UNIVERSITÉ CLAUDE BERNARD-LYON 1

U.F.R D'ODONTOLOGIE

Année 2024 Thèse n°2024 LYO1D 036

THÈSE
POUR LE DIPLOME D'ÉTAT DE DOCTEUR EN CHIRURGIE DENTAIRE

Présentée et soutenue publiquement le 17/06/2024

Par

Bérengère BAYON
Née le 13/07/1995

SEGMENTATION DES LÉSIONS CARIEUSES PAR L'APPRENTISSAGE PROFOND : REVUE SYSTÉMATIQUE DE LITTÉRATURE

JURY

Madame la Professeure Brigitte GROSGOGEAT-BALAYRE	Présidente
Monsieur le Professeur Cyril VILLAT	Assesseur
Madame la Docteure Marie-Agnès GASQUI-SAINT JOACHIM	Assesseure
Monsieur le Docteur Raphaël RICHERT	Assesseur
Monsieur Fabien MILLIOZ	Membre invité du jury

UNIVERSITE CLAUDE BERNARD LYON I

Président de l'Université	Frédéric FLEURY
Président du Conseil Académique et de la Commission Recherche	Hamda BEN HADID
Vice-Président du Conseil d'Administration	Philippe CHEVALIER
Vice-Présidente de la Commission Formation	Céline BROCHIER
Vice-Président Relations Hospitalo-Universitaires	Jean François MORNEX
Directeur général des services	Pierre ROLLAND

SECTEUR SANTE

Doyen de l'UFR de Médecine Lyon-Est	Gilles RODE
Doyen de l'UFR de Médecine et de Maïeutique Lyon Sud - Charles Mérieux	Philippe PAPAREL
Doyen de l'Institut des Sciences Pharmaceutiques et Biologiques (ISPB)	Claude DUSSART
Doyen de l'UFR d'Odontologie	Jean-Christophe MAURIN
Directeur de l'Institut des Sciences & Techniques de Réadaptation (ISTR)	Jacques LUAUTÉ
Présidente du Comité de Coordination des Études Médicales	Carole BURILLON

SECTEUR SCIENCES ET TECHNOLOGIE

Directrice de l'UFR Biosciences	Kathrin GIESELER
Directeur de l'UFR Faculté des Sciences	Bruno ANDRIOLETTI
Directeur de l'UFR Sciences & Techniques des Activités Physiques et Sportives (STAPS)	Guillaume BODET
Directeur de Polytech Lyon	Emmanuel PERRIN
Directeur de l'Institut Universitaire de Technologie Lyon 1 (IUT)	Michel MASSENZIO
Directeur de l'Institut des Science Financière & Assurances (ISFA)	Nicolas LEBOISNE

28 septembre 2023

Directeur de l'Observatoire de Lyon	Bruno GUIDERDONI
Directeur de l'Institut National Supérieur du Professorat & de l'Éducation (INSPÉ)	Pierre CHAREYRON
Directrice du Département-composante Génie Électrique & des Procédés (GEP)	Rosaria FERRIGNO
Directrice du Département-composante Informatique	Saida BOUAZAK BRONDEL
Directeur du Département-composante Mécanique	Marc BUFFAT

28 septembre 2023

FACULTE D'ODONTOLOGIE DE LYON

Doyen : Pr. Jean-Christophe MAURIN, Professeur des Universités-Praticien hospitalier

Vice-Doyens : Pr. Maxime DUCRET, Professeur des Universités - Praticien hospitalier
Pr. Brigitte GROSGOGEAT, Professeure des Universités - Praticien hospitalier
Pr. Cyril VILLAT, Professeur des Universités - Praticien hospitalier

SOUS-SECTION 56-01 : **ODONTOLOGIE PEDIATRIQUE ET ORTHOPEDIE DENTO-FACIALE**

Professeur Emérite des Universités-PH : M. Jean-Jacques MORRIER,
Professeure des Universités-PH : Mme Béatrice THIVICHON-PRINCE
Maîtres de Conférences-PH : Mme Sarah GEBEILE-CHAUTY, Mme Claire PERNIER
Maître de Conférences Associé : Mme Guillemette LIENHART

SOUS-SECTION 56-02 : **PREVENTION – EPIDEMIOLOGIE ECONOMIE DE LA SANTE - ODONTOLOGIE LEGALE**

Professeur des Universités-PH : M. Denis BOURGEOIS
Maître de Conférences-PH : M. Bruno COMTE
Maître de Conférences Associé : M. Laurent LAFOREST

SOUS-SECTION 57-01 : **CHIRURGIE ORALE – PARODONTOLOGIE – BIOLOGIE ORALE**

Professeurs des Universités-PH : M. Jean-Christophe FARGES, Mme Kerstin GRITSCH

Maîtres de Conférences-PH : Mme Doriane CHACUN, M. Thomas FORTIN
M. Arnaud LAFON, Mme Kadiatou SY, M. François VIRARD

SOUS-SECTION 58-01 : **DENTISTERIE RESTAURATRICE, ENDODONTIE, PROTHESE, FONCTION-DYSFONCTION, IMAGERIE, BIOMATERIAUX**

Professeure Émérite des Universités-PH : Mme Dominique SEUX

Professeurs des Universités-PH : M. Maxime DUCRET, M. Pierre FARGE,
Mme Brigitte GROSGOGEAT, M. Christophe JEANNIN
M. Jean-Christophe MAURIN, Mme Catherine MILLET
Mme Sarah MILLOT, M. Olivier ROBIN, M. Cyril VILLAT

Maîtres de Conférences-PH : Mme Marie-Agnès GASQUI DE SAINT-JOACHIM
Mme Marion LUCCHINI, M. Thierry SELLI
Mme Sophie VEYRE, M. Stéphane VIENNOT

Maîtres de Conférences Associés M. Hazem ABOUELLEIL-SAYED, Mme Ina SALIASI

SECTION 87 : **SCIENCES BIOLOGIQUES FONDAMENTALES ET CLINIQUES**

Professeure des Universités-PH : Mme Florence CARROUEL

21 mars 2024

RÉSUMÉ

Objectifs : Cette revue systématique vise à évaluer les études utilisant le deep learning (DL) pour la segmentation des caries en 2D et à étudier les performances des différentes méthodes.
Méthodes : Les recherches bibliographiques ont été effectuées dans MEDLINE, Embase, IEEE explore et Web of science jusqu'en décembre 2023, avec l'équation de recherche suivante : deep learning AND segment* AND (cari* OR decay OR dental cavity OR pulp*). Cette étude a été enregistrée dans PROSPERO.
Résultats : Sur 863 identifiés, nous avons inclus 17 articles traitant de la segmentation automatique utilisant le deep learning, publiés entre 2021 et 2023. Entre 141 et 10 000 données par ensembles de données ont été utilisés pour développer ou tester les modèles, utilisant des radiographies panoramiques (n = 7), péri-apicales (n = 7), et rétro-coronaires (n = 3). La majorité des études ont utilisé une étape de pré-traitement (n=15) (filtre, rotations, augmentation). Un modèle dérivé de l'architecture en forme de U (n=5) était développé, parfois même de U-net (n=5) avec un F1 score allant de 53,50% à 96,47%. Peu d'études ont été validées sur des données externes, et la plupart ont développé leur modèle en utilisant les rayons X provenant d'un seul type d'appareil à rayons X.
Conclusions : Les métriques et l'évaluation des performances pour le deep learning ont révélé des résultats intéressants concernant la segmentation des rayons X. Néanmoins, avant de pouvoir généraliser son application, l'ensemble de données d'entraînement doit être élargi, diversifié et standardisé pendant la phase de prétraitement.

ABSTRACT

Objectives: This systematic review aims to evaluate studies using deep learning (DL) for caries segmentation in 2D and to investigate the performance of the different methods.
Methods: Literature searches will were performed in MEDLINE, Embase, IEEE explore, and Web of science until December 2023, with the following search equation: deep learning AND segment* AND (cari* OR decay OR dental cavity OR pulp*). This study was registered in PROSPERO.
Results: Of 863 identified, we included 17 papers dealing with automatic segmentation using deep learning published in 2021-23. Between 141 and 10000 datasets were used for developing or testing the models, from panoramic (n = 7), periapical (n = 7), bitewing (n = 3). Most studies used a pre-processing step (n=15) (filter, rotations, increase). A derivative of u-shape architecture (n=5) was developed, sometimes even of u-net (n=5) with a F1 score from 53,50% to 96,47%. Few studies were validated on external data, and most developed their model using x-rays from a single type of x-ray machine.
Conclusions: The metrics and performance evaluation for deep learning have revealed interesting results regarding x-ray segmentation. Nevertheless, before being able to generalize its application, the training dataset needs to be expanded, diversified, and standardized during the pre-processing phase.

REMERCIEMENTS

A notre présidente du jury,

Madame la Professeure GROSGOGEAT-BALAYRE Brigitte,
Professeure des Universités à l'UFR d'Odontologie de Lyon - Praticien-Hospitalier
Docteur en Chirurgie Dentaire
Docteur de l'Université Lyon I
Habilitée à Diriger des Recherches
Vice-Doyen à l'UFR d'Odontologie de Lyon

Je vous remercie de l'honneur que vous nous faites en acceptant la présidence de notre jury. Je suis reconnaissante pour votre enseignement tout au long de ces années à l'université et à l'hôpital. Votre implication et votre passion pour l'enseignement sont très appréciées. Ces quelques mots sont destinés à témoigner de toute ma considération et mon respect envers vous.

A notre jury de thèse,

Monsieur le Professeur VILLAT Cyril,
Professeur des Universités à l'UFR d'Odontologie de Lyon Praticien-Hospitalier
Docteur en Chirurgie Dentaire
Ancien Interne en Odontologie
Docteur de l'École Centrale Paris
Habilité à Diriger des Recherches
Vice-Doyen à l'UFR d'Odontologie de Lyon
Responsable du département Dentisterie Restauratrice – Endodontie

Je vous remercie de nous faire le plaisir de votre présence au sein de ce jury. Vos enseignements ont fait naître ma passion pour l'odontologie conservatrice. Veuillez trouver ici l'expression de mon profond respect.

A notre jury de thèse,

Monsieur le Docteur RICHERT Raphaël, Chef de Clinique des Universités
Chef de Clinique des Universités - Assistant Hospitalier
Docteur en Chirurgie Dentaire
Docteur de l'Université de Lyon 1

Votre générosité à faire partie de ce jury est très appréciée et nous en sommes honorés. Je tenais à vous remercier pour la qualité de l'enseignement que vous nous avez dispensé tout au long de notre parcours hospitalo-universitaire. Vos conseils et votre pédagogie m'ont permis d'apprendre énormément à vos côtés. Ce travail témoigne de ma gratitude et de mon respect envers vous.

A notre jury de thèse,

Monsieur Fabien MILLIOZ,
Maître de Conférences Université Lyon 1
Enseignant- Chercheur au laboratoire de CREATIS

Un grand merci à vous, pour votre disponibilité et votre engagement. Votre expertise et votre regard critique ont été essentiels pour affiner et renforcer les arguments présentés dans cette recherche.

A notre directrice de thèse,

Madame la Docteure GASQUI DE SAINT-JOACHIM Marie-Agnès,
Maître de Conférences des Universités à l'UFR d'Odontologie de Lyon - Praticien Hospitalier
Docteur en Chirurgie Dentaire

Je tiens à vous remercier chaleureusement, pour votre soutien constant, vos conseils éclairés et votre encouragement tout au long de ce travail. Votre rigueur scientifique et votre attention aux détails ont permis d'améliorer significativement la qualité de ce travail.

A ma mère,

Merci de m'avoir inculqué les valeurs du travail, de la persévérance, et de m'avoir toujours soutenue dans mes choix.

A mes frères,

Pour votre soutien et l'intérêt que vous portez à mon parcours depuis ses débuts. Twino, je suis profondément reconnaissante pour les moments de complicité et de soutien que nous avons partagés.

