

Dental caries detection in children using intraoral scans and deep learning

Bree Jones ^{a,b,*} , Mathias Lambach ^{c,d}, Tong Chen ^{b,e} , Stavroula Michou ^f , Nicky Kilpatrick ^{a,g} , Nigel Curtis ^{g,h} , David P. Burgner ^{a,g} , Christoph Vannahme ^d , Mihiri Silva ^{a,b}

^a Melbourne Dental School, University of Melbourne, 720 Swanston Street, Carlton, VIC, 3052, Australia

^b Murdoch Children's Research Institute, Royal Children's Hospital Melbourne, 50 Flemington Rd, Parkville, VIC, 3052, Australia

^c University of Southern Denmark, SDU Robotics, Campusvej 55, Odense M, 5230, Denmark

^d 3Shape TRIOS A/S, Copenhagen, Niels Juels Gade 13, 1060, Denmark

^e Clinical Epidemiology & Biostatistics Unit, Murdoch Children's Research Institute, Royal Children's Hospital Melbourne, 50 Flemington Rd, Parkville, VIC, 3052, Australia

^f Department of Odontology, School of Health and Medical Sciences, University of Copenhagen, Nørre Allé 20, 1165, Denmark

^g Department of Paediatrics, University of Melbourne, Royal Children's Hospital Melbourne, 50 Flemington Rd, Parkville, VIC, 3052, Australia

^h Department of General Medicine, Royal Children's Hospital Melbourne, 50 Flemington Rd, Parkville, VIC, 3052, Australia

ARTICLE INFO

Keywords:

Dental Caries
Image Interpretation
Computer-Assisted
Artificial Intelligence
Deep Learning
Quantitative Light Fluorescence
Models
Dental
Imaging
Three-Dimensional

ABSTRACT

Objective: This study aimed to demonstrate the use of deep learning for automating caries detection using intraoral scan data from children and to evaluate diagnostic agreement between the models' predictions and dental practitioner assessments on 3D models.

Methods: Intraoral scans were collected from two cohorts at Murdoch Children's Research Institute. Two researchers annotated scan meshes using the International Caries Classification and Management System. A pre-processing pipeline converted the data into a 2D format. Carious teeth from the first cohort ($n = 332$) were split at the participant level into training ($n = 192$), validation ($n = 63$), and test ($n = 77$) sets. An Attention U-Net was trained to classify initial, moderate, and extensive dental caries. Segmentation and lesion detection performance was evaluated on the test set using the metrics Intersection over Union (IoU), Sensitivity (SE), Specificity (SP), and Precision (P). Carious teeth from the second independent cohort ($n = 119$) were used for external validation. Multilevel logistic regression assessed diagnostic agreement to compare the model performance to dental practitioners across all caries thresholds (initial, moderate and extensive).

Results: For segmentation tasks, the model had the best performance for extensive caries (SE 71 %, P 66 %, IOU 0.55). The model showed overall promising performance for lesion detection (SE 67 %, P 73 %). Performance slightly declined on an external dataset. Diagnostic agreement between the model and dental practitioners was comparable across all disease thresholds: initial (odds ratio OR 0.82, 95 % Confidence Interval (CI) 0.6–1.15), moderate (OR 0.9, 95 % CI 0.5–1.6) and extensive (OR 0.85, 95 % CI 0.42–1.71).

Conclusion: The proof-of-concept demonstrates that deep learning can achieve moderate performance in detecting extensive caries from intraoral scans, though performance was limited for early and moderate lesions. Further research is needed to improve model accuracy and generalisability across all disease stages.

Clinical Significance: This study represents an exploratory effort towards developing AI-assisted caries detection using intraoral scanner data in children. While the long-term potential of such technology could include support for early diagnosis, enhanced caries monitoring, and a reduction in the subjectivity of caries assessment, our current findings indicate that significant model refinement and extensive validation are imperative, especially for the detection of initial carious lesions, before such clinical applications can be realized.

* Corresponding author at: Melbourne Dental School, The University of Melbourne, 720 Swanston St, Carlton 3053, Australia.

E-mail addresses: bree.jones@unimelb.edu.au (B. Jones), msl2610@gmail.com (M. Lambach), tong.chen@mcri.edu.au (T. Chen), stavroula.michou@sund.ku.dk (S. Michou), nicky@bassdata.com.au (N. Kilpatrick), nigel.curtis@rch.org.au (N. Curtis), david.burgner@mcri.edu.au (D.P. Burgner), Christoph.Vannahme@3shape.com (C. Vannahme), mihiri.silva@mcri.edu.au (M. Silva).

1. Introduction

Dental caries is a prevalent issue, affecting 43 % of children worldwide [1]. Early detection and management of dental caries is necessary for timely intervention, and prevention of disease progression.

The use of intraoral scanners in paediatric dentistry is increasing. Studies in children suggest digital impressions are perceived to be more comfortable and have reliable accuracy compared to conventional impressions [2,3]. Recent evidence outlines the potential role of intraoral scanners to support clinical diagnostics [2,4].

Intraoral scanning is emerging as a valuable tool for caries detection, offering potential advantages over traditional visual examination, particularly when combined with advanced imaging and automated analysis. Research suggests that on-screen assessment of intraoral scans by dental practitioners achieves diagnostic performance comparable to that of visual examination [5] although the reliability, has been shown to be influenced by the clinician's level of experience and challenges with identifying colour changes associated with early stages of the disease [6]. Intraoral scanning (IOS) combined with light-induced fluorescence technology can potentially enhance visualisation of early carious lesions [7,8]. Automated rule-based systems using logistic regression have been developed for intraoral scanners that quantify colour and fluorescence loss and compare it to a pre-defined scoring cut-off to indicate dental caries [9]. However, the clinical application of the automated system in primary teeth tends to detect fewer early carious lesions and to underestimate the severity of more advanced carious lesions compared to visual examination [10–12]. Thus, there is a need to explore alternative methods to support reliable and objective dental caries detection.

The International Caries Classification and Management System™ (ICCMST™) offers a standardised and reliable framework for detecting, staging, and managing carious lesions at various levels of severity [13]. It builds upon the foundation established by the International Caries Detection and Assessment System (ICDAS) [14], and combines the ICDAS criteria into three stages: initial non-cavitated enamel lesions, moderate lesions characterised by enamel micro-cavitation or dentinal involvement with intact enamel, and extensive lesions, marked by tooth cavitation with visible underlying dentine. Detecting dental caries in clinical practice can be challenging for clinicians due to the varying characteristics of lesions at different stages of development, each requiring tailored treatment approaches. Initial lesions are reversible and can be managed using non-invasive strategies such as fluoride therapy, dietary modifications, and improved oral hygiene practices. In contrast, cavities involve the destruction of tooth structure and necessitate restorative interventions. Accurately distinguishing between these stages is essential to inform appropriate dental caries management strategies.

Machine Learning (ML) is a potential approach to interpreting intraoral scan data by learning patterns from labelled (annotated) datasets, enabling consistent caries detection across various disease thresholds. Machine learning encompasses a range of algorithms that infer relationships from data without being explicitly programmed [15]. Within ML, deep learning (DL), leverages artificial neural networks, such as convolutional neural networks (CNNs), to process high dimensional inputs such as 3D meshes or 2D projections from intraoral scans. These models are particularly effective for computer vision tasks in dentistry, including: image classification, where the model predicts the presence or absence of dental caries; object detection, which localises and labels carious lesions within an image; semantic segmentation, which assigns a class label (i.e. initial, moderate or extensive) to each pixel to delineate areas of disease; and instance segmentation, which identifies individual lesions as separate and distinct entities. Developing such models involves supervised learning using annotated image datasets. The dataset is typically divided into training, validation and test subsets. During training the model optimises its parameters by minimising a loss function that quantifies prediction errors against known

'ground truth' labels (reference standard). To ensure robustness and generalisability, external validation using data from independent cohorts is recommended [16].

Many studies have proved the concept of using deep learning models for detecting caries using intraoral photographs [17–20]. Unlike intraoral photographs, intraoral scans are polygonal meshwork, that represent the surface geometry of teeth and surrounding structures. While the intraoral scan meshwork exists in three-dimensional space (defined by X, Y, and Z coordinates), it primarily represents the surface geometry of the scanned structures rather than their internal composition or full 3D volume [21]. The surface mesh is made up of facets (triangular polygons) that approximate the surface, defined by vertices and edges, and may also include attributes such as colour, texture, and orientation. Although several algorithms have been described to segment teeth using intraoral scan images [22,23], to our knowledge, caries detection using intraoral scan data and deep learning has not been reported.

