molar: Its calcification, eruption and decay. *Am J Phys Anthropol.* 1955;13(2):253-283. doi:10.1002/ajpa.1330130206
43. Maber M, Liversidge HM, Hector MP. Accuracy of age estimation of radiographic methods using developing teeth. *Forensic Sci Int.* 2006;159(1):68-73. doi:10.1016/j.forsciint.2006.02.019
44. Moorrees CFA, Fanning EA, Hunt EE. Age Variation of Formation Stages for Ten Permanent Teeth. *J Dent Res.* 1963;42(6):1490-1502. doi:10.1177/00220345630420062701
45. Wolanski N. A new method for the evaluation of tooth formation. *Acta Genet Stat Med.* 1966;16(2):186-197. doi:10.1159/000151963
46. Gustafson G, Koch G. Age estimation up to 16 years of age based on dental development. *OdontRev.* 1974;25(3):297-306.
47. Haavikko K. Tooth formation age estimated on a few selected teeth. A simple method for clinical use. *Proc Finnish Dent Soc.* 1974;70(1):15-19.
48. Anderson DL, Thompson GW, Popovich F. Age of attainment of mineralization stages of the permanent dentition. *J Forensic Sci.* 1976;21(1):191-200.
49. Demirjian A, Goldstein H. New systems for dental maturity based on seven and four teeth. *Ann Hum Biol.* 1976;3(5):411-421. doi:10.1080/03014467600001671

50. Carels CE, Kuijpers-Jagtman AM, van der Linden FP, Van't Hof MA. Age reference charts of tooth length in Dutch children. *J Biol Buccale*. 1991;19(4):297-303.
51. Mörnstad H, Staaf V, Welander U. Age estimation with the aid of tooth development: a new method based on objective measurements. *Eur J Oral Sci*. 1994;102(3):137-143.
doi:10.1111/j.1600-0722.1994.tb01169.x
52. Liversidge HM, Molleson TI. Developing permanent tooth length as an estimate of age. *J Forensic Sci*. 1999;44(5):917-920.
53. Teivens A, Mörnstad H. A comparison between dental maturity rate in the Swedish and Korean populations using a modified Demirjian method. *J Forensic Odontostomatol*. 2001;19(2):31-35.
54. Willems G, Ph D, Olmen A Van, et al. Dental Age Estimation in Belgian Children: Demirjian's Technique Revisited. *J Forensic Sci*. 2001;46(4):893-895.
55. Blenkin MRB, Evans W. Age estimation from the teeth using a modified Demirjian system. *J Forensic Sci*. 2010;55(6):1504-1508. doi:10.1111/j.1556-4029.2010.01491.x
56. Cameriere R, Ferrante L, Cingolani M. Age estimation in children by measurement of open apices in teeth. *Int J Legal Med*. 2006;120(1):49-53. doi:10.1007/s00414-005-0047-9
57. AlQahtani SJ, Hector MP, Liversidge HM. Brief communication: The London atlas of human tooth development and eruption. *Am J Phys Anthropol*. 2010;142(3):481-490.
doi:10.1002/ajpa.21258
58. Adams DM, Ralston CE, Sussman RA, Heim K, Bethard JD. Impact of population-specific dental development on age estimation using dental atlases. *Am J Phys Anthropol*. 2018;168(1):190-199. doi:10.1002/ajpa.23735
59. Hostiuc S, Diaconescu I, Rusu MC, Negoii I. Age Estimation Using the Cameriere Methods of Open Apices: A Meta-Analysis. *Healthc (Basel, Switzerland)*. 2021;9(2).
doi:10.3390/healthcare9020237
60. Ribier L, Saint-Martin P, Seignier M, Paré A, Brunereau L, Rérolle C. Cameriere's third molar maturity index in assessing age of majority: a study of a French sample. *Int J Legal Med*. 2019;783-792. doi:10.1007/s00414-019-02123-z
61. Cameriere R, Cunha E, Sassaroli E, Nuzzolese E, Ferrante L. Age estimation by pulp/tooth area