A Philippe,

Tes sacrifices, tes encouragements et ta présence à mes côtés ont été essentiels pour traverser les moments difficiles et mener à bien ce projet.

TABLE DES MATIÈRES

ABRÉVIATIONS

IA: Intelligence Artificielle

ROI: Region of Interest

IoU: Intersection of Union

AUC: Area Under the Curve

CBCT: Cone Beam Computed Tomography

ReLU: Rectified Linear Units

PRISMA-DTA: Preferred Reporting Items for a Systematic Reviews and Meta-Analyses of Diagnosis Test Accuracy Studies

QUADAS-AI: Quality Assessment tool for Artificial Intelligence-centered Diagnostic test Accuracy Studies

CLAIM: Checklist for Artificial Intelligence in Medical Imaging

CLAHE: Contrast Limited Adaptative Histogram Equalization

MCC: Matthew's Correlation Coefficient

OPT: Orthopantomogramme

DSC: Dice Similarity Coefficient

INTRODUCTION

La carie dentaire est une maladie infectieuse chronique affectant la moitié de la population mondiale (1). Le diagnostic de la sévérité et de l'activité des lésions carieuses est nécessaire, afin de déterminer les traitements cliniques appropriés : interventions thérapeutiques visant à contrôler ou traiter les lésions (2).

Il est souvent aidé d'un **examen radiographique** complémentaire de routine (panoramique, rétro-alvéolaire, ou rétro-coronaire), permettant la visualisation de la profondeur des lésions carieuses dentaires proximales.

Des variabilités significatives de lecture inter-examinateurs sont observées, pour l'interprétation d'une même radiographie préopératoire, à la fois pour la détection des lésions carieuses (surtout précoces) et pour l'évaluation de la profondeur des lésions carieuses (3). Ces divergences dans l'accord inter-examinateurs sont dues à des facteurs tels que la qualité des radiographies, les conditions d'observation, l'expérience des dentistes, et la difficulté de lecture des niveaux de gris par l'œil humain (loi du contraste simultané) (4,5).

1. Segmentation

La computer vision permet à une machine d'analyser une image par la classification, la détection, et la segmentation.
La segmentation d'images est une technique de computer vision qui consiste à découper de façon, le plus souvent automatique, une image en zones de pixels appartenant à une même classe d'objets. Elle extrait des régions d'intérêt (ROI) sur les images radiographiques, par l'identification puis la classification de chaque pixel (rétro-alvéolaires, orthopantomogrammes, tomodensitométries à faisceau conique (CBCT)).

La segmentation automatique peut se faire avec, ou sans l'aide du machine learning/intelligence artificielle. Par exemple, le clustering est un algorithme de segmentation automatique sans ML/IA. Les algorithmes par apprentissage profond permettent d'avoir des qualités de segmentation équivalentes à celle d'un spécialiste.

Il en existe plusieurs types (4,6) :

- Segmentation sémantique : elle peut se faire manuellement ou automatiquement. Elle étiquette chaque pixel d'une image et les classifie comme appartenant ou non à une certaine catégorie, mais elle ne différencie pas les objets appartenant à cette même classe ;

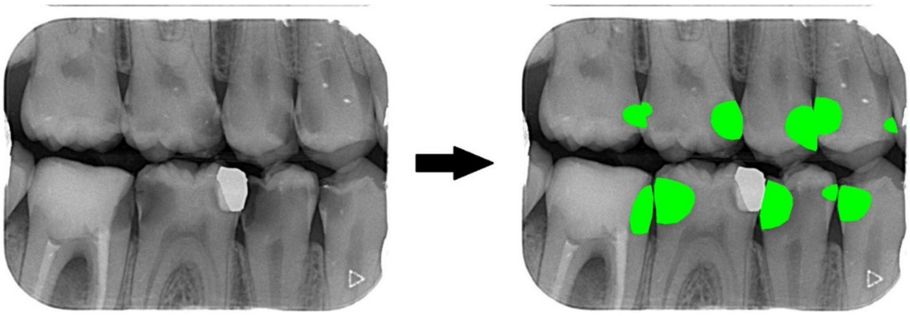

Figure 1 : Segmentation sémantique de lésions carieuses proximales d'une rétro-coronaire (7)

- Segmentation par instance : sépare les différentes instances d'un type d'objet (différencier les ensembles de pixels, comme numéroter les dents), les différents objets d'une même classe sont clairement identifiés ; sur une image d'une rue, la segmentation par instance identifiera les piétons, les voitures, c'est-à-dire les différents types d'objets ;

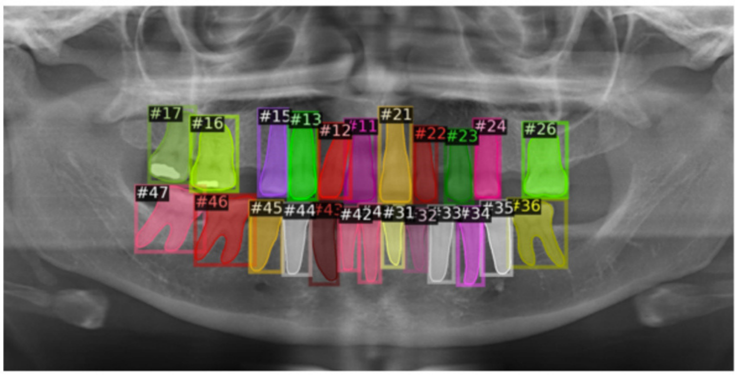

Figure 2 : Segmentation par instance de dents (8)

- Segmentation panoptique : c'est la combinaison de la segmentation sémantique et de la segmentation par instance : elle classifie toutes les zones d'une image, en divisant les instances ; segmente à la fois les objets, et les différents types de fond (ayant une moindre géométrie propre).

Avant l'apparition de l'apprentissage profond, les algorithmes utilisaient d'autres types de segmentation, moins précis, comme (6) :

- Segmentation de seuil : divise les pixels en deux segments en fonction de leur relation avec la valeur seuil. Le seuillage peut être global (image divisée en régions de premier plan et d'arrière-plan, avec une valeur seuil) ou adaptatif (valeur seuil appliquée localement qui dépend des caractéristiques de l'image). C'est une segmentation automatique ;

- Segmentation basée sur la région : regroupe les pixels par des critères similaires (couleur, texture, intensité) ;

- Segmentation basée sur la périphérie : détecte les changements d'intensité ou valeurs de couleur pour marquer les limites.

2. *Intelligence artificielle*

L'**intelligence artificielle** est un procédé logique et automatisé, reposant sur un algorithme, qui simule l'intelligence humaine (9,10).

L'une des branches de l'IA, développée dans les années 80, est l'apprentissage automatique (**machine learning**). Un algorithme détermine les paramètres d'un modèle à partir des données plutôt que de le prédéfinir par des connaissances externes. On cherche à obtenir les paramètres d'un modèle qui seront les plus performants, définis grâce à des données d'entraînement. Une fois l'apprentissage terminé, le modèle est déployé en production (6,11).

Dans les années 2010 naît l'apprentissage profond (**deep learning**), qui, comme le machine learning, apprend de lui-même les caractéristiques les plus pertinentes pour un problème donné, mais avec plus de capacités de calcul et plus de données (6). L'apprentissage profond utilise des réseaux de neurones artificiels interconnectés permettant la résolution de problèmes complexes, comme (12) : le traitement du langage naturel (capacité à interpréter du texte : ChatGPT) ou la **vision par ordinateur**. Il s'agit d'un type particulier d'algorithmes d'apprentissage automatique, caractérisés par un grand nombre de couches de neurones (13). Inspiré d'un neurone biologique : un nœud d'un réseau de plusieurs neurones reçoit plusieurs valeurs d'entrée, et génère une valeur de sortie (3).

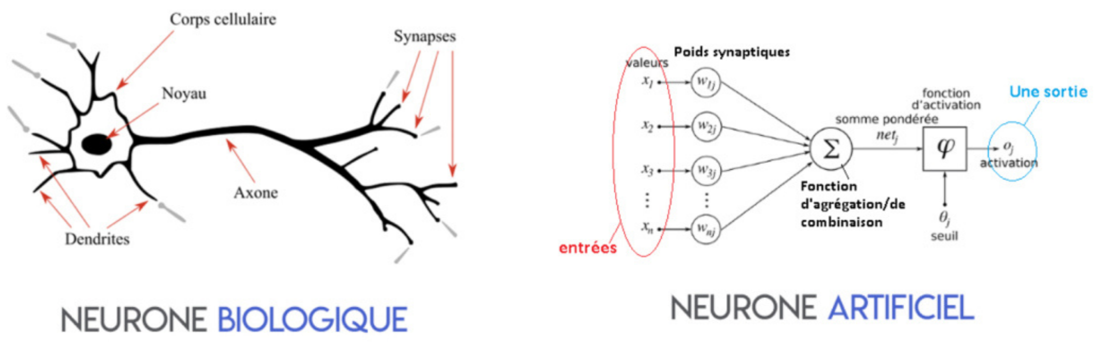

Figure 3 : Neurone biologique humain et neurone artificiel d'IA (14)

Des architectures de réseaux neuronaux profonds se sont succédées, utilisant de plus en plus de couches, afin d'obtenir une faible fonction d'erreur. La précision sur le jeu d'entraînement ne pourra que s'améliorer au fil de l'entraînement, au pire, stagner dans le cas d'overfitting (augmentation de la fonction de perte).

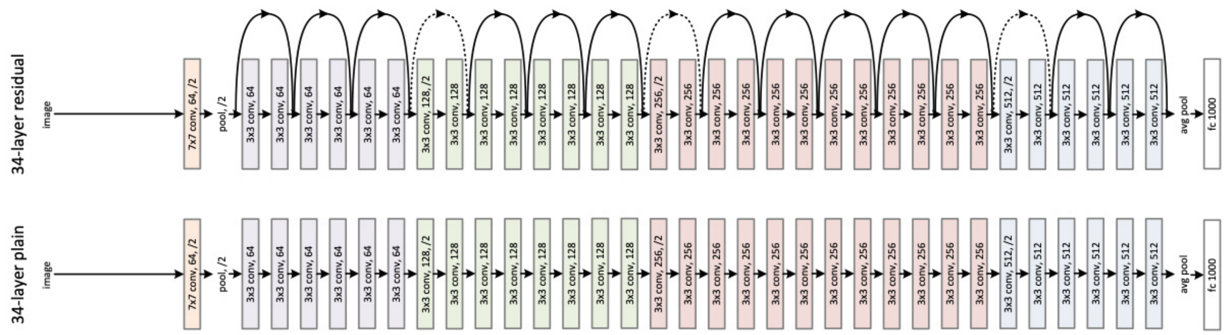

Figure 4 : Architecture d'un réseau simple de 34 couches comparé à ResNet34 (15)

Un réseau neuronal résiduel (**ResNet**) a été conçu. Intégrant des "sauts de connexion" au-dessus de certaines couches, il forme des blocs résiduels. D'autres modèles avec plusieurs connexions en parallèle sont également disponibles, comme DenseNets.

Une couche de réseau apprend une différence par rapport à son entrée, plutôt que sa sortie, quelle que soit la dynamique de cette dernière. Cette architecture est moins susceptible de subir une saturation de précision, améliore la généralisation, et l'activation de réseaux plus profonds de manière plus rapide. Cependant, ils peuvent également augmenter la complexité et les besoins en mémoire du modèle, voire introduire la redondance et le bruit (16).