This proof-of-concept study aimed to use deep learning for caries detection and classification using intraoral scan data obtained from children and explore the generalisability of the model to an external cohort. Additionally, this study aimed to benchmark the model against dental practitioners.

2. Materials and methods

2.1. Study design

This study investigated the feasibility of using Attention U-Net [24] for semantic segmentation of primary coronal carious lesions using intraoral scan data in children. Fig. 1 outlines the study design schema for evaluating model performance; a cross-sectional diagnostic accuracy study design was used [25]. Fig. 2 outlines the study design schema for benchmarking the machine learning model's predictions against dental practitioners; a cross-sectional diagnostic agreement study design was utilised. As this was an exploratory, proof-of-concept study, no formal sample size calculation was performed. The Royal Children's Hospital Human Research Ethics Committee approved the design of this study (RCH HREC No 88,321). The reporting of this study followed the checklist for using artificial intelligence in dental research [26].

2.2. Data acquisition

Intraoral scans were obtained from participants enrolled in the Melbourne Infant Study: BCG for Allergy and Infection Reduction (MIS BAIR, HREC 33,025) and VASCular Changes Following Infectious Diseases (VASCFIND, HREC 36,009). MIS BAIR is a trial investigating the impact of BCG vaccination on childhood allergy [27]. VASCFIND is a cohort study investigating links between childhood infection and cardiovascular health. Children underwent intraoral scanning (TRIOS 4, 3Shape TRIOS A/S Denmark) aided by commercial software (Dental Desktop v7.4, 3Shape Unite v22.1, 3Shape TRIOS A/S, Denmark). Immediately before scanning dental plaque was removed from all tooth surfaces using Puritan swabs, and the teeth were dried before and during scanning with cotton rolls and gauze. The manufacturer recommended scanning strategy was followed. The TRIOS 4 IOS utilises both white and blue-violet light during scanning to create digital models of the teeth and surrounding structures, which can be visualised with an overlay of approximate natural colours and fluorescence. Study visits occurred at Murdoch Children's Research Institute between January 2021 and March 2022 for MIS BAIR and from July 2021 to November 2023 for VASCFIND.

2.3. Annotation procedures

All digital models generated from the intraoral scanners were viewed on a laptop computer (Alienware, DELL) using a custom non-

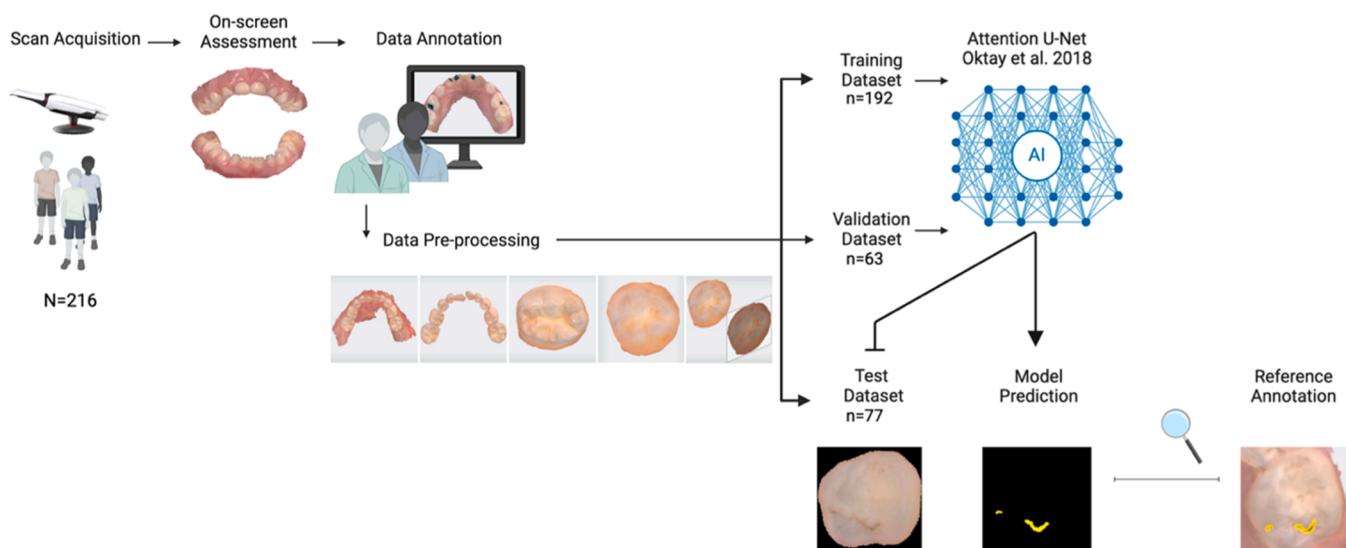


Fig. 1. Study design schema for development of a deep learning model to segment dental caries using intraoral scan data from the MIS BAIR cohort. Left to Right: 216 intraoral scans were collected from the MIS BAIR cohort. Two dental practitioners annotated the scans and data was pre-processed into machine readable format. Teeth with carious lesions were divided into three groups: a training set, used for model development, a validation set for optimising the model during training and an independent test set. The unlabelled test images were inputted into the model, and the model made a prediction mask. The prediction mask was compared to the reference annotation with a variety of metrics. Created in <https://BioRender.com>.

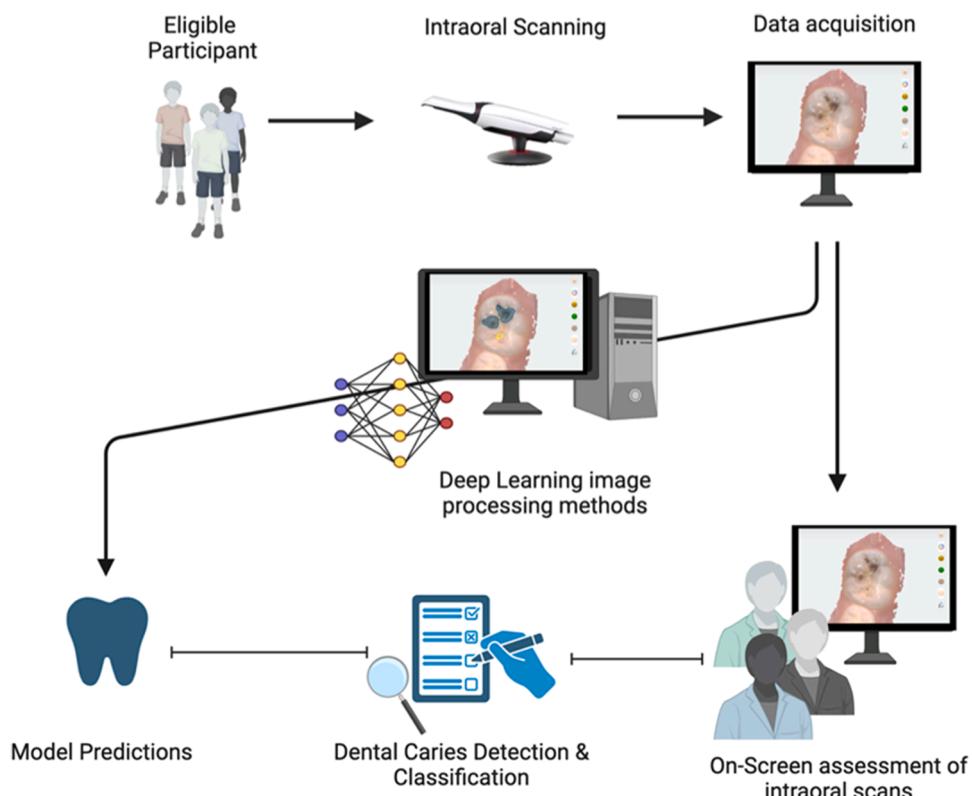


Fig. 2. Study design schema for comparison of machine learning predictions compared to the on-screen assessment of 3D models. Models were examined by three dental practitioners at a single time-point. Created in <https://BioRender.com>.

commercial software with annotation capabilities (internally developed by 3Shape TRIOS A/S). Two trained and calibrated dental practitioners each with >8 years of experience annotated the intraoral scan data.

Before the data labelling process was commenced, the dental practitioners responsible for scan annotation underwent training by an expert examiner to detect and classify dental caries on intraoral scans using the modified International Caries Detection and Assessment

System. After the initial training session, the examiners underwent calibration on a separate intraoral scan dataset, consisting of 200 examination sites across the full spectrum of dental caries which had been confirmed based on histological reference standards. The two examiners repeated the training and calibration procedures until the intra and inter examiner agreement exceeded $\kappa_w > 0.8$.