- ratio in canines: Study of a Portuguese sample to test Cameriere's method. *Forensic Sci Int.* 2009;193(1-3):128.e1-128.e6. doi:10.1016/j.forsciint.2009.09.011
62. Tardivo D, Sastre J, Ruquet M, et al. Three-dimensional modeling of the various volumes of canines to determine age and sex: A preliminary study. *J Forensic Sci.* 2011;56(3):766-770. doi:10.1111/j.1556-4029.2011.01720.x
 63. Cameriere R, Ferrante L, Cingolani M. Variations in pulp/tooth area ratio as an indicator of age: a preliminary study. *J Forensic Sci.* 2004;49(2):317-319.
 64. Sironi E, Gallidabino M, Weyermann C, Taroni F. Probabilistic graphical models to deal with age estimation of living persons. *Int J Legal Med.* 2016;130(2):475-488. doi:10.1007/s00414-015-1173-7
 65. Willems G. A review of the most commonly used dental age estimation techniques. *J Forensic Odontostomatol.* 2001;19(1):9-17.
 66. Wang J, Ji F, Zhai Y, Park H, Tao J. Is Willems method universal for age estimation: A systematic review and meta-analysis. *J Forensic Leg Med.* 2017;52:130-136. doi:10.1016/j.jflm.2017.09.003
 67. Hostiuc S, Edison S-E, Diaconescu I, Negoii I, Isaila O-M. Accuracy of the Demirjian's method for assessing the age in children, from 1973 to 2020. A meta-analysis. *Leg Med (Tokyo).* 2021;52:101901. doi:10.1016/j.legalmed.2021.101901
 68. Jayaraman J, Wong HM, King NM, Roberts GJ. The French–Canadian data set of Demirjian for dental age estimation: A systematic review and meta-analysis. *J Forensic Leg Med.* 2013;20(5):373-381. doi:10.1016/j.jflm.2013.03.015
 69. Yan J, Lou X, Xie L, Yu D, Shen G, Wang Y. Assessment of dental age of children aged 3.5 to 16.9 years using Demirjian's method: A meta-analysis based on 26 studies. *PLoS One.* 2013;8(12):1-10. doi:10.1371/journal.pone.0084672
 70. Sehrawat JS, Singh M. Willems method of dental age estimation in children: A systematic review and meta-analysis. *J Forensic Leg Med.* 2017;52:122-129. doi:10.1016/j.jflm.2017.08.017
 71. Liversidge HM, Herdeg B, Rosing FW. Dental Age Estimation of Non-Adults. A Review of Methods and Principles. In: Kurt W, Rosing FW, Teschler-Nicola M, eds. *Dental Anthropology: Fundamentals, Limits and Prospects.* SpringerWi. Vienna: Springer Vienna; 1998:419-442.

doi:10.1007/978-3-7091-7496-8

72. Thevissen PW, Fieuws S, Willems G. Human dental age estimation using third molar developmental stages: Does a Bayesian approach outperform regression models to discriminate between juveniles and adults? *Int J Legal Med.* 2010;124(1):35-42. doi:10.1007/s00414-009-0329-8
73. Tao J, Wang J, Wang A, et al. *Dental Age Estimation : A Machine Learning Perspective.* Vol 7. Springer International Publishing; 2020. doi:10.1007/978-3-030-14118-9
74. Sironi E, Gallidabino M, Weyermann C, et al. Bayesian networks of age estimation and classification based on dental evidence: A study on the third molar mineralization. *J Forensic Leg Med.* 2018;55(2):23-32. doi:10.1016/j.jflm.2018.02.005
75. Star H, Thevissen P, Jacobs R, Fieuws S, Solheim T, Willems G. Human Dental Age Estimation by Calculation of Pulp-Tooth Volume Ratios Yielded on Clinically Acquired Cone Beam Computed Tomography Images of Monoradicular Teeth. *J Forensic Sci.* 2011;56(SUPPL. 1). doi:10.1111/j.1556-4029.2010.01633.x
76. Štepanovský M, Ibrová A, Buk Z, Velemínská J. Novel age estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods. *Forensic Sci Int.* 2017;279(10):1779-1790. doi:10.1002/sim.6448
77. Howard J. Artificial intelligence: Implications for the future of work. *Am J Ind Med.* 2019;62(11):917-926. doi:10.1002/AJIM.23037
78. Meyer B. John McCarthy. *Communications of the ACM.*
79. Le Cun Y. Machines apprenantes simples. In: *Quand l Machine Apprend. La Révolution Des Neurones Artificiels et de l'apprentissage Prfond.* Odile Jaco. Paris; 2019:81-120.
80. Chang AC. Basic Concepts of Artificial Intelligence. In: Elsevier, ed. *Intelligence-Based Medicine.* London: Academic Press is an imprint of Elsevier; 2020:47-75.
81. Quinlan JR. Combining Instance-Based and Model-Based Learning. *Mach Learn Proc 1993.* 1993;93:236-243. doi:10.1016/b978-1-55860-307-3.50037-x
82. Chang AC. Machine and Deep Learning. In: Elsevier, ed. *Intelligence-Based Medicine.* London: Academic Press is an imprint of Elsevier; 2020:150-276.
83. Claesen M, De Moor B. Hyperparameter Search in Machine Learning. *MIC 2015 XI*