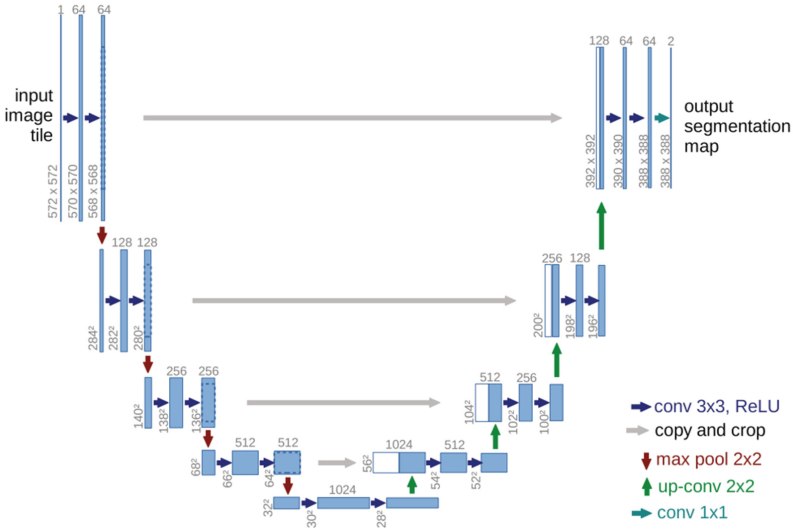

Figure 5 : Architecture de U-Net (17)

En 2015, **U-Net** fut développé. Il est aujourd'hui le plus largement utilisé dans l'imagerie médicale. Sa forme distinctive en "U" comprend :

➔ un bloc downsample / encodeur / contraction / sous-échantillonnage :

- un assemblage de couches de **convolutions** : application de filtres successifs sur tous les pixels de l'image pour obtenir une nouvelle image. Résultat d'un calcul (multiplication matricielle) (18) qui définit la valeur du pixel de l'image résultat, à la position centrale du filtre. Elles extraient des caractéristiques de l'image à différents niveaux et voient des portions de l'image de plus en plus grande, permettant d'utiliser le contexte de l'image (6) ;

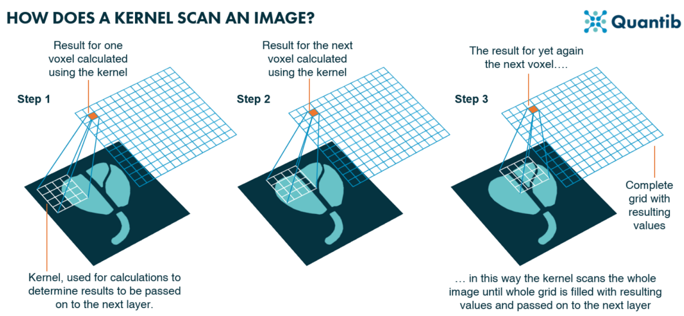

Figure 6 : Une matrice de convolution (18)

- une couche d'activation **ReLU** (Rectified Linear Units) (6) : tous les réseaux de neurones ont une fonction d'activation à chaque couche, que ce soit ReLU ou autre. Cette non-linéarité nécessaire, est similaire à un neurone classique (activé/non activé). Sinon le réseau serait équivalent à une seule couche, où la sortie serait directement une fonction de l'entrée ;

- puis une couche de **max pooling** : diminue le nombre de pixels, réduisant les dimensions spatiales et augmentant la profondeur des cartes de caractéristiques. Cela permet de réduire le temps de calcul et d'être peu coûteux en mémoire de stockage.

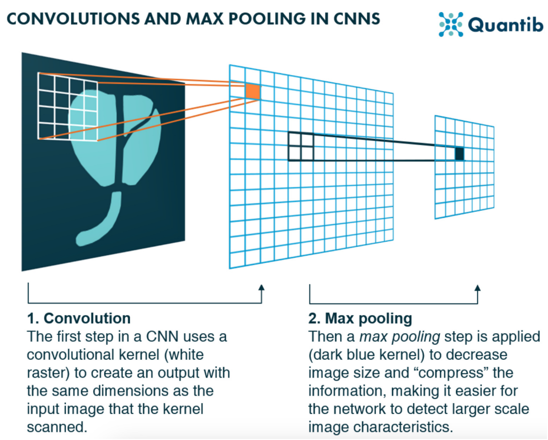

Figure 7 : Sous-échantillonnage par réduction de dimension (19)

➔ un pont : relie les deux chemins ;

➔ un bloc upsample / décodeur / expansion / sur-échantillonnage : même chemin inverse pour aboutir de nouveau à l'image de taille donnée (19–21).

Ce modèle présente une symétrie, ce qui lui permet de fusionner les caractéristiques contextuelles provenant du premier bloc avec des informations de localisation plus détaillées provenant du second bloc : grâce aux **connexions résiduelles**, pour éviter de perdre des informations précises, comme la localisation, au cours de l'expansion (21).

3. *Entraînement et évaluation des modèles*

Le modèle est entraîné par un processus itératif, une mise à jour répétée du modèle en fonction du calcul de l'erreur, entre la sortie de la fonction escomptée (segmentation par des experts ou vérité terrain) et la sortie réelle (segmentation par le réseau) : fonction de perte ou loss. Le but étant que l'erreur obtenue devienne faible.

Les paramètres internes sont optimisés lors de l'entraînement. A chaque itération, les métriques sur le jeu d'entrainement et sur un jeu de données différentes de celles d'entrainement (jeu de validation), sont calculées pour vérifier qu'il n'y est pas d'overfitting (surentraîner sur les données d'entraînement) et que l'entraînement se passe bien en définissant les hyperparamètres (early stopping = hyperparamètre). Enfin l'estimation des performances est définie lors du test sur un troisième jeu de données dédié.

Une fois entraîné, de nouvelles données sont entrées, et il en résulte des prédictions de la part du modèle (inférence) (6).

Les performances des modèles sont évaluées par des métriques (6,22):

- **Justesse** (accuracy) *:* $\frac{Proportion\ de\ pixels\ correctement\ classés}{Nombre\ total\ de\ pixels}$, doit être >0,5 pour être acceptable, proche de 1 pour être qualitatif ;
- Rappel : $\frac{Vrais\ positifs\ (pixels\ correctement\ segmentés)}{Vrais\ positifs+faux\ négatifs\ (tous\ les\ pixels)}$;
- Précision : $\frac{Vrais\ positifs\ (pixels\ correctement\ segmentés)}{Vrais\ positifs+faux\ positifs\ (tous\ les\ pixels\ segmentés)}$, proportion de prédictions correctes avec un jeu de données test (doit être proche de 1, sans être =1 (= surapprentissage, test avec données d'entrainement)) ;
- **Score F1** : $2\ x\ \frac{Précision \times Rappel}{Précision+Rappel}$, moyenne harmonique de la précision et du rappel (doit être proche de 1) ;
- Indice de Jaccard (IoU) : $\frac{Taille\ de\ l'intersection\ entre\ l'objet\ et\ sa\ segmentation}{Taille\ de\ l'union\ entre\ l'objet\ et\ sa\ segmentation}$, nombre de pixels appartenant à la fois à la segmentation et à l'objet réel, divisé par le nombre de pixels appartenant soit à la détection, soit à l'objet réel (0= aucun recouvrement, 1= parfait recouvrement entre la détection et l'objet) ;
- **Dice** : $\frac{2 \times Taille\ de\ l'intersection\ entre\ l'objet\ et\ sa\ segmentation}{Taille\ de\ l'objet + Taille\ de\ sa\ segmentation}$, le dice sert à voir si la segmentation de manière globale est bonne quand on compare la segmentation vérité terrain à la segmentation faite par l'algorithme ;
- **Distance de Hausdorff** : (non utilisée dans les articles inclus mais est pourtant très complémentaire de la mesure précédente) la distance de Hausdorff entre deux ensembles de points est la distance maximale d'un point d'un sous-ensemble au point le plus proche de l'autre sous-ensemble (23). Autrement dit, c'est la mesure les écarts

de 2 surfaces en rapport d'une tangente, deux surfaces peuvent se superposer parfaitement, sauf en un point éloigné (outlyers).

Le travail de cette thèse d'exercice a consisté à l'élaboration d'une revue systématique de littérature lors d'un travail collaboratif. En réponse à une question précise de recherche, ici "(deep learning) AND segment* AND (cari* OR decay OR dental cavity OR pulp*), ce travail s'attache à screener la littérature scientifique à la recherche preuves suffisamment solides pour y répondre. Ici, quelles sont les méthodes répertoriées pour segmenter une lésion carieuse à l'aide du deep learning, et leur pertinence.

SYSTEMATIC REVIEW OF CARIOUS LESION SEGMENTATION USING DEEP LEARNING TECHNIQUES

Amadou Diaw Ndiaye [a], Marie Agnès Gasqui [b,c], Fabien Millioz [d], Matthieu Perard [e,f], Bérengère Bayon [g]

Fatou Leye Benoist [a], Brigitte Grosgogeat [b, c*],

[a] Service d'Odontologie Conservatrice-Endodontie, Université Cheikh Anta Diop, Dakar, Senegal

[b] Laboratoire des Multimatériaux et Interfaces (LMI), UMR CNRS 5615, Université Claude Bernard Lyon 1, Lyon, France

[c] Hospices Civils de Lyon, France

[d] CREATIS (Centre de Recherche en Acquisition et Traitement de l'Image pour la Santé) - CNRS UMR 5220 – INSERM U1294 – Université Claude Bernard Lyon 1 – INSA Lyon - Université Jean Monnet Saint-Etienne, France.

[e] Univ Rennes, INSERM, LTSI – UMR 1099, Rennes, F35000, France

[f] CHU, Rennes, France

[g] Université Claude Bernard Lyon 1, Lyon, France, BP: 69372

* Corresponding author at: Laboratoire des Multimatériaux et Interfaces (LMI), UMR CNRS 5615, Université Claude Bernard Lyon 1, Lyon, France, BP: 69372.

E-mail address: brigitte.grosgogeat@univ-lyon1.fr

INTRODUCTION

Caries lesion constitutes the most common pathology in the world. In 2017, untreated caries on permanent teeth remained the most common health problem, affecting, according to the Global Burden of Disease, 2.3 billion people and more than 530 million children when considering deciduous teeth (24).

The most common routine clinical imaging methods used as complementary diagnostic tools to detect, identify and classify carious lesions are intra-oral radiographs (bitewing and periapical) and dental panoramic imaging, as well as three-dimensional (3D) cone-beam computed tomography (CBCT).

Radiographic detection of advanced carious lesions has limited risks of false positives (high sensitivity and specificity) (25). However, there are significant variabilities in inter-examiner reading for caries lesion detection, even when using the same radiograph (26). These discrepancies in the inter-examiner agreement are due to factors such as radiograph quality, observation conditions and dentists' experience (27,28).

Thanks to pioneering studies focused on artificial intelligence in dentistry, several deep learning systems oriented towards the detection of carious lesions using radiographs are currently available (Mertens 2021, Ezhov 2021, García-Cañas 2022). In addition, recent systematic reviews have shown the satisfactory performance of machine learning and deep learning models in the detection and classification of caries lesions (Foros 2021, Reyes 2022, Mohammad-Rahimi 2022, Talpur 2022, Albano 2024). Furthermore, bitewing radiography is the most commonly used imaging modality and detection remains the most explored computer vision task (Ndiaye et al. 2023).

Apart from classification and detection, image segmentation is an essential area of computer vision, as illustrated by the numerous research projects involving both algorithms based on image processing and learning techniques. Segmentation is one of the most challenging tasks in image processing (29). Through segmentation, the image is fragmented into a series of regions and objects of interest that can be extracted. It can also be used to extract useful information from images for medical imaging applications, and thus, many medical image segmentation applications have been introduced (4,30–32). However, to date, no systematic review has been carried out specifically on the segmentation of carious lesions using deep learning.

This systematic review aimed to evaluate studies using deep learning (DL) for caries segmentation in 2D and to investigate the performance of the different methods.

MATERIAL AND METHODS

The methodology of this protocol was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of diagnostic test accuracy studies (PRISMA-DTA) (Page et al. 2020) (33). The review had been registered and had a protocol published in PROSPERO (International prospective register of systematic reviews: CRD42023475420).

1. Eligibility criteria

This systematic review answers the following PIRD question (Population, Index test, Reference test, and Diagnosis of interest): What are the methods used and the performance of deep learning (index test) in the segmentation of caries lesions (diagnosis of interest) in imaging compared with human experts (if available, reference test)?