To create the 'ground truth reference standard' in this study the

calibrated dental practitioners independently undertook on-screen assessment of the digital models supplemented with fluorescence to detect and classify caries. Dental caries was recorded using a previously validated International Caries Detection and Assessment System (ICDAS) where each of the ICDAS scoring criteria had been modified to incorporate fluorescence features [5,28,29]. For this study, the moderate caries classes ICDAS 03 and ICDAS 04 were combined into a single class, resulting in the following dental caries classes: sound tooth surface, initial dental caries, moderate dental caries and extensive dental caries (Table 1). These classes are based on the ICCMS criteria, and represent clinically relevant disease thresholds [13]. Dental caries was recorded using the Research Electronic Data Capture (Vanderbilt University) tool hosted at Murdoch Children's Research Institute [30,31]. There were 21 disagreements in total. A consensus-based approach was used to resolve disagreements for the presence of dental caries per tooth surface.

Annotation for semantic segmentation tasks is the process of assigning a dental caries class label to every pixel in the image. After surfaces with caries were agreed upon, consensus regarding carious lesion/s location within the tooth surface and the dimensions/area of the carious lesion was established amongst the two examiners. This was annotated onto the digital models to assign a dental caries class label to each facet on the meshwork affected by caries using a custom-built software (developed by 3Shape). Initial, moderate and extensive dental caries were annotated with yellow, red and blue segmentation masks respectively. One or multiple lesions could be annotated per tooth (Fig. 3).

2.4. Dataset characteristics

Table 2 provides an overview of the dataset characteristics. The MIS BAIR dataset consisted of upper and lower jaw scans from 216 participants (mean age 5.6 years), with 49.0 % males and 51.0 % females. The VASCFIND dataset consisted of upper and lower jaw scans from 51 participants (mean age 9.7 years), 58.2 % males and 41.8 % females.

Teeth were included for model training and evaluation if they were (1) visible on the scan and (2) had at least one carious lesion. Of the total 4396 teeth in the MIS BAIR dataset, 534 carious lesions were identified, with 332 teeth meeting the inclusion criteria. These were split into three partitions (training, validation, and test sets) to ensure that (1) caries

Table 1

Caries classification and annotation criteria for on-screen assessment of 3D models. The column on the right shows the colour assigned to the prediction mask for each dental caries class.

ICCMS™ Criteria	Description	Corresponding ICDAS Code	Prediction mask colour
Sound	Sound tooth surfaces which show no evidence of visible caries	ICDAS 00	No Colour
Initial	First or distinct visual changes in enamel, with orange-red fluorescence inconsistent with sound enamel	ICDAS 01 or ICDAS 02	Yellow
Moderate	White or brown spot lesion with localised micro-cavitation without visible dentine exposure, with distinct fluorescence, or Dark shadow visible through intact or micro-cavitated enamel, with poorly delineated distinct fluorescence change	ICDAS 03 ICDAS 04	Red
Extensive	A distinct cavity in opaque or discoloured enamel with visible fluorescence changes and exposed dentine	ICDAS 05 or ICDAS 06	Blue

stages were relatively evenly distributed in each set, and (2) all scans from each participant were allocated to a single set. The training set contained 192 carious teeth and was used for model development; the validation set contained 63 carious teeth, and this partition was used for tuning model parameters and monitoring overfitting; and the test set, which included 77 carious teeth, was reserved for independent internal evaluation (see Table 3 for distribution).

For external validation, the VASCFIND dataset was used, comprising 1294 teeth, of which 119 were carious, was used exclusively to evaluate generalisability. This cohort, collected under a distinct study protocol, characterised by a higher proportion of permanent molars (37.8 % vs 3.9 in the MIS BAIR test dataset) provided an independent assessment of the model's performance on unseen and demographically different data (Table 2).

2.5. Data pre-processing

Intraoral scan data is 2.5D, with more depth than a flat 2D plane but less complexity than volumetric 3D data. The intraoral scan data was pre-processed to generate a machine-readable representation suitable for input into the 2D convolutional neural network. First, hard tissues were segmented from soft tissues, and individual teeth were segmented. The separation of teeth from soft tissues and the categorisation of individual teeth were carried out by an instance segmentation model that is available commercially in the TRIOS Patient Monitoring software (Version 2.3 3Shape TRIOS A/S, Denmark). Next, we flattened the 2.5D surface mesh into a 2D planar representation (Fig. 4). The process employed geometric parameterization techniques, drawing on principles of conformal and quasi-conformal mapping, to minimize angular distortion and thereby preserve feature integrity. Conceptually, this involves defining and parameterizing the mesh boundary, then solving partial differential equations to interpolate the mesh, yielding a 2D representation with minimal angle distortion.

The model used for this experiment was the Attention U-Net [24], a variant of U-Net [32] designed for medical image segmentation (Fig. 5). The Attention U-Net architecture is based on the classical U-Net framework which employs an encoder-decoder architecture [32]. The encoder path progressively downsamples the input image to capture semantic and contextual features, resulting in a feature map representation. Next, the decoder path upsamples this feature map, combining it with high-resolution features from the encoder path through skip connections, ultimately generating a pixel-wise segmentation mask [32]. The distinguishing feature of the Attention U-Net architecture is the incorporation of attention mechanisms within the skip connections between the encoder and decoder paths, allowing the model to focus on specific areas.

We used an adaptive moment estimation (ADAM) W6 optimiser [33] with an initial learning rate of $10^{(-4)}$, trained with a batch size of 12. Several image transformations and distortion techniques were implemented to increase the size of the relatively small dataset. The augmentations applied were 1) Horizontal and vertical flips, 2) Elastic transformations, 3) Grid distortion, 4) Affine transformations, 5) Optical distortion, 6) Gaussian blurring, 7) Gaussian noise, 8) Multiplicative noise, 9) Pixel dropout, and 10) Channel dropout. All augmentations were applied to training images during training with a probability of 0.3 and implemented using the Python Augmentations library [34]. The dropout rate was 0.25. The pre-defined stopping condition was to stop training when the loss had not improved on the validation set after 15 epochs. We utilised a class-weighted focal loss due to the imbalanced dataset [35]. Training was implemented using a Pytorch framework [36] on a single Nvidia 4090 Graphics Processing Unit (GPU).

2.6. Model evaluation metrics

The performance of Attention U-Net for pixel-wise segmentation for each dental caries class was calculated using the following metrics from

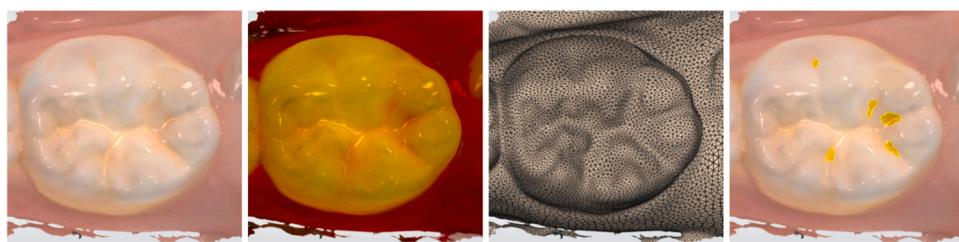


Fig. 3. Example of a tooth with annotations. From left to right, the images show a primary second molar with five identified *initial* caries lesions in white light reflective colours, in fluorescent colours, without any colours and with annotated lesions, respectively. The annotations cover the facets in the mesh that the annotators believe best represents the extent of the lesions.

Table 2

Characteristics amongst children in the study sample. Descriptive statistics are reported as the mean (standard deviation) or frequency (percent). Ethnicity is based on the child's maternal grandmother. Socio-economic Indexes for Areas (SEIFA) ranks areas in Australia according to relative socioeconomic disadvantage. Q1 most disadvantaged Q5 least disadvantaged.

	MIS BAIR	VASC FIND
Eligible Participants	216	56
Age (years)	5.6 (0.4)	9.7 (3.2)
Sex		
Male	106 (49.1)	32 (58.2)
Female	110 (50.9)	23 (41.8)
Ethnicity [†]		
Anglo-Celtic	124 (57.4)	27 (48.2)
European	38 (17.6)	6 (10.7)
Asian	33 (15.3)	5 (8.9)
Middle East	4 (1.9)	3 (5.4)
African	4 (1.9)	2 (3.8)
Aboriginal/ Torres Strait Islander	2 (0.9)	1 (1.8)
Other	11 (5.1)	12 (21.4)
SEIFA [‡]		
Quintile 1	5 (2.31)	1 (1.8)
Quintile 2	31 (14.4)	9 (16.1)
Quintile 3	42 (19.4)	11 (19.6)
Quintile 4	49 (22.7)	16 (28.6)
Quintile 5	89 (41.2)	19 (33.9)
Total Teeth	4396	1294
Participants with caries	113	34
Teeth with at least one carious lesion	332	119
Total carious lesions	534	237

Table 3

Distribution of teeth with carious lesions across data partitions.