Metaheuristics Int Conf. February 2015:10-14. <http://arxiv.org/abs/1502.02127>.

84. Chicco D. Ten quick tips for machine learning in computational biology. *BioData Min*. 2017;10(1):1-17. doi:10.1186/s13040-017-0155-3
85. Hegde S, Patodia A, Shah K, Dixit U. The applicability of the Demirjian, Willems and Chaillet standards to age estimation of 5-15 year old Indian children. *J Forensic Odontostomatol*. 2019;37(1):40-50.
86. Luts J, Ojeda F, Van de Plas Raf R, De Moor B, Van Huffel S, Suykens JAK. A tutorial on support vector machine-based methods for classification problems in chemometrics. *Anal Chim Acta*. 2010;665(2):129-145. doi:10.1016/j.aca.2010.03.030
87. Van Belle V, Van Calster B, Van Huffel S, Suykens JAK, Lisboa P. Explaining Support Vector Machines: A Color Based Nomogram. *PLoS One*. 2016;11(10):e0164568. doi:10.1371/journal.pone.0164568
88. Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD. An introduction to decision tree modeling. *J Chemom*. 2004;18(6):275-285. doi:10.1002/cem.873
89. Yang Y, Yang Y. Hybrid prediction method for wind speed combining ensemble empirical mode decomposition and bayesian ridge regression. *IEEE Access*. 2020;8:71206-71218. doi:10.1109/ACCESS.2020.2984020
90. Ren Y, Zhang L, Suganthan PN. Ensemble Classification and Regression-Recent Developments, Applications and Future Directions [Review Article]. *IEEE Comput Intell Mag*. 2016;11(1):41-53. doi:10.1109/MCI.2015.2471235
91. Opitz D, Maclin R. Popular Ensemble Methods: An Empirical Study. *J Artif Intell Res*. 1999;11:169-198. doi:10.1613/jair.614
92. Wang G, Hao J, Ma J, Jiang H. A comparative assessment of ensemble learning for credit scoring. *Expert Syst Appl*. 2011;38(1):223-230. doi:10.1016/J.ESWA.2010.06.048
93. Chang AC. History of Artificial Intelligence. In: Elsevier, ed. *Intelligence-Based Medicine*. London: Academic Press is an imprint of Elsevier; 2020:76-83.
94. Turing AM. Computing Machinery and Intelligence. In: *Parsing the Turing Test*. Dordrecht: Springer Netherlands; 2009:23-65. doi:10.1007/978-1-4020-6710-5_3
95. Le Cun Y. Brève histoire de l'IA... et de ma carrière. In: *Quand l'Machine Apprend*. La