- **Inclusion criteria**
  This systematic review identifies all primary publications that have used deep learning to segment caries lesions using 2D radiography.

- **Non-inclusion criteria**
  Studies that used CBCT and photography as the radiographic database were not included. Articles on optical coherence tomography or transillumination were also not included.

- **Exclusion criteria**
  Studies focusing on detection and classification as computer vision spots were excluded. Studies that focused on traditional machine learning techniques and those that did not have segmentation as a primary task were excluded. Studies in which the segmented object was not the carious lesion and those containing duplicates were also excluded.

2. Information sources

Literature searches were performed in MEDLINE, Embase, IEEE explore, and Web of science until December 2023, without language restrictions. Manual searches of the bibliographies, citations, and related articles (PubMed function) of the included studies were performed to identify any other relevant elements not previously identified. The reference lists of the selected studies and those of previous systematic reviews were used to complete the literature search.

3. *Search strategy and study selection*

The following search criterion was the individualized strategy used for each database: "(deep learning) AND segment* AND (cari* OR decay OR dental cavity OR pulp*)".

After exclusion of duplicates, titles and abstracts of records identified through literature search were independently screened for eligibility for inclusion with the help of Rayyan collaboration and research tool (34) for screening. Only clearly irrelevant articles were excluded. Subsequently, full texts of records deemed eligible were retrieved and further assessed for inclusion by the same investigators who resolved any discrepancies through consensus (MAG & BB). A third reviewer (ADN) was consulted when any disagreements or discrepancies persisted between the two reviewers. The PRISMA flowchart was used to illustrate the process of identification, selection and inclusion/exclusion of articles.

4. *Data collection process and data extraction*

Two authors (MAG & BB) independently extracted the main characteristics of the included studies using an Excel extraction sheet before being cross-checked. A third reviewer (ADN) was consulted again to resolve conflicts. Data were extracted as follow:

- Study parameters: authors, year, data source, imaging modality, dataset;
- Imaging parameters: type of radiograph, data size, preprocessing/augmentation segmented object;
- Deep learning parameters: reference test/standard, pre-train, AI algorithm,/segmentation method, external validation, metrics and measure of performance.

5. *Risk of bias and applicability*

Methodological quality was assessed by two raters independently (MAG & BB) using the QUADAS-AI tool for AI-focused diagnostic precision/accuracy studies (Sounderajah et al. 2021) (35) and using personalized signaling questions. Discrepancies or disagreements were resolved by a third researcher (AND).

6. Data reporting

A narrative summary of the included studies was produced. The results of the various studies are presented in tabular form and figures.

RESULTS

1. Study selection

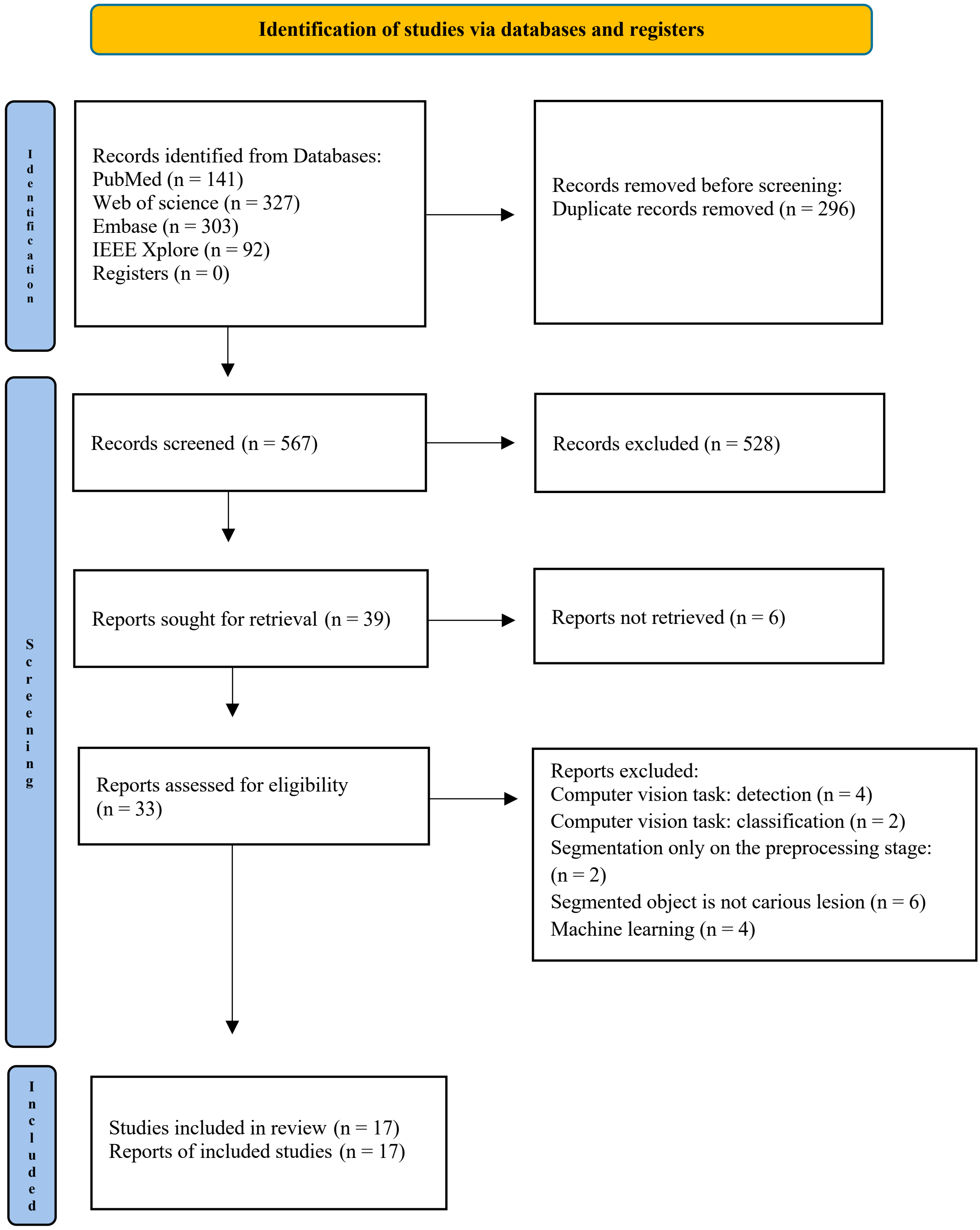

Figure 1: Flow diagram according PRISMA 2020

A PRISMA flowchart diagram illustrating the inclusion workflow is shown in Figure 1. A total of 863 articles were found in the databases, of which 567 were screened after duplicates were removed. Due to various reasons, 528 articles were excluded, such as studies that used detection or classification as a computer vision task, book chapters, conference papers, reviews, studies focused on other pathologies such as cancers, tooth segmentation, bone segmentation, maxillofacial segmentation and others that have nothing to do with caries segmentation. Thus, 33 full manuscripts were individually assessed. During screening with full texts, 16 articles were excluded due to various reasons: the computer vision task was either detection (Fariza 2022, Haghanifar 2020, Bui 2022, Vinh 2023) or classification (Imak 2022, Megalan Leo 2021), studies whose segmented object is not the caries lesion (Megalan Leo 2020, Muresan 2020, Salih 2021, Schneider, 2022), studies used machine learning techniques (Geetha 2020, Majanga CIN 2021, Ramana 2022, Salih 2023) and studies that used segmentation only in the pre-processing stage (Al kheraif et al., Majanga MDPI 2021).

2. *Study characteristics*

Finally, 17 papers dealing with automatic segmentation using deep learning were included in this study. The extracted characteristics from all included articles are provided in Table 1, 2 and 3.

3. *Study and clinical parameters*

Table 1 : Clinical parameters

Reference	Year	Data source	Imaging modality	Dataset
Ahmed et al. (36)	2023	Saudi Arabia	bitewing	554
Baydar et al. (37)	2023	Turkey	bitewing	500
Chen et al. (38)	2023	China	panoramic	1100
Dayi et al. (39)	2023	Turkey	panoramic	504
Gardiyanoglu et al. (40)	2023	Cyprus	panoramic	8138
Qayyum et al. (41)	2023	United Arab Emirates	periapical	141
Ramezanzade et al. (42)	2023	From a previously published randomized clinical trial, multicenter-based, Sweden	periapical	292
Wang et al. (43)	2023	China	panoramic	1000
Zhu et al. (44)	2023	China	panoramic	1159
Ari et al. (17)	2022	Finland	periapical	1169
Lin et al. (45)	2022	China	periapical	3165
Saleh al Ansari et al. (46)	2022	NM	periapical	10000
Velusamy et al. (47)	2022	NM	panoramic	600
Ying et al. (48)	2022	China	periapical	153
Bayrakdar et al. (7)	2021	Turkey	bitewing	621
Lian et al. (49)	2021	China	panoramic	1160
Rajee et al. (50)	2021	Iran, India	periapical	2455

All studies were published between 2021 and 2023 and the majority came from Asia (70,6%). The type of radiograph that were investigated were panoramic (n = 7), periapical (n = 7), bitewing (n = 3). In term of data size, between 141 and 10000 radiographs were used for developing or testing the models in the respective studies (shown in fig. 2).

4. *Imaging parameters*

Table 2: Imaging parameters

	Author	Datasize (train/valid/test)	Data prepocessing	Segmented objects	
bitewing	Ahmed et al.	80% training / 10% testing / 10% validation	- cleaning and augmentation - resized (512x512) - converted to JPEG - contrast was increased	- S1: lesion within the outer half of enamel - S2: lesion within the inner half of enamel - S3: lesion within the outer half of dentin - S4: lesion within the inner half of dentin	
bitewing	Baydar et al.	80% training / 10% testing / 10% validation	NM	- dental caries - dental crown - dental pulp	- dental restorative filling material - dental root canal filling material
bitewing	Bayrakdar et al.	Augmented to 2292 for segmentation model 1968 training / 64 testing / 260 validation	- resized (512 × 512) - horizontal and vertical flip - CLAHE	carious lesion	
periapical	Ari et al.	80% training / 10% testing / 10% validation	- resized (512x512) - intensity normalization - Contrast Limited Adaptive Histogram Equalization (CLAHE)	- carious lesion - crown - dental pulp	- dental filling - root canal filling - periapical lesion
periapical	Lin et al.	Reduced to 800 after exclusion 600 training / 200 testing	- cropped into images containing two posterior teeth - converted to grayscale images	- level 0: non-proximal caries - level 1: proximal caries limited to the outer half of the enamel - level 2: proximal caries limited to the inner half of the enamel - level 3: proximal caries limited to the dento–enamel junction - level 4: proximal caries limited to the outer half of the dentin - level 5: proximal caries limited to the inner half of the dentin	
periapical	Qayyum et al.	141 labelled images (90% training / 10% testing) and 88 unlabeled images	- centroid cropping	carious lesion	
periapical	Ramezanzade et al.	NM	- downscaled by a factor of two - pixel intensity values were normalized to the 0–1 range - random horizontal flipping, affine transformations, and perspective changes	carious lesion	