	Initial Carious Lesions	Moderate Carious lesions	Extensive Carious Lesions	Total
MIS BAIR Training Set	230	33	30	293
MIS BAIR Validation Set	72	14	17	103
MIS BAIR Test Set	102	16	20	138
VASC FIND Test Set	180	18	39	237

the Python TorchMetrics library [37] well suited to evaluate accuracy in multi-class segmentation tasks:

- Precision: Precision measures the proportion of true positive predictions out of all the positive predictions made by the model: $\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives})$
- Recall: Recall measures the proportion of true positive pixels/lesions correctly identified by the model: $\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False Negatives})$

3. Specificity: Specificity measures the proportion of true negative pixels/lesions correctly identified by the model: $\text{Specificity} = \text{True Negatives} / (\text{True Negatives} + \text{False Positives})$

4. Intersection over Union (IoU): IoU is a metric used to measure the extent of overlap between the predicted segmentation mask (P) and the ground truth mask (M) by calculating the ratio of the intersection area ($P \cap M$) to the total union area ($P \cup M$) of the two masks (Fig. 6). This metric ranges from 0 to 1 (0–100 %), where 0 indicates no overlap, and 1 represents perfectly overlapping segmentation. $\text{IoU} = |P \cap M| / |P \cup M|$

IoU is a useful metric for evaluating the accuracy segmentation models because it considers both the true positive area (intersection) and the false positive and false negative areas (union minus intersection). However, even if the model is correctly localising a carious lesion, the false positive and false negative areas beyond the intersection areas can contribute to lower performance metrics. Therefore, a lesion level analysis was conducted using an evaluation method commonly employed in instance segmentation, where the model's predictions are compared to the ground truth annotations on an instance-by-instance basis (i.e., its ability to correctly localise a carious lesion) rather than using the IoU as a pixel-level accuracy metric. Specifically, (1) the IoU was calculated between the prediction and ground truth reference masks, (2) a detection was considered a True Positive if it exceeded a threshold, and if below this threshold, it was considered a False Positive. Any carious lesions present on the ground truth reference standard but not detected by the model were considered False Negatives. Precision and Sensitivity were then calculated using the OpenCV library [38]. The threshold used was 0.25 [39]. Because manual annotations for small lesions often exhibit variability in defining boundaries, the lower IoU threshold in this proof-of-concept study accounts for this uncertainty and acknowledges plausible detections.

Model performance was evaluated in the MIS BAIR test dataset, and then in the VASC FIND dataset to understand how well the model generalised, when trained on a relatively small dataset.

2.7. Dental clinician comparison

We undertook an additional validation step to answer the question; how well can the model interpret intraoral scans compared to clinicians viewing the same scans in isolation of the clinical data? The model's predictions for each tooth surface in the MIS BAIR test dataset were compared to the on-screen assessment of the intraoral scans supplemented with fluorescence for caries detection and classification by three independent dental practitioners who had been previously trained and calibrated in caries detection intraoral scan data following the aforementioned protocol.

The three dental practitioners with >20 years of combined experience examined the intraoral scans and were asked to classify each tooth surface using the criteria presented in Table 1. The dental practitioners were blinded to the model predictions and each other's assessments. A systematic visual examination protocol was used for on-screen

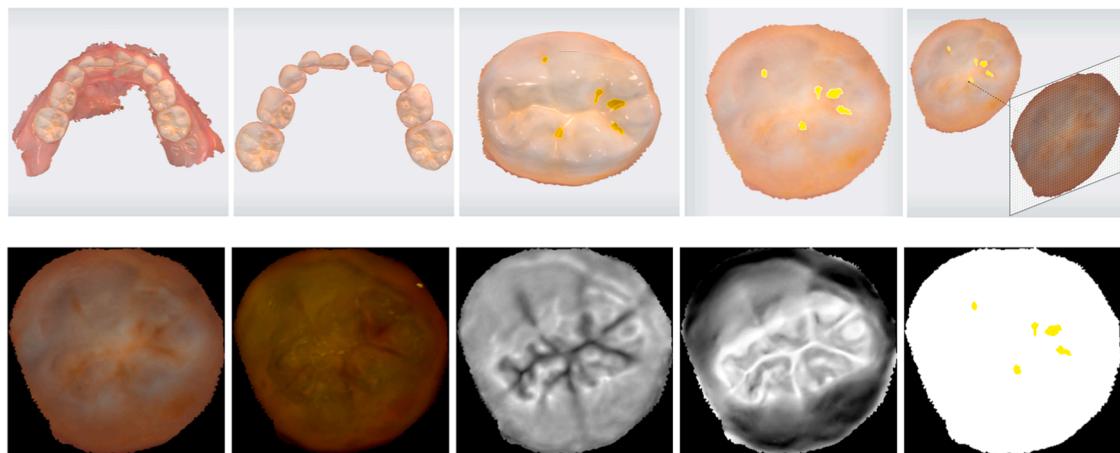


Fig. 4. Data pre-processing. The intraoral scans must undergo a pre-processing pipeline before they can be input to Attention UNet 2D Top row, left to right: We start with a complete jaw scan, the soft tissues are first segmented, then individual teeth meshworks are segmented, the 2.5D meshwork is flattened with minimal mesh distortion. Next, the flattened tooth meshes were projected onto appropriate 2D image planes, recording all relevant data. The projection is carried out with a rasterization rendering routine and the respective 2D data channels are generated. Bottom row, left to right: The pre-processing produces different data channels used as input to the model for each tooth. The resulting data channels were colour (red, green, and blue), fluorescence (red and green), surface curvature, facet normals, as well as an image mask, which indicates the carious pixels, and their corresponding dental caries class. Note the 2D projection approach does not inherently exclude proximal surfaces if visible features (e.g., cavitations, dentine shadows on marginal ridges) are captured, but annotation and model training prioritized features reliably identifiable from scan-derived data alone.

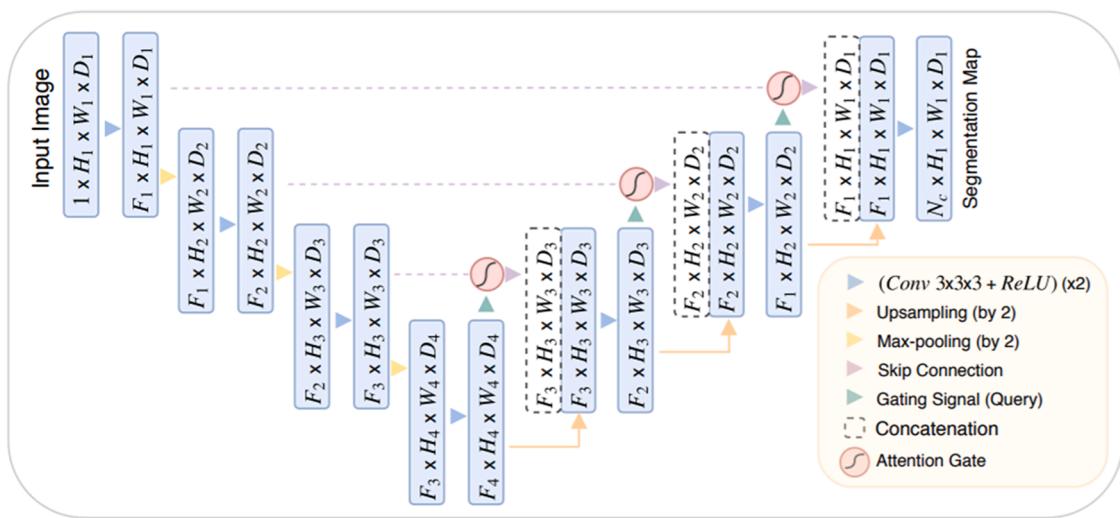


Fig. 5. Attention U-Net architecture. The figure illustrates the architecture of an Attention U-Net model, designed for medical image segmentation. The input image ($H \times W \times D_1$)—where H is the height, W is the width, and D_1 is the number of input channels—is processed through an encoder-decoder structure. The encoder (left) progressively downsamples the image using convolutional layers ($3 \times 3 \times 3$ kernels with ReLU activation) and max-pooling operations, capturing hierarchical feature representations. The bottleneck layer at the bottom extracts abstract features. The decoder (right) reconstructs the segmentation map through upsampling and convolutional layers, incorporating skip connections from the encoder for detailed spatial information. Attention gates (circular nodes) are integrated into the skip connections to filter out irrelevant features by focusing on salient regions, guided by a gating signal from the decoder. The output is a segmentation map ($H \times W \times N$)—where N is the number of segmentation classes—with the same spatial dimensions as the input. Key operations are denoted by arrows: max-pooling (yellow triangles), upsampling (orange arrows), and skip connections (dashed purple lines), with concatenation operations represented by dashed boxes. This architecture combines U-Net’s feature integration with attention mechanisms for enhanced segmentation performance. Published in Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B. 2018. Attention U-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.0999.

assessments of the intraoral scans supplemented with fluorescence. The scans were examined in their postprocessed formats on laptops using the Windows 11 operating system with 15-inch monitors. During the examination, if multiple lesions were present on the same tooth's surface, the caries classification recorded was the most severe clinical presentation.