- Révolution Des Neurones Artificiels et de l'apprentissage Prfond*. Odile Jaco. Paris; 2019:13-79.
96. Le Cun Y. Réseaux profonds et rétropropagation. In: *Quand I Machine Apprend. La Révolution Des Neurones Artificiels et de l'apprentissage Prfond*. Odile Jaco. paris; 2019:166-193.
 97. Cardon D, Cointet JP, Mazières A. *La Revanche Des Neurones: L'invention Des Machines Inductives et La Controverse de l'intelligence Artificielle*. Vol 211.; 2018.
doi:10.3917/res.211.0173
 98. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int J Comput Vis*. 2020;128(2):336-359. doi:10.1007/s11263-019-01228-7
 99. Le Cun Y. les réseaux convolutifs, piliers de l'IA. In: *Quand I Machine Apprend. La Révolution Des Neurones Artificiels et de l'apprentissage Prfond*. Odile Jaco. Paris; 2019:194, 225.
 100. Miki Y, Muramatsu C, Hayashi T, et al. Classification of teeth in cone-beam CT using deep convolutional neural network. *Comput Biol Med*. 2017;80(September 2016):24-29.
doi:10.1016/j.compbiomed.2016.11.003
 101. Sykes L, Bhayat A, Bernitz H. The Effects of the Refugee Crisis on Age Estimation Analysis over the Past 10 Years: A 16-Country Survey. 2017. doi:10.3390/ijerph14060630
 102. Franco A, de Oliveira MN, Campos Vidigal MT, Blumenberg C, Pinheiro AA, Paranhos LR. Assessment of dental age estimation methods applied to Brazilian children: a systematic review and meta-analysis. *Dentomaxillofacial Radiol*. 2021;50(2):20200128.
doi:10.1259/dmfr.20200128
 103. De Donno A, Angrisani C, Mele F, Introna F, Santoro V. Dental age estimation: Demirjian's versus the other methods in different populations. A literature review. *Med Sci Law*. 2021;61(1_suppl):125-129. doi:10.1177/0025802420934253
 104. Hostiuc S, Edison S-E, Diaconescu I, Negoii I, Isailă O-M. Accuracy of the Demirjian's method for assessing the age in children, from 1973 to 2020. A meta-analysis. *Leg Med*. 2021;52:101901. doi:10.1016/j.legalmed.2021.101901
 105. Diaconescu I, Isailă O-M, Hostiuc S. Accuracy of the Chaillet's Method for Assessing the Age in Subadults. A Meta-Analysis. *Curr Heal Sci J*. 2021;47(2):196-203. doi:10.12865/CHSJ.47.02.08

106. Merdietio Boedi R, Shepherd S, Mânica S, Franco A. CBCT in dental age estimation: A systematic review and meta analysis. *Dentomaxillofac Radiol.* January 2022.
doi:10.1259/DMFR.20210335
107. Molina A, Bravo M, Fonseca GM, Márquez-Grant N, Martín-de-Las-Heras S. Dental age estimation based on pulp chamber/crown volume ratio measured on CBCT images in a Spanish population. *Int J Legal Med.* 2021;135(1):359-364. doi:10.1007/s00414-020-02377-y
108. Timme M, Borkert J, Nagelmann N, Streeter A, Karch A, Schmeling A. Age-dependent decrease in dental pulp cavity volume as a feature for age assessment: a comparative in vitro study using 9.4-T UTE-MRI and CBCT 3D imaging. *Int J Legal Med.* 2021;135(4):1599-1609.
doi:10.1007/s00414-021-02603-1
109. Vossoughi M, Movahhedian N, Ghafoori A. The impact of age mimicry bias on the accuracy of methods for age estimation based on Kvaal's pulp/tooth ratios: a bootstrap study. *Int J Legal Med.* July 2021. doi:10.1007/s00414-021-02651-7
110. Pires AC, Vargas de Sousa Santos RF, Pereira CP. Dental age assessment by the pulp/tooth area proportion in cone beam computed tomography: is medico-legal application for age estimation reliable? *J Forensic Odontostomatol.* 2021;2(39):2-14.
111. Al-Omoush SA, Alhadidi A, Al-Kayed A, Saoud H, Alsoleihat F. Do upper third molars provide more accurate age estimation in the adult based on the pulp-to-tooth ratio than lower third molars? A cone-beam CT study. *Saudi Dent J.* 2021;33(7):702-706.
doi:10.1016/j.sdentj.2020.04.006
112. Kim S, Lee Y-H, Noh Y-K, Park FC, Auh Q-S. Age-group determination of living individuals using first molar images based on artificial intelligence. *Sci Rep.* 2021;11(1):1073.
doi:10.1038/s41598-020-80182-8
113. Banar N, Bertels J, Laurent F, et al. Towards fully automated third molar development staging in panoramic radiographs. *Int J Legal Med.* 2020;134(5):1831-1841. doi:10.1007/s00414-020-02283-3
114. Zheng Q, Ge Z, Du H, Li G. Age estimation based on 3D pulp chamber segmentation of first molars from cone-beam-computed tomography by integrated deep learning and level set. *Int J Legal Med.* 2021;135(1):365-373. doi:10.1007/s00414-020-02459-x
115. Shen S, Liu Z, Wang J, Fan L, Ji F, Tao J. Machine learning assisted Cameriere method for