	Saleh Al Ansari et al.	5000 training/ 5000 testing	- Wiener filter	carious lesion
	Ying et al.	40 testing The remaining 113 images were then augmented to 800	- translation, flip, mirror, rotation, vertical perspective deformation, oblique quadrangular perspective deformation, region warping, and shear mapping transformation (with the help of a library package Augmentor) - resized (224 × 224)	carious lesion
	Rajee et al.	NM	- RedGreenBlue (RGB) to grey - Binary Histogram Equalized Image (BHEI)	- periodontal disease - enamel caries - pericoronal disease - periapical diseases
panoramic	Lian et al.	1071 training and validation / 89 testing	- a 300 × 400 ROI image for each caries area was cut - horizontal flip, vertical flip, horizontal vertical flip, and random rotation (within 0 and 15 degrees)	- D0: sound - D1: caries radiolucency in enamel or in the outer third of dentin - D2: caries radiolucency in the middle third of dentin - D3: caries radiolucency in the inner third of dentin with or without apparent pulp involvement
	Dayi et al.	Augmented to 746 occlusal caries, 1627 proximal caries, 378 cervical caries 75% training / 25% testing	- cropped at 540 × 1300 - reduced to 256 × 512	- type I : occlusal caries - type II : proximal caries - type III : cervical caries
	Wang et al.	Augmented to 7500 caries lesions	- crop the entire training dataset into 384 square areas	- shallow caries - middle caries - deep caries
	Velusamy et al.	NM	- resized (416x416)	- cavitated - non cavitated caries
	Gardiyanoglu et al.	80% training / 10% testing / 10% validation	- converted into PNG files and transferred to Computer Vision Annotation Tool (CVAT), - resized (512x1280)	- teeth - crown - bridge restorations - dental implants - root canal fillings - composite - amalgam - dental caries - residual roots
	Chen et al.	NM	- Mask R-CNN	- tooth - outlines of enamel - dentine - pulp - crown restoration - caries grades
	Zhu et al.	Augmented to 3217 caries regions	NM	- shallow carie - moderate carie - deep carie

Six studies only segmented the caries, five segmented and compared different depths of the caries, one segmented different locations of caries (occlusal, proximal and cervical), finally five other studies also segmented other structures such as: enamel, dentine, pulp, crown restauration, composite, dental implants, residual roots, supernumerary teeth, jaws position…

Regarding pre-processing, we found a resizing in nine studies, with a dimension of 512x512 for three studies. Some applied image enhancement techniques such as Contrast Limited Adaptive Histogram Equalization (CLAHE) (Bayrakdar et al. 2022, Ari et al. 2022) to make caries more visible by improving contrasts (7). For Rajee et al., the preprocessing technique used a binary histogram equalized image: a method that employs multiple thresholds to segment the image into different areas of brightness, such as a background and various other objects (Rajee et al. 2021) (52).

Others simply used cleaning and augmentation, increased contrast (Ahmed et al. 2023), random rotation flipping (horizontal-vertical-mirror) and perspective deformation (Ramezanzade et al. 2023, Ying et al. 2022, Lian et al. 2021) to simulate different viewing conditions (42).

To minimize the noise in dental caries radiography images, Saleh Al Ansari et al. (2022) used the developed Wiener filter: a linear filter used to conceal visual noise.

Chen et al. (2023) could crop each tooth from panoramic image, using the pixel-level individual tooth mask obtained by MasK R-CNN, for subsequent processing. Some have cropped the images, particularly for studies involving panoramic, to reduce the image to the region of interest, but also to increase the data sample (artificial augmentation of data). Dayi et al. (2023), Wang et al. (2023) and Zhu et al. (2023) increased respectively 504, 1000 and 1159 x-rays to 2759, 7500 and 3217 caries regions. Furthermore, Bayrakdar et al. (2022) obtained 2292 images (from 621) by applying horizontal and vertical flip and both.

Whilst seven studies used a ratio of 80% of the data for training, 10% for testing, and 10% for validation; the remaining studies used a different ratio as 50%/50% (Saleh Al Ansari et al. 2022) or it is not mentioned precisely.

5. *Deep learning parameters*

Table 3: Deep learning parameters

	Author	Reference test	Pre-train	Segmentation method/ AI algorithm	External validation	Metrics and measure of performance
bitewing	Ahmed et al.	2 restorative dentists were calibrated	3 encoders pre-trained on Image Net weights transfer learning	U-Net	NM	Mean F1 score 0.535 Mean IOU 0.551 - S1: 0.21 - S2: 0.23 - S3: 0.35 - S4: 0.41
	Baydar et al.	2: 3 years experienced research assistant – 11 year experienced oral and maxilla-facial radiologist	Using 200 epochs	U-Net	NM	F1 score 0.8818 Recall 0.8235 Precision 0.9491
	Bayrakdar et al.	2 experienced dento-maxillofacial radiologists (for segmentation)	With 100 epochs	VGG-16 (detection) U-Net (segmentation)	5 different observers were evaluated to external datasets for the compare AI model performance	F1 score 0.84 Precision 0.86 Sensitivity 0.81
periapical	Ari et al.	2: a research assistant (2 years of experience) / a dento-maxillofacial radiologist (12 years)	Using different epoch values for each situation (800 for caries)	U-Net	NM	F1 score (carious lesion) 0.82 Precision 0.82 Sensitivity 0.82
	Lin et al.	3 examiners (5 years of clinical experience)	Yes	3 preprocessing strategies: - Image Recognition (without preprocessing) - Image Segmentation - Edge Extraction Cifar-10Net CNN network (for detection)	NM	Accuracy: - IR 82.1% - EE 85.9% - IS 60.6% - human eyes 78% Performance IS : 90,3% level 0, 14,3% level 3, 16,9% level 4, 23,8% level 5 Performance human eyes : 84,5% level 0, 54,3% level 3, 63,9% level 4, 87,8% level 5

	Qayyum et al.	A team of expert train by an expert dental radiologist (20 years)	Microsoft's COCO dataset	SSL : - generative models (Deeplab-v326, FCN, LRASPP) - classifier models (ResNet-5029, ResNet-10129, Mobilenet-v330)	NM	For Deeplabv3-ResNet-101 Accuracy 98.38 mIoU 50.18 Dice 0.50
	Ramezanzade et al.	1 dentist (3 years of clinical experience); supervised by 2 experienced clinicians (25 years)	On the ImageNet database	ResNet50 and a neural network trained on the distance between the pulp and lesion	NM	F1 score 0.71 Accuracy 0.78 Sensitivity 0.62 Specificity 0.83 AUC 0.73
	Saleh Al Ansari et al.		Yes	BatWhaleOptimization (segmentation) GLCM is used for feature extraction CNN (classification)	NM	F1 score 91% Recall 92% Accuracy 99.12%
	Ying et al.	2: professional stomatologist (more than 10 years clinical experience); a senior stomatologist (20 years) reviewer	With ImageNet (150 epochs)	Draws on the implementation of UNet and Trans-UNet	NM	Precision 0.7443 Dice 0.7487
	Rajee et al.		NM	Curvilinear Semantic D-CNN (segmentation) Inception Resnet V2 (classification)	NM	Accuracy 0.95 (enamel caries segmentation) DICE 0.83 Jaccard 0.98 MCC 0.72
	Lian et al.	3: dental experts (3-15 clinical experience); a fourth reviewed	nnU-Net automatically configures itself, including training	nnU-Net (segmentation as preprocessing) DenseNet121 (classification)	NM	F1 score 0.902 IoU 0.785 Accuracy 0.986 Precision 1.000 Dice 0.663
panoramic	Dayi et al.	2: 8 years restorative dental specialist / 5 year oral and maxilla-facial radiologist	With TensorFlow-Keras library	DentalCariesDetectionNet (with different backbone networks : MobileNetV2, VGG16, ResNet50, EfficientNet, Inception)	NM	Best weighted averages of F1 score (for RestNet50-DCDNet) 62.79 - type I: 70.79 - type II: 67.65 - type III: 18.64 Time consumption 0.0983 s
	Wang et al.	5 dentists	On PyTorch Lightning with a NVIDIAGeForce RTX 3090 GPU	SSL : MULA network (ResNet34 encoder and U-Shape network backbone)	NM	DICE 71,12% Accuracy 76,94% under 530 slices

	Velusamy et al.		With 200 iterations	YOLOv3 (segmentation) Faster R-CNN (detection)	NM	F1 score 96.467% Recall (on 600 data) 96.657% Precision 92.759% Execution time 2,56 ms
	Gardiyanoglu et al.	2 dento-maxillofacial radiologists	Trained separately for each of the structures	U-Net	NM	Dice similarity coefficient (caries) 0.88
	Chen et al.	2 dentists	Using 200 epochs	Mask R-CNN (ResNet101 to extract individual tooth) Modified U-Net	NM	F1 score 0.780 Recall 0.823 Precision 0.741 Overall accuracy 0.782 Dice coefficient (caries) 0.7248
	Zhu et al.	3 dentists; checked by a fourth	Using 200 epochs	CariesNet (U-Shape network)	NM	F1 score 92.87% Accuracy 93.61% Precision 94.09% DSC 93.64% Recall 86.01%

The algorithm most often used in studies was U-Net (n=5) (7,17,36,37,40), and the other studies were inspired by algorithms to create theirs, such as Dayi et al. with DentalCariesDetectionNet (with different backbone networks: MobileNetV2, VGG16, ResNet50, EfficientNet, Inception). The encoder part consisted of pre-trained backbone network architectures such as VGG16, MobileNet and EfficientNet; and the last part contained a Multi-Predicted Output (MPO) structure, the final feature map was divided into three different paths for detecting occlusal, proximal and cervical caries (Dayi et al. 2023) (39). The highest performance was obtained with the ResNet50-DCDNet architecture with F1 score (62.79%) and produced effective results for Type I (occlusal) and Type II (proximal) caries detection. According to Dayi et al., the DCDNet architecture they suggested struggled to effectively learn type III (cervical) caries due to limitations in the dataset concerning this particular type and probably due to difficulty of diagnosis of cervical caries with periapical radiographs (superposition).

Ramezanzade et al. also used a ResNet architecture, to examine how offering dental students AI-enhanced radiographic data compared to conventional radiographic information and evaluate its impact on their forecasts regarding pulp exposure post-carious tissue removal. A multi-path neural network was implemented: the first was a CNN based on ResNet-50 architecture, who analyzes the dental radiograph; the second path was a neural network trained on the distance between the pulp and lesion extracted from X-ray segmentations, treatment type, pain and age of the patient, who analyzes numerical clinical features. The AI model had much better performance than all groups (GroupX-ray, GX-ray+clinical data, GX-ray+AI, GX-ray+clinical data+AI), with accuracy 0.78. Finally, "the more agreeable students were to AI predictions, the better their overall performance was" (Ramezanzade et al. 2023) (42).

Semi-supervised learning approach was used by Wang et al. and Qayyum et al. (2023): "a teacher model is trained in a fully supervised learning fashion on real data (to guarantee high-quality pseudo-label generation), which is then used to generate pseudo labels for unlabelled images for training student model. Lastly, the student model is trained on both the real and pseudo labels to ensure better generalization" (41). Qayyum et al. assessed six state-of-the-art deep learning-based segmentation models and determined that the exceptional performance of

Deeplabv3 with ResNet101 backbone was primarily due to the ResNet101 architecture, achieving an accuracy of 98.38% (Qayyum et al. 2023).

Zhu et al. collected 3127 panoramic radiographs containing precisely outlined caries lesions, spanning shallow, moderate, and deep stages. Subsequently, they developed CariesNet, comprising a U-shape network with the additional full-scale axial attention module, aimed at segmenting these three types of caries. CariesNet shows limited performance on moderate caries (DSC 0.694). The models tend to misclassify moderate caries as shallow caries or deep caries, because either the boundaries between deep caries and moderate caries or the boundaries between shallow caries and moderate caries are relatively blurred (Zhu et al. 2023) (44).

Lian et al. used nnU-Net for segmentation as preprocessing, with an overall accuracy of 0.986. The mean accuracy of the dentists was lower than that of the model but not significantly (0.955). For D1 lesions, the recall rate of the model was 0.765, while it was 0.466 for the dentists. For D2 lesions, the recall rate of the model was 0.652, while it was 0.539 for the dentists. For D3 lesions, the recall rate of the model was 0.918, while it was 0.954 for the dentists. The model exhibited greater sensitivity in identifying D1 and D2 lesions compared to dentists, whereas both the model and dentists demonstrated higher efficiency in detecting D3 lesions (deep caries) (Lian et al. 2021) (49).

Velusamy et al. (2022) designed You Only Look Once Version 3 (YOLOv3) as a U-shaped network with a large-scale axial attention module, the study showed it performs best across all data sizes.