The Attention U-Net predictions could be visualised on the intraoral scans as a segmentation mask which indicated which facets belonged to a given caries class using the custom software (developed by 3Shape). To

facilitate the direct comparison of the model’s predictions to the dental practitioners’ on-screen assessments of the intraoral scans, the deep learning predictions needed to be consolidated into a single score per tooth surface. This involved aggregating the model’s predictions, which may have identified multiple potential lesion areas on a single surface, into a single surface-level score. The most severe score was recorded if multiple lesions were on one surface. Surfaces were excluded if they were (1) restored (2) proximal (2) partially erupted or (4) predictions involved less than five facets. The latter was considered as the model

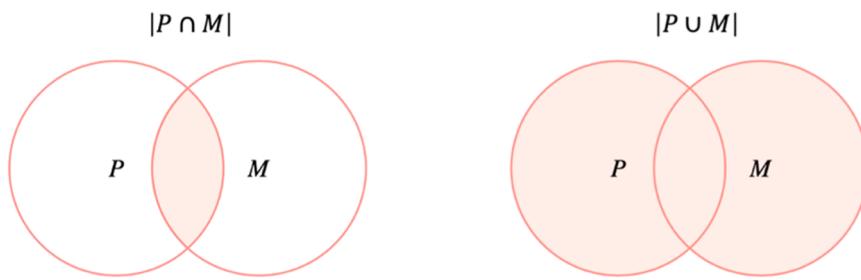


Fig. 6. Intersection over Union. Left: Represents the area of the intersection between the predicted P and ground truth masks M. Right: Represents the area of the union of the predicted and ground truth masks.

lacked a post-processing step, and these would typically be applied to remove such ‘noise’. Additionally, while proximal surfaces with clearly visible features (e.g., cavitations, dentine shadows) were included in model training, they were excluded here because earlier lesions are not visible on intraoral scans, making consistent and reliable agreement assessment across all lesion stages unfeasible.

To evaluate the agreement between the Attention U-Net predictions and the dental practitioners for caries detection and classification for visible tooth surfaces, a binary variable was created to represent each disease threshold. Specifically, caries data were dichotomised at initial ($ICDAS \geq 1$), moderate ($ICDAS \geq 3$), and extensive disease thresholds ($ICDAS \geq 5$). Next, multilevel logistic regression models were used to estimate the likelihood of initial caries detection at the tooth surface level. This approach was selected to account for the within mouth, and within tooth clustering of the dental caries data, which precludes the use of traditional measures of agreement. The likelihood of detecting dental caries between the model and the dental practitioners was estimated by including an indicator variable for the caries detection method as a fixed effect. This analysis was repeated at the moderate and extensive thresholds for the MIS BAIR dataset. The agreement analyses were conducted using R [40].

3. Results

3.1. Model performance

The evaluation metrics for the model’s performance on the MIS BAIR test set and VASCFIND dataset are presented in Table 4.

Attention U-Net had the highest recall and precision for the sound class. In the MIS BAIR test set the recall for initial, moderate, and extensive caries was 44 %, 20 % and 71 %, respectively. For the VASCFIND dataset, it was 29 %, 22 % and 88 %. Attention U-Net classified extensive caries with the highest precision relative to the other classes, 66 % for MIS BAIR and 47 % for VASCFIND. The model could localise extensive caries more accurately than other caries classes, and localisation performance was higher for MIS BAIR (IoU 0.55) than the VASCFIND dataset (IoU 0.45). A sample of the segmentation masks and respective reference annotations is presented (Fig. 7).

Table 4

Attention U-Net performance metrics. Performance metrics of the test groups for Attention U-Net for caries detection and classification by the 4 class ICCMS system and for lesion level classification at the initial disease threshold of MIS BAIR (MB) and VASC (VASC) datasets.

Attention U-Net Performance Metrics	Precision		Recall		Specificity		IoU	
	MB	VASC	MB	VASC	MB	VASC	MB	VASC
Semantic Segmentation 4-Class ICCMS™								
Sound	0.99	0.99	0.98	0.96	0.70	0.81	0.98	0.95
Initial	0.27	0.33	0.44	0.29	0.99	0.99	0.20	0.19
Moderate	0.27	0.09	0.20	0.22	0.99	0.99	0.13	0.07
Extensive	0.66	0.47	0.71	0.88	0.98	0.96	0.55	0.45
Instance Segmentation- Initial disease threshold								
Binary Classification	0.67	0.55	0.73	0.65				

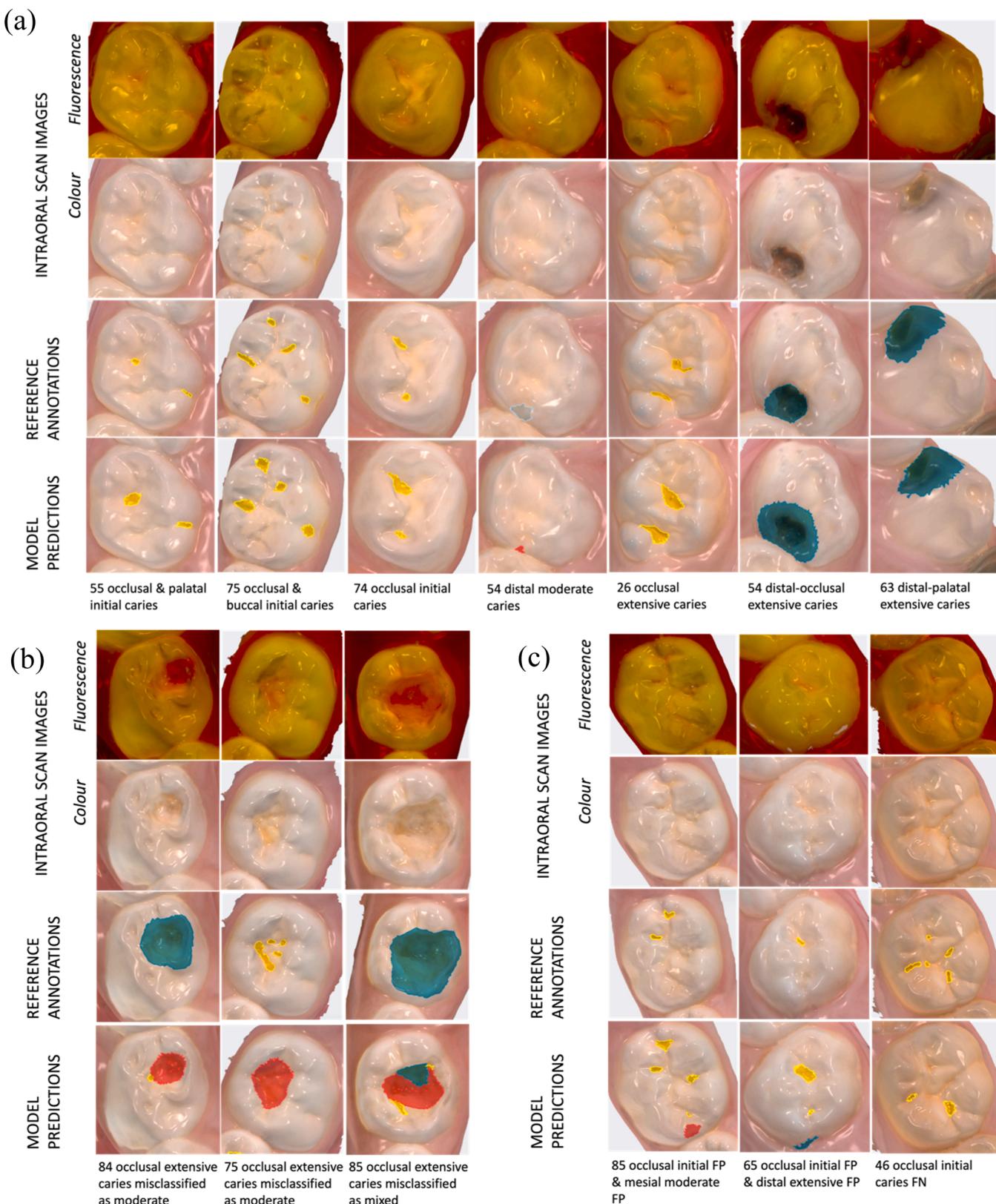


Fig. 7. A sample of the 3D models, the reference annotations, and corresponding model predictions. The teeth are numbered using FDI notation. (a) Shows examples of true positive lesions, with some examples of under segmentation of lesion borders compared to the ground truth reference standard. (b) Shows examples of misclassifications. (c) Shows examples of false positive and negative predictions. False Positive (FP), False Negative (FN).