- dental age estimation. *BMC Oral Health*. 2021;21(1):641. doi:10.1186/S12903-021-01996-0
116. Guo Y-C, Han M, Chi Y, et al. Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images. *Int J Legal Med*. 2021;135(4):1589-1597. doi:10.1007/s00414-021-02542-x
117. Zaborowicz K, Biedziak B, Olszewska A, Zaborowicz M. Tooth and bone parameters in the assessment of the chronological age of children and adolescents using neural modelling methods. *Sensors*. 2021;21(18):1-18. doi:10.3390/s21186008
118. Khanagar SB, Vishwanathaiah S, Naik S, et al. Application and performance of artificial intelligence technology in forensic odontology – A systematic review. *Leg Med*. 2021;48:101826. doi:10.1016/J.LEGALMED.2020.101826
119. Laniado N. A Call to End the Use of Dental Radiographs for Age Estimation. *Am J Public Health*. 2021;111(5):799-800. doi:10.2105/AJPH.2021.306223
120. Franco A, de Oliveira MN, Campos Vidigal MT, Blumenberg C, Pinheiro AA, Paranhos LR. Assessment of dental age estimation methods applied to Brazilian children: a systematic review and meta-analysis. *Dentomaxillofac Radiol*. 2021;50(2):20200128. doi:10.1259/dmfr.20200128
121. Rodriguez B, Cucina A, Vega E, Hoyos R, Cardoso HFV. A critical test of twelve methods for estimating age using radiographic staging of developing teeth on a sample of 6- to 15-year-old children from Mérida, Yucatán (México). *Int J Legal Med*. 2021;135(6):2457-2467. doi:10.1007/S00414-021-02689-7
122. Brotons A, Remy F, Foti B, Philip-Alliez C. Concordances and correlations between chronological, dental and bone ages: A retrospective study in French individuals. *Forensic Sci Int*. 2021;331:111150. doi:10.1016/J.FORSCIINT.2021.111150
123. Wallraff S, Vesal S, Syben C, Lutz R, Maier A. Age Estimation on Panoramic Dental X-ray Images using Deep Learning. *Inform aktuell*. 2021:186-191. doi:10.1007/978-3-658-33198-6_46
124. Štern D, Payer C, Giuliani N, Urschler M. Automatic Age Estimation and Majority Age Classification from Multi-Factorial MRI Data. *IEEE J Biomed Heal Informatics*. 2019;23(4):1392-1403. doi:10.1109/JBHI.2018.2869606
125. Halabi SS, Prevedello LM, Kalpathy-Cramer J, et al. The RSNA Pediatric Bone Age Machine

- Learning Challenge. *Radiology*. 2019;290(2):498-503. doi:10.1148/radiol.2018180736
126. Dallora AL, Anderberg P, Kvist O, Mendes E, Ruiz SD, Berglund JS. Bone age assessment with various machine learning techniques: A systematic literature review and meta-analysis. *PLoS One*. 2019;14(7):1-22. doi:10.1371/journal.pone.0220242
127. Galibourg A, Dumoncel J, Telmon N, Calvet A, Michetti J, Maret D. Assessment of automatic segmentation of teeth using a watershed-based method. *Dentomaxillofacial Radiol*. 2018;47(1):20170220. doi:10.1259/dmfr.20170220
128. Hemalatha B, Rajkumar N. A versatile approach for dental age estimation using fuzzy neural network with teaching learning - based optimization classification. *Multimed Tools Appl*. 2020;79(5-6):3645-3665. doi:10.1007/s11042-018-6434-2
129. Vila-Blanco N, Carreira MJ, Varas-Quintana P, Balsa-Castro C, Tomas I. Deep Neural Networks for Chronological Age Estimation from OPG Images. *IEEE Trans Med Imaging*. 2020;39(7):2374-2384. doi:10.1109/TMI.2020.2968765
130. De Tobel J, Radesh P, Vandermeulen D, Thevissen PW. An automated technique to stage lower third molar development on panoramic radiographs for age estimation: A pilot study. *J Forensic Odontostomatol*. 2017;35(2):42-54.
131. Ueda D, Yamamoto A, Shimazaki A, et al. Artificial intelligence-supported lung cancer detection by multi-institutional readers with multi-vendor chest radiographs: a retrospective clinical validation study. *BMC Cancer*. 2021;21(1):1120. doi:10.1186/s12885-021-08847-9
132. Gardner GG, Keating D, Williamson TH, Elliott AT. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *Br J Ophthalmol*. 1996;80(11):940-944. doi:10.1136/bjo.80.11.940
133. Galibourg A, Cussat-Blanc S, Dumoncel J, Telmon N, Monsarrat P, Maret D. Comparison of different machine learning approaches to predict dental age using Demirjian's staging approach. *Int J Legal Med*. 2021;135(2):665-675. doi:10.1007/s00414-020-02489-5
134. Grzybowski A, Brona P, Lim G, et al. Artificial intelligence for diabetic retinopathy screening: a review. *Eye*. 2020;34(3):451-460. doi:10.1038/s41433-019-0566-0
135. Sayres R, Taly A, Rahimy E, et al. Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy. *Ophthalmology*. 2019;126(4):552-564. doi:10.1016/j.ophtha.2018.11.016