Lin et al. (2022) used different training strategies to train their Cifar-10Net CNN to detect proximal caries lesions: 2 machine learning techniques : image recognition (IR) and edge extraction (EE), compared to image segmentation (IS). The EE strategy having the highest values, more sensitive (86.9%) than human eyes (69%), especially for proximal caries limited to the outer half of enamel. The poor performance of the IS recognition mode may be attributed to excessive extrema and noise (45).

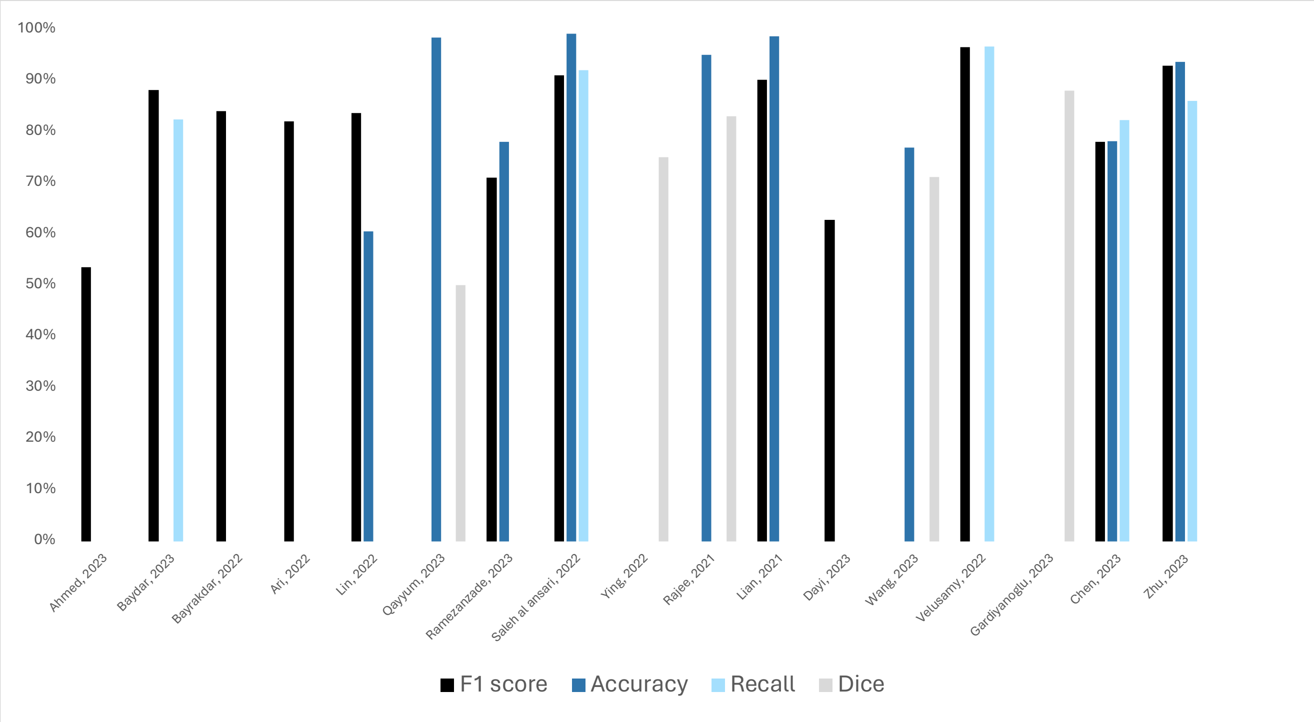

Figure 2: Metrics and measure of performance

The performance outcomes are presented in figure 2. For n=12 studies, the F1 score was presented, and that conducted by Velusamy et al., using YOLOv3 for caries segmentation on 600 panoramic, concluded with the highest score of 96.467%. It also had the best rate of recall on 600 data (96,657%). Ahmed et al. achieved the lowest F1 score of 53.5% in the segmentation of different caries depths in 554 bitewing images.

Saleh et al., showed the best accuracy score (99,12%), with their BatWhaleOptimization, which segmented 10 000 carious lesions in periapical.

For four studies (Baydar et al., Chen et al., Zhu et al., Velusamy et al.,) the training was 200 epochs.

Most reference tests, which are clinical dentists or radiologists, have between 2 and 25 years of experience. There were at least two of them, sometimes supervised by another even more experienced clinician (n=5).

6. Bias analysis

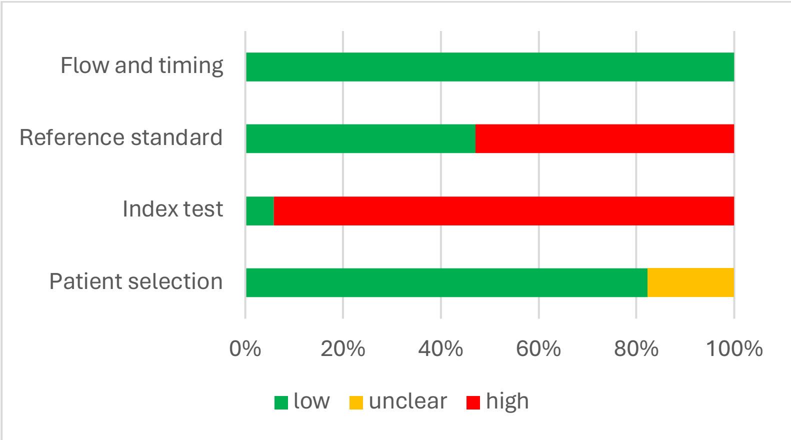

Figure 3: Risk of bias

The index test presented a high risk of bias of 95%, due to the few studies validated on external data. 55% high risk of bias was also present due to some studies not using an appropriate reference standard (the number of readers who label the radiographs).

Patient selection and flow revealed a low risk of bias, due to fairly detailed criteria in the studies, such as: inclusion/exclusion criteria, origin and number of data, preprocessing steps, annotation by the same people.

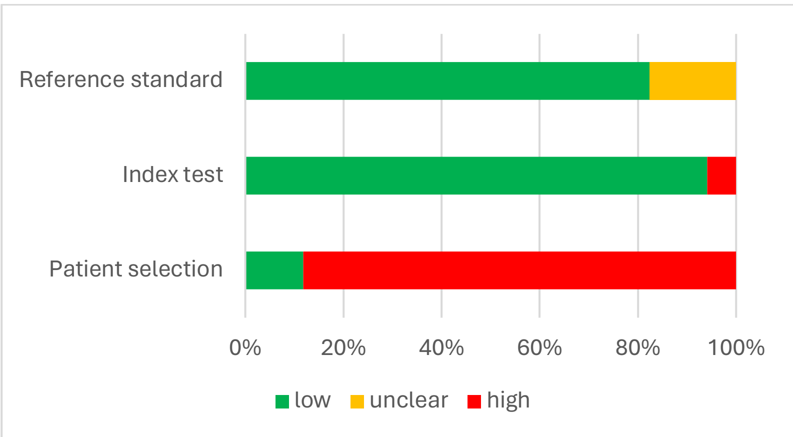

Figure 4: Applicability concerns

Most studies developed their model using x-rays from a single type of x-ray machine, leading to a high risk of bias rate of 90%.

The architectures of the models were well developed in the articles (index test), and the annotating dentists seemed to be able to segment caries correctly.

DISCUSSION

Interest in artificial intelligence increases significantly, according to the number of recent studies, from 2021 to 2023. The majority of these single-center retrospective studies carried out a pre-processing step (n=15). The algorithms inspired by U-net and ResNet, performed better than previous models.

1. Data repartition

Compared to natural images, medical images are thought to have unique characteristics and are well fitted to deep learning (45). The included studies presented datasets between 141 and 10,000 radiographs, before augmentation for some.

The most performed x-ray imaging techniques in dentistry are intraoral x-rays which include bitewing (BW) views and periapical (intraoroal radiography). The studies reflect reality, with 58,8% of bitewing and periapical radiography.

Majanga et al. has identified several limitations that hinder performances of segmentation techniques:

- the limited amount of dental images available for training deep learning systems;
- the majority of dental images exhibit low contrast, unclear or indistinct edge boundaries, and noise (parasitic fluctuation or degradation that the image undergoes from the moment of its acquisition until its recording);
- these images typically vary in size, resolution, and scale;
- and many current segmentation methods demand substantial computational resources, which restricts their practical application (51).

For this reason, initials 120 bitewing images have been augmented to give 11114 images in the dataset. Large datasets let the models have more sophisticated architectures, including more parameters. Advanced models can handle more complicated features and detect dental caries in the early stages (44). In the study conducted by Zhu et al., the 1159 panoramic images made it possible to obtain 3217 images including caries regions with cropping of subzones.

The clinical diagnosis of caries would be mainly based on panoramic radiographs, according to Molander, 1996 (52). But there is a lot of noise coming from patients' head swinging, quality of equipment and the operational experience of medical staff, who are responsible for a large number of artifacts and great disturbance to recognition based on low-density shadows (53).

Datasets compiled from multiple dental facilities across Denmark and Sweden have a stronger capacity for generalization ref. Our review lacked multicentered studies, whereas a diverse dataset, combined with data augmentation, could have improved the applicability of the models.

A large volume of training data allows the model to encounter more examples. A large amount of validation data provides a greater diversity of information to select the best hyperparameters and the most suitable model. An ample supply of test data ensures a more reliable assessment of the model's ability to generalize to unobserved data.

In case of limited availability of data, it is better to devote it mainly to learning. A split such as 80% for training, 10% for validation, and 10% for testing is most common. However, it should be noted that this approach could bias the estimate of the final model performance (6).

2. *Medical imaging*

The results indicate no noticeable differences between the different categories of X-rays. With compared averages for F1 scores of 75.22% for studies using bitewing, 83.58% for periapical, and 82.53% for panoramic. The best rate of F1 score was represented by the study of Velusamy et al. (96,47%) on panoramic, and the two best accuracy rates were represented by studies using periapical (Saleh Al Ansari et al. for 99% and Qayyum et al. for 98,38%).

The constraints of traditional radiography primarily stem from representing caries lesions in 2D, whereas they are 3D structures in reality. This discrepancy may result in the loss of critical information. Furthermore, the radiographic depiction of a lesion is based on the selected projection geometry. Transitioning from film to digital detectors does not alleviate these fundamental limitations. Prior research indicates that CBCT (cone beam computed tomography) surpasses intraoral digital imaging systems as a superior imaging modality.

In Esmaeilyfard's study, the precision of classifying the location of dental caries was superior to that of Dayi et al., particularly regarding cervical caries. They acquired three images per case (axial, sagittal, and coronal) from 382 molar teeth with caries and 403 molar cases without caries. These images were then fed into a multiple-input convolutional neural network (CNN) to classify the extent and location of dental caries. The F1 score achieved was 93.2% (54).

As Gardiyanoglu pointed out, improvements are necessary in standardization, such as patient positioning in OPT scans, acquiring larger datasets (>1000) from various institutions, increasing computational power, adopting unsupervised/semi-supervised learning methods, transitioning to prospectively collected data (vs retrospective such as in n=17 studies), and conducting randomized controlled trials, to become truly dependable aids in diagnosis (40). Achieving this would also entail harmonizing parameters, such as utilizing uniform x-ray devices and image enhancement techniques.

3. *Preprocessing*

Another issue developed in a study, is that most dental images used are characterized with low contrast, false or fuzzy edge borders, and noise. And they are mostly multisized, multiresolution, and multiscaled in nature (51).

As the gathered images may have noise or any contract-related problems, it requires the pre-processing process for providing superior performance in dental caries segmentation and delivering best image quality possible to the network to increase its performance. According to Ramana et al., CLAHE illustrated maximum efficiency in final results: the image consists of a set of pixels, where the best image is attained through a linear cumulative histogram. A digital image is determined with a gray level in a range (55).