Table 5

Diagnostic agreement between model predictions and three dental practitioners.

Disease threshold	Model Prediction		Dental Practitioner 1		Dental Practitioner 2		Dental Practitioner 3		Diagnostic Agreement	
	No	Yes	No	Yes	No	Yes	No	Yes	OR (95 % CI)	P
Initial threshold	109	98	113	94	130	77	111	96	0.82 0.60	1.15 0.27
Moderate threshold	173	34	170	37	179	28	175	32	0.9 0.50	1.60 0.71
Extensive threshold	182	25	181	26	184	23	186	21	0.85 0.42	1.71 0.65

detection using intraoral scan data, a previously unreported approach.

Attention U-Net model demonstrated moderate performance on a relatively small dataset. For semantic segmentation tasks, the model demonstrates the highest accuracy for lesions which have cavitated into dentine but poor localisation of initial and moderate caries. The latter may be attributed to the small size of initial lesions and micro-cavitations, with irregular segmentation boundaries, which may be difficult for the model to detect with precision. The poorer performance at this threshold also be a result of the limited diversity in the clinical presentations of dental caries for different tooth types and surfaces, which may have constrained the model to fully learn their morphological variations across the full spectrum of disease. Attention U-Net demonstrated consistently high performance in localising sound tooth structures. However, as sound tooth structure constituted the dominant class, the model likely exhibited a bias toward this category. Regarding generalisability, for semantic segmentation tasks Attention U-Net showed similar trends in the VASCFIND dataset for segmentation, with a tendency towards lower precision estimates. For the instance segmentation task (lesion-level performance) Attention U-Net showed moderate performance for the MIS BAIR dataset, and the higher recall suggests that many instances of caries were detected, though the moderate precision indicates the deep learning model was still prone to false positive indications. In contrast, the VASCFIND dataset demonstrated lower precision and recall, highlighting the model's decreased performance when generalising to a separate dataset. The comparison to dental practitioner results are interesting and suggest that a model with moderate performance for segmentation tasks, may have comparable performance to dental practitioners for dental caries detection and classification. The odds ratio for all disease thresholds included the value 1, suggesting there was no difference between the methods. Evaluating the model at a surface level allowed for positional errors in the location of a carious lesion within a single tooth surface and may partially explain the high agreement. However, the wide confidence intervals around the agreement estimate indicated a high level of uncertainty, and consequently the agreement results should be interpreted with caution.

Currently, no research allows for direct comparison of these results; however, the performance of existing models on intraoral photographs and bitewing radiographs suggests segmentation approaches tend to perform better for detecting advanced carious lesions rather than early disease. A recent systematic review on the application of artificial intelligence methods to photographs for dental caries detection, identified that model performance varied significantly based on lesion severity and computer vision task [41]. A study by Moutselos et al. [42] employed R-CNN with super pixel-based segmentation to detect and classify dental caries across the ICDAS scale using a small dataset of 88 *in vivo* images. Their model performed well for advanced lesions, achieving precision of 0.78, and recall of 0.66, however it had difficulty discriminating initial-stage dental caries (ICDAS 2), achieving a recall of only 0.27 [42]. A study by Yoon et al. [43] also used R-CNN on a substantially larger dataset of 24,578 images, to segment caries on intraoral photographs, and demonstrated a recall of 0.71 for advanced lesions, but did not report on early disease thresholds. A study by Panyarak et al. [44] investigated the application of YOLOv3 in bitewing radiographs for instance segmentation of dental caries across the four ICCMS™ classes (sound, initial, moderate and extensive caries). The model demonstrated inconsistent performance in detecting initial-stage caries on the

enamel's outermost surface and suggested the imbalanced dataset as a factor impacting model performance [44]. A study by Ahmed et al. [45] investigating a U-Net architecture with encoders pre-trained on Image Net, for segmentation of caries on bitewing radiographs, reported better performance for more advanced lesions, but lower IoU for initial caries (range 0.19–0.26). They also reported dataset size as a study limitation possibly contributing to the lower performance at earlier disease thresholds [45]. In overcoming the challenges associated with smaller datasets, Tareq et al. [46] demonstrated the application of YOLO combined with transfer learning using VGG16, can significantly enhance the performance of the model for early caries detection. As an alternate approach, Felsch et al. [18] have described the use of a vision transformer-based model for caries segmentation and achieved pixel-wise IoU scores of 0.630 for non-cavitated caries and 0.692 for dentine cavities, with F1 scores (a metric calculated based on precision and recall) of 0.773 and 0.818, respectively. The model benefited from a large dataset (18,179 images), but precise delineation of early-stage caries remains challenging, reflecting modest pixel-level performance. Collectively, these results highlight the challenges in detecting and segmenting early dental caries using machine learning methods.

Our study exhibited several strengths. It presents a workflow for processing intraoral scan data, starting with entire arch intraoral scans, and segmenting individual teeth into a machine-readable format for 2D deep learning models, finishing with a process to enable prediction masks to be superimposed onto the scans for visualisation. It includes an external validation step for evaluating the performance of the model for segmentation tasks, which is infrequently described in the literature. Additionally, it presents the clinical agreement between the model predictions and three dental practitioners, which enhances the clinical interpretability of the results.

Our study has several limitations that warrant consideration. First, as a proof-of-concept, our dataset did not provide the necessary sample size or distribution of caries across the disease spectrum, anatomical surfaces and tooth types. Second, analysing model performance by tooth surface or type would have been clinically valuable but required a larger, more representative dataset and additional segmentation algorithms beyond the scope of this study. Furthermore, relying solely on intraoral scan data, as in this study, presents a key limitation for specific surfaces; namely, the inherent difficulty in detecting non-cavitated or early proximal carious lesions. Third, the inherent characteristics of dental caries present unique challenges for semantic segmentation. Lesions can be small, morphologically variable, and context-dependent in appearance and location, making them difficult to consistently delineate and reducing segmentation performance [47]. Fourth, our dataset exhibited class imbalance, a common problem in medical segmentation tasks, with higher proportions of the sound (background) class compared to caries classes. This imbalance can introduce bias into the model, favouring the prediction of sound pixels over carious pixels, and result in suboptimal segmentation metrics. Interestingly, despite the relative scarcity of training data for all caries classes compared to the background, the model performed well on the extensive class. This likely reflects the more pronounced visual features of extensive caries, which aid in consistent identification, unlike initial and moderate lesions. Fifth, the reference standard annotation itself, based on expert interpretation of intraoral scans may have introduced bias into this study [48,49]. Although we had two examiners independently assess the dataset at a surface level, and perform annotation of the scan meshwork together,

there was no triangulation with definitive clinical data. As such, the reported performance reflects agreement with scan-based annotations and may not directly translate to accuracy in detecting clinically confirmed caries. Sixth, our comparison to dental practitioner's approach presents additional limitations. Because we lacked access to clinical data, we compared the model's performance to practitioners interpreting intraoral scans in isolation, rather than comprehensive clinical examinations incorporating tactile assessment and radiographic information. Given that our model was trained solely on scan data, comparison against multimodal clinical assessment would be premature at this developmental stage. Our reported metrics therefore reflect scan interpretation accuracy rather than comprehensive diagnostic performance against a clinically accepted standard. Further, our IoU threshold for lesion segmentation may have introduced bias into the lesion-level performance metrics. Also, we did not apply a post-processing step in this study, which influenced the performance metrics. Our model sometimes classifies different pixels within the same lesion into different classes, however clinically we would instead assess the entire lesion based on the most severe characteristics present. Lastly, while the deep learning-based approach to caries detection is promising, it may be fundamentally limited using processed intraoral scan data rather than raw image inputs. Intraoral scanners capture thousands of RGB and fluorescent images across multiple light spectra, but this rich information is substantially reduced through algorithmic processing to generate 3D meshworks and mapped colour approximations. Training the model on these processed outputs likely constrained its ability to detect the subtle visual and textural features characteristic of early caries. Despite employing an Attention U-Net architecture, chosen for its effectiveness in medical image segmentation and capacity to focus on relevant regions, alongside an AdamW optimizer, extensive data augmentation, and a class-weighted focal loss to manage class imbalance, the model performed sub-optimally for initial and moderate lesions. This suggests that the additional transformation of high-dimensional scan data into lower-resolution 2D projections may have further obscured critical discriminatory signals, limiting the model's ability to distinguish early demineralization from sound enamel, even with attention mechanisms in place.