136. Selvaraju RR, Das A, Vedantam R, Cogswell M, Parikh D, Batra D. Grad-CAM: Why did you say that? 2016:1-4. <http://arxiv.org/abs/1611.07450>.
137. Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. *Proc - 2018 IEEE Winter Conf Appl Comput Vision, WACV 2018*. 2018;2018-Janua:839-847.
doi:10.1109/WACV.2018.00097
138. Zhou B, Khosla A, Lapedriza À, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. *CoRR*. 2015;abs/1512.0. <http://arxiv.org/abs/1512.04150>.
139. Chow AR. What Are NFTs and Why They Are Shaking Up the Art World? Time. <https://time.com/5947720/nft-art/>. Published 2021. Accessed March 29, 2021.
140. Chang AC. Health-care Data and Databases. In: Elsevier, ed. *Intelligence-Based Medicine*. London: Academic Press is an imprint of Elsevier; 2020:117-149.
141. Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data. *IEEE Intell Syst*. 2009;24(2):8-12. doi:10.1109/MIS.2009.36
142. Yang K, Qinami K, Fei-Fei L, Deng J, Russakovsky O. Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ; 2020:547-558.
143. Olze A, Reisinger W, Geserick G, Schmeling A. Age estimation of unaccompanied minors. Part II. Dental aspects. *Forensic Sci Int*. 2006;159 Suppl:S65-7. doi:10.1016/j.forsciint.2006.02.018
144. Zaborowicz K, Biedziak B, Olszewska A, Zaborowicz M, Ling S. Tooth and Bone Parameters in the Assessment of the Chronological Age of Children and Adolescents Using Neural Modelling Methods. 2021. doi:10.3390/s21186008
145. Chaillet N, Nyström M, Demirjian A. Comparison of Dental Maturity in Children of Different Ethnic Origins: International Maturity Curves for Clinicians. *J Forensic Sci*. 2005;50(5):1-11. doi:10.1520/jfs2005020
146. De Micco F, Martino F, Campobasso C Pietro. Ethical issues in age assessment by the third molar development. *Aust J Forensic Sci*. 2020;00(00):1-12. doi:10.1080/00450618.2020.1789220
147. Pan J, Shen C, Yang Z, et al. A modified dental age assessment method for 5- to 16-year-old

- eastern Chinese children. *Clin Oral Investig*. 2021;25(6):3463-3474. doi:10.1007/s00784-020-03668-9
148. Palmela Pereira C, Rodrigues A, Santos A, Salvado F, Santos R, Cameriere R. Cut-off for the legal ages in the Portuguese Population by Third Maturity Index: Measures of Accuracy. *Arch Oral Biol*. 2021;125:105089. doi:10.1016/j.archoralbio.2021.105089
149. Brotons A, Remy F, Foti B, Philip-Alliez C. Concordances and correlations between chronological, dental and bone ages: A retrospective study in French individuals. *Forensic Sci Int*. 2022;331:111150. doi:10.1016/j.forsciint.2021.111150
150. Gaêta-Araujo H, Oliveira-Santos N, Nascimento EHL, et al. A new model of classification of third molars development and its correlation with chronological age in a Brazilian subpopulation. *Int J Legal Med*. 2021;135(2):639-648. doi:10.1007/s00414-020-02401-1
151. Guo Y-C, Han M, Chi Y, et al. Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images. *Int J Legal Med*. 2021;135(4):1589-1597. doi:10.1007/s00414-021-02542-x
152. Sezer B, Çarıkçioğlu B. Accuracy of the London Atlas, Haavikko's Method and Cameriere's European Formula of dental age estimation in Turkish children. *Leg Med (Tokyo)*. 2022;54:101991. doi:10.1016/j.legalmed.2021.101991
153. Shen S, Liu Z, Wang J, Fan L, Ji F, Tao J. Machine learning assisted Cameriere method for dental age estimation. *BMC Oral Health*. 2021;21(1):641. doi:10.1186/s12903-021-01996-0
154. Shi L, Zhou Y, Lu T, et al. Dental age estimation of Tibetan children and adolescents: Comparison of Demirjian, Willems methods and a newly modified Demirjian method. *Leg Med (Tokyo)*. 2022;55:102013. doi:10.1016/j.legalmed.2022.102013
155. Kanchan T, Chugh V, Chugh A, et al. Age estimation using third molar maturation based on Demirjian's criteria. *Leg Med (Tokyo)*. 2021;53:101959. doi:10.1016/j.legalmed.2021.101959
156. Ramli UIS, Muhd US, Mohd Yusof MYP. Accuracy of Kvaal's radiographic and translucent dentinal root techniques of extracted teeth in Malay adults for dental age estimation. *J Forensic Odontostomatol*. 2021;2(39):38-44.
157. De Micco F, Martino F, Velandia Palacio LA, Cingolani M, Campobasso C Pietro. Third molar maturity index and legal age in different ethnic populations: Accuracy of Cameriere's method. *Med Sci Law*. 2021;61(1_suppl):105-112. doi:10.1177/0025802419900686