Geetha et al. applied a Laplacian filter to obtain a sharper image, the edges of the image are highlighted, and the low frequency components are removed. Each subimage is smoothed using a Gaussian filter (56).

4. *Augmentation*

Many studies of dental issue detection through X-rays suffer from inadequate image quantities in their datasets. Having large datasets allows models to adopt more complex architectures with increased parameters. Consequently, these advanced models can effectively manage intricate features and identify subtle irregularities within tooth textures, such as early-stage dental caries (57). This is why Qayyum et al. used an augmentation to increase their sample: centroid cropping. Centroids cropped-images have significantly smaller dimensions compared to the original (high-dimensional) images, yet they have demonstrated superior performance to models trained on original data in self-supervised learning scenarios (41).

5. *Segmentation*

It is generally expensive to acquire annotated medical images, generating accurate labels requires expertise and high labor costs (53). Typically carried out by specialists in the field, such as dentists or radiologists (44). Perceptions vary between them, influenced by their individual experiences and their distinct modes of observation. According to Duggan et al., a panel of three experts is a commonly used method to establish reference standards when ground truth is not available for AI validation (58).

For Ramezanzade et al., ground truth (who reflecting the actual outcome of the treatment : pulp exposure or not (42)) was used as a reference instead of depending solely on expert ratings. This allows us to overcome this bias, when the reference standard is not appropriate (55% risk of bias in our study).

Deep learning has made significant strides in image segmentation tasks as the amount of annotated data has grown. According to Wang et al. the semi-supervised approach emerges as a significant development. To leverage unlabeled data, the pseudo-label learning technique employs an initial model to generate preliminary pseudo-labels. These are then merged with the labeled data to refine the segmentation model, resulting in more precise pseudo-labels. Subsequently, the method iterates through this process multiple times to enhance the model further. However, this iterative process often consumes a considerable amount of time (53). Self-training has demonstrated considerable success in utilizing unlabeled data and have delivered competitive outcomes when compared to fully supervised learning methods (57).

In this analysis, U-Net appeared to be the most utilized. This preference stems from its utilization of sequential convolutional layers, enabling more precise segmentation, particularly with limited training data. Consequently, it is generally favored for image segmentation within the medical domain (17). As the number of layers grows, so does the mathematical computational power. The effectiveness of these networks hinges on both the quantity and quality of the training data provided. Without losing precision during training (due to saturation), in fact, the models offered today avoids this problem by skipping connections like ResNet.

U-Net also provides a considerable advantage in reducing calculation time. As presented in the study by Dayi et al., the time consumption was 23ms for U-Net compared to 98.3ms for ResNet50-DCDNet. In fact, he uses the max pooling step to decrease image size, making it easier for the network to detect larger scale image characteristics. However, the study also demonstrates that ResNet50-DCDNet remains the model with the best F1 score (41).

In the study by Rajee et al., the comparison of different segmentations shows that the best precision is via deep learning. The Fuzzy C-mean clustering method is a machine learning technique in which each data point is separated into different clusters and then assigned a probability score of being in that cluster (59,60). In the region of interest, the affected area is separated from other regions (52). The curvilinear image separates the foreground from the background, then groups together the areas affected by the disease. It does not make it possible to precisely identify the affected area.

Through the various iteration processes, the model proposed by Rajee et al., clusters the affected area optimally, and the resulting cluster centers increasingly converge with the real clusters.

AI's capacity for generating outcomes that are more sensitive, rapid, and dependable when compared to humans is attributed to its notable advantages, including robust mathematical computational capabilities, expansive storage capacity, versatility in task execution, and the capability for continuous training on extensive datasets. The study conducted by Lin et al. shows us this aspect with 3165 periapical, by comparing the performance of human eyes and different segmentation methods (Image Segmentation, Image recognition, Edge extraction). "The EE recognition mode was significantly more sensitive in detecting both enamel and dentin caries than human eyes (p all < 0.05)" (45). In contrast, the IS relied on an algorithm applied to grayscale images, where pixels with high intensity (white) represented vertices and those with low intensity (black) indicated valley bottoms (61). However, it yielded the poorest outcomes in this investigation with 54,9% of AUCs compared to humans with 76,7% (47).

Bayrakdar et al. utilized VGG-16 and U-Net architectures for detecting and segmenting dental caries in bitewing radiographs. Their findings revealed that the AI model surpassed the research assistants, achieving an F1 score of 0.81 in dental caries segmentation (7).

Ari et al. concluded that their model, developed using the U-Net model, improved the success rate of caries lesion segmentation. However, like most studies, this one had certain limitations. It did not use an external dataset, which limits the generalizability of the results to other populations. Additionally, the study lacked observers with different experiences, which may introduce, again, bias in the evaluation of model performance. Finally, the study did not compare their model with different convolutional neural network (CNN) models, which could provide more complete context on the relative effectiveness of their approach (18).

It would be necessary to examine the accuracy of neural networks and their appropriate use in clinical settings. This correct use includes how dentists adopt and use these tools, improving diagnostic procedures, and the impact of these tools on treatment decision protocols (51).

CONCLUSION

In busy medical settings, errors or instances of under-diagnosis may occur due to variations in clinician's expertise and attention. Such challenges could be mitigated by incorporating AI software into the analysis of radiographs, aiding clinicians in both diagnosis and treatment procedures.

The metrics and performance evaluation for deep learning have revealed interesting results regarding x-ray segmentation. Nevertheless, for broadening its application, the training dataset needs to be expanded, diversified, and standardized during the pre-processing phase.

PERSPECTIVES

Cette revue systématique de littérature a examiné les résultats des études disponibles, concernant la segmentation des lésions carieuses sur les radiographies 2D, par des techniques de deep learning.

Ce long travail nous a permis de comprendre l'exigence de l'élaboration d'une revue. Qui doit être menée de manière rigoureuse, structurée, et dont la méthodologie explicite est indispensable pour réduire les biais à chaque étape.

Traitant d'articles publiés relativement récemment, elle montre l'intérêt grandissant pour ces nouvelles technologies. Mais le peu d'études à ce sujet, seulement 17 articles inclus dans cette revue, après exclusion de 550 autres, nous amène à prendre du recul.

Les résultats indiquent que l'IA promet d'être une aide fiable pour le praticien. En particulier, U-Net, un modèle récent de deep learning, notamment adapté pour des tâches de vision par ordinateur, telles que les images médicales. Ils suggèrent que ces modèles amélioreraient nos diagnostics, du fait de leurs remarquables résultats comparés à l'œil humain. Le deep learning a une capacité de généralisation supérieure grâce des bases de données plus grandes et des méthodes de calcul puis puissantes que le machine learning. Ces modèles peuvent prédire des résultats pour des données qu'ils n'ont jamais vues auparavant.

Cependant, plusieurs limitations doivent être reconnues. Premièrement, la majorité des études incluses étaient des études uni-centriques, cette hétérogénéité des données étudiées rendant difficile la généralisation des résultats. Deuxièmement, les modèles n'étaient pas entraîné à l'aide de données externes, ce qui limite leur validation finale. Enfin, trop peu de données étaient utilisé. Le deep learning nécessite des quantités massives de données, en raison de la complexité des réseaux de neurones multicouches.

En conséquence, 87% des projets ne parviendrait jamais à être déployés en production selon une étude (6).

Gardons à l'esprit qu'il s'agit de modèles statistiques, dont on doit optimiser les paramètres avant de pouvoir leur faire véritablement confiance. En diversifiant les données d'entrainement (études multicentriques, appareils radiographiques et patients différents).

BIBLIOGRAPHIE

1. Frencken JE, Sharma P, Stenhouse L, Green D, Laverty D, Dietrich T. Global epidemiology of dental caries and severe periodontitis - a comprehensive review. J Clin Periodontol. 2017 Mar;44 Suppl 18:S94–105.
2. Gugnani N, Pandit I, Srivastava N, Gupta M, Sharma M. International Caries Detection and Assessment System (ICDAS): A New Concept. Int J Clin Pediatr Dent. 2011;4(2):93–100.
3. Neurone artificiel [Internet]. [cited 2024 Apr 8]. Available from: https://www.cnil.fr/fr/definition/neurone-artificiel
4. Buhl Nikolaj. Medical Image Segmentation: A Complete Guide [Internet]. [cited 2024 Apr 8]. Available from: https://encord.com/blog/medical-image-segmentation/
5. Gouvernement du Canada SC. Segmentation d'image en imagerie médicale [Internet]. 2021 [cited 2024 Apr 8]. Available from: https://www.statcan.gc.ca/fr/science-donnees/reseau/segmentation-image
6. Wallach D. Le Deep Learning pour le traitement d'images Classification, détection et segmentation avec Python et TensorFlow. EIHS-DEEPLEAR. ENI; 2024. 536 p.
7. Bayrakdar IS, Orhan K, Akarsu S, Çelik Ö, Atasoy S, Pekince A, et al. Deep-learning approach for caries detection and segmentation on dental bitewing radiographs. Oral Radiol. 2022 Oct;38(4):468–79.
8. ResearchGate [Internet]. [cited 2024 Jun 11]. Figure 5. Visualization of tooth instance segmentation model result. Available from: https://www.researchgate.net/figure/Visualization-of-tooth-instance-segmentation-model-result_fig5_358289548
9. Intelligence artificielle [Internet]. [cited 2024 Apr 8]. Available from: https://www.cnil.fr/fr/definition/intelligence-artificielle
10. Intelligence artificielle — Wikipédia [Internet]. [cited 2024 Jun 11]. Available from: https://fr.wikipedia.org/wiki/Intelligence_artificielle
11. Apprentissage automatique [Internet]. [cited 2024 Apr 8]. Available from: https://www.cnil.fr/fr/definition/apprentissage-automatique
12. Apprentissage profond (deep learning) [Internet]. [cited 2024 Apr 8]. Available from: https://www.cnil.fr/fr/definition/apprentissage-profond-deep-learning
13. Réseau de neurones artificiels (artificial neural network) [Internet]. [cited 2024 Apr 8]. Available from: https://www.cnil.fr/fr/definition/reseau-de-neurones-artificiels-artificial-neural-network

14. Maurice B. Fonctionnement du neurone artificiel [Internet]. Deeply Learning. 2018 [cited 2024 Jun 1]. Available from: https://deeplylearning.fr/cours-theoriques-deep-learning/fonctionnement-du-neurone-artificiel/
15. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition [Internet]. arXiv; 2015 [cited 2024 May 9]. Available from: http://arxiv.org/abs/1512.03385
16. Quels sont les avantages et les inconvénients de l'utilisation de connexions skip dans les CNN ? [Internet]. [cited 2024 May 9]. Available from: https://fr.linkedin.com/advice/0/what-benefits-drawbacks-using-skip-connections?lang=fr
17. Ari T, Sağlam H, Öksüzoğlu H, Kazan O, Bayrakdar İŞ, Duman SB, et al. Automatic Feature Segmentation in Dental Periapical Radiographs. Diagn Basel Switz. 2022 Dec 7;12(12):3081.
18. Fortunati V. Deep learning radiology: the secret of convolutional neural networks [Internet]. [cited 2024 May 9]. Available from: https://www.quantib.com/blog/convolutional-neural-networks-in-deep-learning-radiology
19. Formation Tech et Data en ligne | Blent.ai [Internet]. [cited 2024 May 9]. U-Net : le réseau de neurones populaire en Computer Vision. Available from: https://blent.ai/blog/a/unet-computer-vision
20. Hoang A. Apprendre le Deep Learning. 2023 [cited 2024 May 9]. Tout savoir sur U-Net : l'architecture révolutionnaire pour la segmentation d'images. Available from: https://apprendre-le-deep-learning.com/u-net-une-architecture-pour-la-segmentation-d-images/
21. Benlahmar. Architecture U-Net: Une explication détaillée [Internet]. [cited 2024 May 9]. Available from: https://datasciencetoday.net/index.php/en-us/deep-learning/228-unet
22. luisquintanilla. Métriques ML.NET - ML.NET [Internet]. 2023 [cited 2024 May 9]. Available from: https://learn.microsoft.com/fr-fr/dotnet/machine-learning/resources/metrics
23. Benhabiles H, Vandeborre JP, Lavoué G. Une collection de modèles 3D avec vérité-terrain pour l'évaluation objective des algorithmes de segmentation.
24. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Lond Engl. 2018 Nov 10;392(10159):1789–858.
25. Schwendicke F, Stolpe M, Meyer-Lueckel H, Paris S. Detecting and treating occlusal caries lesions: a cost-effectiveness analysis. J Dent Res. 2015 Feb;94(2):272–80.
26. Estay J, Bersezio C, Arias R, Fernández E, Oliveira Junior O, Andrade M, et al. Effect

of Clinical Experience on Accuracy and Reliability of Radiographic Caries Detection. Int J Odontostomatol. 2017 Jun 20;1:1.