Future research on caries detection using intraoral scan data should diversify and expand dataset size [50]. Triangulating the annotation process against both comprehensive clinical findings and radiographic data, would significantly improve the reference standard quality used for model development. Ultimately, future work could develop multi-modal systems that process both intraoral scan data with near-infrared transillumination images for early proximal caries detection; a key direction to create a tool with greater clinical benefit. For improved model performance, one important step could be the implementation of customised post-processing procedures. This could evaluate all pixels within the lesion and assign a lesion-level classification based on the majority class present, the most severe class detected, or other appropriate thresholds [42]. Beyond post-processing, several algorithmic and architectural enhancements warrant exploration. Transfer learning approaches could be extended beyond standard ImageNet pre-training to include larger, domain-specific dental datasets, or self-supervised learning on unannotated intraoral scan data to build more robust feature representations [51]. 3D U-Net, PointCNN or graph neural networks which have been successful for tooth segmentation using intraoral scan data could be explored [22,23]. Alternative architectures, including Vision Transformers, may also offer improved contextual understanding of subtle features due to their global attention mechanisms [52]. Lastly, given the central approach to caries identification is based on the mapping of colour changes to the tooth surface, a more promising future direction involves training models directly on the raw image data captured by the intraoral scanner. By preserving the full spectral detail and image fidelity of the original images, training on the raw data, could provide representations of caries truer to the actual tooth surface. This approach would not only potentially capture more

subtle presentations but would substantially increase the size of the dataset by including the full stream of captured images, rather than a single processed STL or PLY file.

5. Conclusion

To our knowledge, this is the first study to describe the application of deep learning for the segmentation of dental caries using intraoral scan data. Within the limitations of this proof-of-concept study, the Attention U-Net model, trained on a relatively small dataset, showed encouraging performance for detecting extensive caries cavitated into dentine but performed poorly for early and moderate lesions. These findings underscore the technical and clinical challenges of scan-based AI diagnosis in its current form and highlight critical areas for future refinement.

Disclaimer

The content presented in this work is for research purposes only, it belongs solely to authors and does not reflect the views, positions, or products of 3Shape A/S, or any of its affiliates.

Data availability

The data sets generated during and/or analysed during this study are not publicly available due to consent not obtained from participants for public data sharing but are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Bree Jones: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mathias Lambach:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tong Chen:** Writing – review & editing, Writing – original draft, Formal analysis. **Stavroula Michou:** Writing – review & editing, Resources. **Nicky Kilpatrick:** Writing – review & editing, Supervision. **Nigel Curtis:** Writing – review & editing, Supervision. **David P. Burgner:** Writing – review & editing, Supervision. **Christoph Vannahme:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Mihiri Silva:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Christoph Vannahme reports a relationship with 3 Shape TRIOS A/S at the time that this research was undertaken that includes: employment. Stavroula Michou reports a relationship with 3 Shape TRIS A/S at the time the research was undertaken that includes: employment. Mathias Lambach reports a relationship with 3 Shape TRIS A/S at the time the research was undertaken that includes: Industrial PhD Candidate. Mathia Lambach reports a relationship with Innovation Fund Denmark that includes: funding grants. Mihiri Silva reports a relationship with Murdoch Children's Research Institute that includes: employment and funding grants. David Burgner reports a relationship with Australia National Health and Medical Research Council Investigator Grant that includes: funding grants.

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

Funding

3Shape TRIOS A/S Denmark funds the PhD stipend of author BJ according to an industrial PhD agreement with Murdoch Children's Research Institute. ML was funded by an Innovation Fund Denmark. MS is funded by a Melbourne Children's Clinician-Scientist Fellowship. DB is supported by the Australia National Health and Medical Research Council Investigator Grant. Co-author NC is supported by a National Health and Medical Research Council (NHMRC) Investigator Grant (GNT1197117). The Victorian Government's Operational Infrastructure Support Program supports research at Murdoch Children's Research Institute.

Acknowledgements

We want to thank all families and participants of the MIS BAIR and VASCFIND studies. Thank you to Jenn Copley (JC) for supporting the annotation and scoring process.