158. Cadenas de Llano-Pérula M, Kihara E, Thevissen P, et al. Validating dental age estimation in Kenyan black children and adolescents using the Willems method. *Med Sci Law*. 2021;61(3):180-185. doi:10.1177/0025802420977379
159. Naik S, Rajadhyaksha N, Ghule KD. Evaluating the accuracy of two different age estimation methods in growing children in Navi Mumbai using Orthopantomograph: A radiographic study. *J Indian Soc Pedod Prev Dent*. 2021;39(4):363-368. doi:10.4103/jisppd.jisppd_227_21
160. Kim S, Lee Y-H, Noh Y-K, Park FC, Auh Q-S. Age-group determination of living individuals using first molar images based on artificial intelligence. *Sci Rep*. 2021;11(1):1073. doi:10.1038/s41598-020-80182-8
161. Hostiuc S, Diaconescu I, Rusu MC, Negoii I. Age estimation using the cameriere methods of open apices: A meta-analysis. *Healthc*. 2021;9(2):1-14. doi:10.3390/healthcare9020237
162. Wang J, Wang M, Shen S, et al. Testing the nonlinear equations for dental age evaluation in a population of eastern China. *Leg Med (Tokyo)*. 2021;48:101793. doi:10.1016/j.legalmed.2020.101793
163. Thomas LA, Thomas LR, Balla SB, et al. Above or below 14 years? An orthopantomographic study based on chronological course of eruption of mandibular premolars and second molars in a sample of south Indian children. *Leg Med (Tokyo)*. 2021;48:101814. doi:10.1016/j.legalmed.2020.101814
164. Kanchan T, Chugh V, Chugh A, et al. Age estimation using third molar maturation based on Demirjian's criteria. *Leg Med*. 2021;53(February):101959. doi:10.1016/j.legalmed.2021.101959
165. Matteussi GT, Jacometti V, Franco A, da Silva RHA. Age estimation in humans through the analysis of aspartic acid racemization from teeth: A scoping review of methods, outcomes, and open research questions. *Forensic Sci Int*. 2022;331:111154. doi:10.1016/j.forsciint.2021.111154
166. Alqerban A, Alrashed M, Alaskar Z, Alqahtani K. Age estimation based on Willems method versus country specific model in Saudi Arabia children and adolescents. *BMC Oral Health*. 2021;21(1):341. doi:10.1186/s12903-021-01707-9
167. Putul M, Konwar R, Dutta M, et al. Assessment of Age at the Stages of the Eruption of Third Molar Teeth among the People of North-Eastern India. *Biomed Res Int*. 2021;2021:9714121.