27. Parker JM, Mol A, Rivera EM, Tawil PZ. Cone-beam Computed Tomography Uses in Clinical Endodontics: Observer Variability in Detecting Periapical Lesions. J Endod. 2017 Feb;43(2):184–7.

28. Gasqui MA, Pérard M, Decup F, Monsarrat P, Turpin YL, Villat C, et al. Place of a new radiological index in predicting pulp exposure before intervention for deep carious lesions. Oral Radiol. 2022 Jan 1;38(1):89–98.

29. Gonzalez RC, Woods RE. Digital image processing. 2. ed., internat. ed. Upper Saddle River, NJ: Prentice-Hall; 2002. 793 p.

30. Noble JA, Boukerroui D. Ultrasound image segmentation: a survey. IEEE Trans Med Imaging. 2006 Aug;25(8):987–1010.

31. Gerth S, Claußen J, Eggert A, Wörlein N, Waininger M, Wittenberg T, et al. Semiautomated 3D Root Segmentation and Evaluation Based on X-Ray CT Imagery. Plant Phenomics Wash DC. 2021;2021:8747930.

32. Teixeira LO, Pereira RM, Bertolini D, Oliveira LS, Nanni L, Cavalcanti GDC, et al. Impact of Lung Segmentation on the Diagnosis and Explanation of COVID-19 in Chest X-ray Images. Sensors. 2021 Oct 27;21(21):7116.

33. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, and the PRISMA-DTA Group, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388–96.

34. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016 Dec 5;5(1):210.

35. Sounderajah V, Ashrafian H, Rose S, Shah NH, Ghassemi M, Golub R, et al. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. Nat Med. 2021 Oct;27(10):1663–5.

36. Ahmed WM, Azhari AA, Fawaz KA, Ahmed HM, Alsadah ZM, Majumdar A, et al. Artificial intelligence in the detection and classification of dental caries. J Prosthet Dent. 2023 Aug 26;S0022-3913(23)00478-X.

37. Baydar O, Różyło-Kalinowska I, Futyma-Gąbka K, Sağlam H. The U-Net Approaches to Evaluation of Dental Bite-Wing Radiographs: An Artificial Intelligence Study. Diagn Basel Switz. 2023 Jan 26;13(3):453.

38. Chen Q, Huang J, Zhu H, Lian L, Wei K, Lai X. Automatic and visualized grading of

dental caries using deep learning on panoramic radiographs. Multimed Tools Appl. 2022 Nov 16;82(15):23709–34.

39. Dayı B, Üzen H, Çiçek İB, Duman ŞB. A Novel Deep Learning-Based Approach for Segmentation of Different Type Caries Lesions on Panoramic Radiographs. Diagn Basel Switz. 2023 Jan 5;13(2):202.

40. Gardiyanoğlu E, Ünsal G, Akkaya N, Aksoy S, Orhan K. Automatic Segmentation of Teeth, Crown-Bridge Restorations, Dental Implants, Restorative Fillings, Dental Caries, Residual Roots, and Root Canal Fillings on Orthopantomographs: Convenience and Pitfalls. Diagn Basel Switz. 2023 Apr 20;13(8):1487.

41. Qayyum A, Tahir A, Butt MA, Luke A, Abbas HT, Qadir J, et al. Dental caries detection using a semi-supervised learning approach. Sci Rep. 2023 Jan 13;13(1):749.

42. Ramezanzade S, Dascalu TL, Ibragimov B, Bakhshandeh A, Bjørndal L. Prediction of pulp exposure before caries excavation using artificial intelligence: Deep learning-based image data versus standard dental radiographs. J Dent. 2023 Nov;138:104732.

43. Wang Y, Xia W, Yan Z, Zhao L, Bian X, Liu C, et al. Root canal treatment planning by automatic tooth and root canal segmentation in dental CBCT with deep multi-task feature learning. Med Image Anal. 2023 Apr;85:102750.

44. Zhu H, Cao Z, Lian L, Ye G, Gao H, Wu J. CariesNet: a deep learning approach for segmentation of multi-stage caries lesion from oral panoramic X-ray image. Neural Comput Appl. 2022 Jan 7;1–9.

45. Lin X, Hong D, Zhang D, Huang M, Yu H. Detecting Proximal Caries on Periapical Radiographs Using Convolutional Neural Networks with Different Training Strategies on Small Datasets. Diagn Basel Switz. 2022 Apr 21;12(5):1047.

46. Chawla R, Krishna KH, Deshmukh AA, Sagar KVD, Ansari MSA, Taloba AI. A Hybrid Optimization Approach with Deep Learning Technique for the Classification of Dental Caries. Int J Adv Comput Sci Appl IJACSA [Internet]. 2022 34/30 [cited 2024 Apr 3];13(12). Available from: https://thesai.org/Publications/ViewPaper?Volume=13&Issue=12&Code=IJACSA&SerialNo=41

47. Velusamy J, Rajajegan T, Alex S, Ashok M, Mayuri AV, Kiran S. Faster Region-based Convolutional Neural Networks with You Only Look Once multi-stage caries lesion from oral panoramic X-ray images. Expert Syst. 2023 May 13;

48. Ying S, Wang B, Zhu H, Liu W, Huang F. Caries segmentation on tooth X-ray images with a deep network. J Dent. 2022 Apr 1;119:104076.

49. Lian L, Zhu T, Zhu F, Zhu H. Deep Learning for Caries Detection and Classification. Diagn Basel Switz. 2021 Sep 13;11(9):1672.

50. Rajee MV, Mythili C. Dental Image Segmentation and Classification Using Inception Resnetv2. IETE J Res. 2023 Sep 20;69(8):4972–88.

51. Majanga V, Viriri S. Dental Images' Segmentation Using Threshold Connected Component Analysis. Comput Intell Neurosci. 2021;2021:2921508.

52. Molander B. Panoramic radiography in dental diagnostics. Swed Dent J Suppl. 1996;119:1–26.

53. Wang X, Gao S, Jiang K, Zhang H, Wang L, Chen F, et al. Multi-level uncertainty aware learning for semi-supervised dental panoramic caries segmentation. Neurocomputing. 2023 Jul 1;540:126208.

54. Esmaeilyfard R, Bonyadifard H, Paknahad M. Dental Caries Detection and Classification in CBCT Images Using Deep Learning. Int Dent J. 2024 Apr 1;74(2):328–34.

55. Ramana Kumari A, Nagaraja Rao S, Ramana Reddy P. Design of hybrid dental caries segmentation and caries detection with meta-heuristic-based ResneXt-RNN. Biomed Signal Process Control. 2022 Sep 1;78:103961.

56. Geetha V, Aprameya KS, Hinduja DM. Dental caries diagnosis in digital radiographs using back-propagation neural network. Health Inf Sci Syst. 2020 Dec;8(1):8.

57. Zhu Y, Sapra K, Reda FA, Shih KJ, Newsam S, Tao A, et al. Improving Semantic Segmentation via Video Propagation and Label Relaxation [Internet]. arXiv; 2019 [cited 2024 May 9]. Available from: http://arxiv.org/abs/1812.01593

58. Duggan GE, Reicher JJ, Liu Y, Tse D, Shetty S. Improving reference standards for validation of AI-based radiography. Br J Radiol. 2021 Jul 1;94(1123):20210435.

59. Frédéric D, Brahim EK. Etude de méthodes de Clustering pour la segmentation d'images en couleurs.

60. Satyam Kumar. C-Means Clustering Explained | Built In [Internet]. 2022 [cited 2024 May 17]. Available from: https://builtin.com/data-science/c-means

61. Plateforme Logicielle pour l'Analyse des Signaux et leur Traitement, l'Intégration de données et des Connaissances. La segmentation des images [Internet]. Available from: https://pfl-cepia.hub.inrae.fr/axe-images/tutoriel/la-segmentation-des-images

ANNEXE

	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Ahmed, 2023	+	−	−	+	−	+	+
Ari, 2022	+	−	−	+	−	+	+
Baydar, 2023	+	−	−	+	−	+	+
Bayrakdar, 2022	+	+	−	+	−	+	+
Chen, 2023	+	−	−	+	−	+	+
Dayi, 2023	+	−	−	+	−	+	+
Gardiyanoglu, 2023	+	−	−	+	+	−	+
Lian, 2021	+	−	+	+	−	+	+
Lin, 2022	+	−	+	+	−	+	+
Qayyum, 2023	+	−	+	+	−	+	+
Rajee, 2021	?	−	+	+	+	+	?
Ramezanzade, 2023	+	−	+	+	+	+	+
Saleh al ansari, 2022	?	−	−	+	−	+	?
Velusamy, 2022	?	−	+	+	−	+	?
Wang, 2023	+	−	+	+	−	+	+
Ying, 2022	+	−	−	+	−	+	+
Zhu, 2023	+	−	+	+	−	+	+

− High   ? Unclear   + Low

Tableau 4 : Analyse des biais selon QUADAS-AI et CLAIM

N°2024 LYO1D 036

Bérengère Bayon – Segmentation des lésions carieuses par l'apprentissage profond : Revue systématique de littérature

Résumé :

Objectifs : Cette revue systématique vise à évaluer les études utilisant le deep learning (DL) pour la segmentation des caries en 2D et à étudier les performances des différentes méthodes.

Méthodes : Les recherches bibliographiques ont été effectuées dans MEDLINE, Embase, IEEE explore et Web of science jusqu'en décembre 2023, avec l'équation de recherche suivante : deep learning AND segment* AND (cari* OR decay OR dental cavity OR pulp*). Cette étude a été enregistrée dans PROSPERO.

Résultats : Sur 863 identifiés, nous avons inclus 17 articles traitant de la segmentation automatique utilisant le deep learning, publiés entre 2021 et 2023. Entre 141 et 10 000 données par ensembles de données ont été utilisés pour développer ou tester les modèles, utilisant des radiographies panoramiques (n = 7), péri-apicales (n = 7), et rétro-coronaires (n = 3). La majorité des études ont utilisé une étape de pré-traitement (n=15) (filtre, rotations, augmentation). Un modèle dérivé de l'architecture en forme de U (n=5) était développé, parfois même de U-net (n=5) avec un F1 score allant de 53,50% à 96,47%. Peu d'études ont été validées sur des données externes, et la plupart ont développé leur modèle en utilisant les rayons X provenant d'un seul type d'appareil à rayons X.

Conclusions : Les métriques et l'évaluation des performances pour le deep learning ont révélé des résultats intéressants concernant la segmentation des rayons X. Néanmoins, pour élargir son application, l'ensemble de données d'entraînement doit être élargi, diversifié et standardisé pendant la phase de prétraitement.

Mots clés :

- Carie
- Radiographie
- Segmentation
- Deep learning

Jury :

Présidente : Madame la Professeure Brigitte GROSGOGEAT-BALAYRE

Assesseurs : Monsieur le Professeur Cyril VILLAT

Madame la Docteure Marie-Agnès GASQUI-SAINT JOACHIM

Monsieur le Docteur Raphaël RICHERT

Membre invité : Monsieur Fabien MILLIOZ

Adresse de l'auteur : dr.berengere.bayon@gmail.com