doi:10.1155/2021/9714121

168. Rodriguez B, Cucina A, Vega E, Hoyos R, Cardoso HF V. A critical test of twelve methods for estimating age using radiographic staging of developing teeth on a sample of 6- to 15-year-old children from Mérida, Yucatán (México). *Int J Legal Med.* 2021;135(6):2457-2467. doi:10.1007/s00414-021-02689-7
169. Angelakopoulos N, Galić I, Balla SB, et al. Comparison of the third molar maturity index (I(3M)) between left and right lower third molars to assess the age of majority: a multi-ethnic study sample. *Int J Legal Med.* 2021;135(6):2423-2436. doi:10.1007/s00414-021-02656-2
170. Augusto D, Pereira CP, Rodrigues A, Cameriere R, Salvado F, Santos R. Dental Age Assessment by I(2M) and I(3M): Portuguese Legal Age Thresholds of 12 and 14 Year Olds. *Acta Stomatol Croat.* 2021;55(1):45-55. doi:10.15644/asc55/1/6
171. Andrews SE, Roberts G, Set P, Warburton F, Gilbert FJ. Third molar development in a London population of White British and Black British or other Black ethnicity. *J Forensic Sci.* 2022;67(1):229-242. doi:10.1111/1556-4029.14928
172. Zirk M, Zoeller JE, Lentzen M-P, Bergeest L, Buller J, Zinser M. Comparison of two established 2D staging techniques to their appliance in 3D cone beam computer-tomography for dental age estimation. *Sci Rep.* 2021;11(1):9024. doi:10.1038/s41598-021-88379-1
173. Pyata JR, Kandukuri BA, Gangavarapu U, et al. Accuracy of four dental age estimation methods in determining the legal age threshold of 18 years among South Indian adolescents and young. *J Forensic Odontostomatol.* 2021;3(39):2-15.
174. Suvarna M, Lingam S, Balla SB, et al. Measurement of the open apices of mandibular first and second premolars to test the chronological age over 14 years: Study on a sample of south Indian children. *Leg Med (Tokyo).* 2021;49:101835. doi:10.1016/j.legalmed.2020.101835
175. Shivakumar B, Bankur R, Chinna SK, Niveditha B, Jain V, Sundar D. Application of Willems Method as an Age Estimator in Early Adolescents of Bangalore Population. *J Pharm Bioallied Sci.* 2021;13(Suppl 1):S595-S597. doi:10.4103/jpbs.JPBS_693_20
176. Zaborowicz M, Zaborowicz K, Biedziak B, Garbowski T. Deep Learning Neural Modelling as a Precise Method in the Assessment of the Chronological Age of Children and Adolescents Using Tooth and Bone Parameters. *Sensors (Basel).* 2022;22(2). doi:10.3390/s22020637
177. Karimi A, Qudeimat MA, Lucas VS, Roberts G. Dental age estimation: Development and

- validation of a reference data set for Kuwaiti children, adolescents, and young adults. *Arch Oral Biol.* 2021;127:105130. doi:10.1016/j.archoralbio.2021.105130
178. Suvarna M, Lingam S, Balla SB, et al. Measurement of the open apices of mandibular first and second premolars to test the chronological age over 14 years: Study on a sample of south Indian children. *Leg Med.* 2021;49(December 2020):101835. doi:10.1016/j.legalmed.2020.101835
179. Sartori V, Franco A, Linden M-S, et al. Testing international techniques for the radiographic assessment of third molar maturation. *J Clin Exp Dent.* 2021;13(12):e1182-e1188. doi:10.4317/jced.58916
180. Shen S, Guo Y, Wang M, et al. A quick method of determining the age of 8 years old: Based on the first premolars on Eastern Chinese population. *Leg Med.* 2021;53(July):101950. doi:10.1016/j.legalmed.2021.101950
181. Matteussi GT, Jacometti V, Franco A, da Silva RHA. Age estimation in humans through the analysis of aspartic acid racemization from teeth: A scoping review of methods, outcomes, and open research questions. *Forensic Sci Int.* 2022;331:111154. doi:10.1016/j.forsciint.2021.111154
182. Koç A, Özlek E, Öner Talmaç AG. Accuracy of the London atlas, Willems, and Nolla methods for dental age estimation: a cross-sectional study on Eastern Turkish children. *Clin Oral Investig.* 2021;25(8):4833-4840. doi:10.1007/s00784-021-03788-w
183. Crawley H. *When Is a Child Not a Child? Asylum, Age Disputes and the Process.* London: Immigration Law Practitioners' Association (ILPA); 2007.
184. Santiago BM, Almeida L, Cavalcanti YW, Magno MB, Maia LC. Accuracy of the third molar maturity index in assessing the legal age of 18 years: a systematic review and meta-analysis. *Int J Legal Med.* 2018;132(4):1167-1184. doi:10.1007/s00414-017-1766-4