References

- [1] W.H. Organization., Global oral health status report: towards universal health coverage for oral health by 2030., (2022).
- [2] D. Serrano-Velasco, A. Martín-Vacas, M.M. Paz-Cortés, G. Giovannini, P. Cintora-López, J.M. Aragoneses, Intraoral scanners in children: evaluation of the patient perception, reliability and reproducibility, and chairside time-a systematic review, *Front. Pediatr.* 11 (2023) 1213072.
- [3] J.R. Ye, S.H. Park, H. Lee, S.J. Hong, Y.K. Chae, K.E. Lee, H.S. Lee, S.C. Choi, O. H. Nam, Influence of limited mouth opening in children on intraoral scanning accuracy: an *in vitro* study, *Int. J. Paediatr. Dent.* 34 (6) (2024) 755–763.
- [4] W.S. Lin, A. Alfaraj, F. Lippert, C.C. Yang, Performance of the caries diagnosis feature of intraoral scanners and near-infrared imaging technology-a narrative review, *J. Prosthodont.* 32 (S2) (2023) 114–124.
- [5] P. Ntovas, S. Michou, A.R. Benetti, A. Bakhshandeh, K. Ekstrand, C. Rahiotis, A. Kakaboura, Occlusal caries detection on 3D models obtained with an intraoral scanner. A validation study, *J. Dent.* 131 (2023) 104457.
- [6] I. Porumb, D.-C. Leucuta, M. Nigoghossian, B. Culic, P.O. Lucaci, C. Culic, I. C. Badea, A.-N. Leghezeu, A.G. Nicoara, M.-R. Simu, Caries lesion assessment using 3D virtual models by examiners with different degrees of clinical experience, *Medicina (B Aires)* 59 (12) (2023) 2157.
- [7] S. Michou, A.R. Benetti, C. Vannahme, P.G. Hermannsson, A. Bakhshandeh, K. R. Ekstrand, Development of a fluorescence-based caries scoring system for an intraoral scanner: an *in vitro* study, *Caries Res.* 54 (4) (2020) 324–335.
- [8] S. Michou, M.S. Lambach, P. Ntovas, A.R. Benetti, A. Bakhshandeh, C. Rahiotis, K. R. Ekstrand, C. Vannahme, Automated caries detection *in vivo* using a 3D intraoral scanner, *Sci. Rep.* 11 (1) (2021) 21276.
- [9] S. Michou, M.S. Lambach, P. Ntovas, A.R. Benetti, A. Bakhshandeh, C. Rahiotis, K. R. Ekstrand, C. Vannahme, Automated caries detection *in vivo* using a 3D intraoral scanner, *Sci. Rep.* 11 (1) (2021).
- [10] B. Jones, T. Chen, S. Michou, N. Kilpatrick, D.P. Burgner, C. Vannahme, M. Silva, Diagnostic agreement between visual examination and an automated scanner system with fluorescence for detecting and classifying occlusal carious lesions in primary teeth, *J. Dent.* 149 (2024) 105279.
- [11] S. Michou, A. Tsakanikou, A. Bakhshandeh, K.R. Ekstrand, C. Rahiotis, A. R. Benetti, Occlusal caries detection and monitoring using a 3D intraoral scanner system. An *in vivo* assessment, *J. Dent.* 143 (2024) 104900.
- [12] N. Schulz-Weidner, M. Gruber, B. Wöstrmann, C.F. Uebereck, N. Krämer, M. A. Schlenz, Occlusal caries detection with intraoral scanners in pediatric dentistry: a comparative clinical study, *J. Clin. Med.* 13 (4) (2024) 925.
- [13] I.A. Pitts NB, Martignon S., et al., ICCMSTM guide for Practitioners and educators., 2014.
- [14] A.I. Ismail, W. Sohn, M. Tellez, A. Amaya, A. Sen, H. Hasson, N.B. Pitts, The international caries detection and assessment system (ICDAS): an integrated system for measuring dental caries, *Commun. Dent. Oral Epidemiol.* 35 (3) (2007) 170–178.
- [15] I.S. Organisation, Information Technology — Artificial intelligence — Artificial intelligence Concepts and Terminology International Standard 22989:2022, Terms and Definitions, International Standards Organisation and International Electrotechnical Commission, 2022, p. 60.
- [16] G.S. Collins, K.G.M. Moons, P. Dhiman, R.D. Riley, A.L. Beam, B. Van Calster, M. Ghassemi, X. Liu, J.B. Reitsma, M. Van Smeden, A.-L. Boulesteix, J. C. Camaradou, L.A. Celi, S. Denaxas, A.K. Denniston, B. Glocker, R.M. Golub, H. Harvey, G. Heinze, M.M. Hoffman, A.P. Kengne, E. Lam, N. Lee, E.W. Loder, L. Maier-Hein, B.A. Mateen, M.D. McCradden, L. Oakden-Rayner, J. Ordish, R. Parnell, S. Rose, K. Singh, L. Wynants, P. Logullo, TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods, *BMJ* (2024) e078378.
- [17] H. Askar, J. Krois, C. Rohrer, S. Mertens, K. Elhennawy, L. Ottolenghi, M. Mazur, S. Paris, F. Schwendicke, Detecting white spot lesions on dental photography using deep learning: a pilot study, *J. Dent.* 107 (2021) 103615.
- [18] M. Felsch, O. Meyer, A. Schlickenrieder, P. Engels, J. Schönewolf, F. Zöllner, R. Heinrich-Weltzien, M. Hesenius, R. Hickel, V. Gruhn, J. Kühnisch, Detection and localization of caries and hypomineralization on dental photographs with a vision transformer model, *NPJ. Digit. Med.* 6 (1) (2023).
- [19] J. Kühnisch, O. Meyer, M. Hesenius, R. Hickel, V. Gruhn, Caries detection on intraoral images using artificial intelligence, *J. Dent. Res.* 101 (2) (2022) 158–165.
- [20] L. Zhang, R. Tanno, M. Xu, Y. Huang, K. Bronik, C. Jin, J. Jacob, Y. Zheng, L. Shao, O. Ciccarelli, F. Barkhof, D.C. Alexander, Learning from multiple annotators for medical image segmentation, *Pattern. Recognit.* 138 (2023). None.
- [21] R. Richert, A. Goujat, L. Venet, G. Viguié, S. Viennot, P. Robinson, J.-C. Farges, M. Fages, M. Ducret, Intraoral scanner technologies: a review to make a successful impression, *J. Heal. Eng.* 2017 (1) (2017) 8427595.
- [22] F.G. Zanjani, A. Pourtaherian, S. Zinger, D.A. Moin, F. Claessen, T. Cherici, S. Parinussa, P.H.N. de With, Mask-MCNet: tooth instance segmentation in 3D point clouds of intra-oral scans, *Neurocomputing* 453 (2021) 286–298.
- [23] S. Vinayahalingam, S. Kempers, J. Schoepf, T.-M.H. Hsu, D.A. Moin, B. van Ginneken, T. Flügge, M. Hanisch, T. Xi, Intra-oral scan segmentation using deep learning, *BMC. Oral Health* 23 (1) (2023) 643.
- [24] O. Oktay, J. Schlemper, LL. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, Attention u-net: learning where to look for the pancreas, *arXiv preprint arXiv:1804.03999* (2018).
- [25] T. Mathes, D. Pieper, An algorithm for the classification of study designs to assess diagnostic, prognostic and predictive test accuracy in systematic reviews, *Syst. Rev.* 8 (1) (2019) 226.
- [26] F. Schwendicke, T. Singh, J.-H. Lee, R. Gaudin, A. Chaurasia, T. Wiegand, S. Uribe, J. Krois, Artificial intelligence in dental research: checklist for authors, reviewers, readers, *J. Dent.* 107 (2021) 103610.
- [27] N.L. Messina, K. Gardiner, S. Donath, K. Flanagan, A.-L. Ponsonby, F. Shann, R. Robins-Browne, B. Freyne, V. Abruzzo, C. Morison, L. Cox, S. Germano, C. Zufferey, P. Zimmermann, K.J. Allen, P. Vuillermin, M. South, D. Casalaz, N. Curtis, Study protocol for the Melbourne Infant Study: BCG for Allergy and Infection Reduction (MIS BAIR), a randomised controlled trial to determine the non-specific effects of neonatal BCG vaccination in a low-mortality setting, *BMJ Open*. 9 (12) (2019) e032844.
- [28] G. Sá, S. Michou, M. Bönecker, F. Mendes, B. Amarante, K. Ekstrand, Diagnostic validity of ICDAS clinical criteria on digital 3D models, *J. Dent.* 149 (2024) 105274.
- [29] A. Ferreira Zandonà, E. Santiago, G. Eckert, M. Fontana, M. Ando, D.T. Zero, Use of ICDAS combined with quantitative light-induced fluorescence as a caries detection method, *Caries Res.* 44 (3) (2010) 317–322.
- [30] P.A. Harris, R. Taylor, B.L. Minor, V. Elliott, M. Fernandez, L. O'Neal, L. McLeod, G. Delacqua, F. Delacqua, J. Kirby, S.N. Duda, The REDCap consortium: building an international community of software platform partners, *J. Biomed. Inf.* 95 (2019) 103208.
- [31] P.A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, J.G. Conde, Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support, *J. Biomed. Inf.* 42 (2) (2009) 377–381.
- [32] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, Springer*, 2015, pp. 234–241. October 5–9, 2015, proceedings, part III 18.
- [33] P.K. Diederik, Adam: a method for stochastic optimization, (No Title) (2014).
- [34] A. Buslaev, V.I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, A.A. Kalinin, Albumentations: fast and flexible image augmentations, *Information* 11 (2) (2020) 125.
- [35] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2999–3007.
- [36] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, PyTorch: an imperative style, high-performance deep learning library, *ArXiv*. (2019) arXiv preprint arXiv:1912.01703 10 (1912).
- [37] N.S. Detlefsen, J. Borovec, J. Schlock, A.H. Jha, T. Koker, L. Di Liello, D. Stanci, C. Quan, M. Grechkin, W. Falcon, Torchmetrics-measuring reproducibility in pytorch, *J. Open. Source Softw.* 7 (70) (2022) 4101.
- [38] G. Bradski, A. Kaehler, *Learning OpenCV: Computer vision With the OpenCV Library*, O'Reilly Media, Inc, 2008.
- [39] A. Sonavane, R. Kohar, Dental cavity detection using yolo, in: *Proceedings of Data Analytics and Management: ICDAM 2021 2*, Springer, 2022, pp. 141–152.
- [40] R Development Core Team, R: a language and environment for statistical computing, R foundation for statistical computing, Vienna, Austria., 2023.
- [41] M. Moharrami, J. Farmer, S. Singhal, E. Watson, M. Glogauer, A.E.W. Johnson, F. Schwendicke, C. Quinonez, Detecting dental caries on oral photographs using artificial intelligence: a systematic review, *Oral Dis.* 30 (4) (2024) 1765–1783.
- [42] K. Moutslos, E. Berdouses, C. Oulis, I. Maglogiannis, Recognizing occlusal caries in dental intraoral images using deep learning, in: *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, IEEE, 2019, pp. 1617–1620.
- [43] K. Yoon, H.M. Jeong, J.W. Kim, J.H. Park, J. Choi, AI-based dental caries and tooth number detection in intraoral photos: model development and performance evaluation, *J. Dent.* 141 (2024) 104821.
- [44] W. Panyarak, W. Suttipak, K. Wantanajittikul, A. Charuakkra, S. Papayatasok, Assessment of YOLOv3 for caries detection in bitewing radiographs based on the ICCMS™ radiographic scoring system, *Clin. Oral. Investig.* 27 (4) (2023) 1731–1742.

- [45] W.M. Ahmed, A.A. Azhari, K.A. Fawaz, H.M. Ahmed, Z.M. Alsadah, A. Majumdar, R.M. Carvalho, Artificial intelligence in the detection and classification of dental caries, *J. Prosthet. Dent.* (2023).
- [46] A. Tareq, M.I. Faisal, M.S. Islam, N.S. Rafa, T. Chowdhury, S. Ahmed, T.H. Farook, N. Mohammed, J. Dudley, Visual diagnostics of dental caries through deep learning of non-standardised photographs using a hybrid YOLO ensemble and transfer learning model, *Int. J. Env. Res. Public Health* 20 (7) (2023) 5351.
- [47] D. Müller, I. Soto-Rey, F. Kramer, Towards a guideline for evaluation metrics in medical image segmentation, *BMC. Res. Notes.* 15 (1) (2022) 210.
- [48] L.T. Arsiwala-Scheppach, A. Chaurasia, A. Mueller, J. Krois, F. Schwendicke, Machine learning in dentistry: a scoping review, *J. Clin. Med.* 12 (3) (2023) 937.
- [49] T. Walsh, R. Macey, P. Riley, A.M. Glenny, F. Schwendicke, H.V. Worthington, J. E. Clarkson, D. Ricketts, T.L. Su, A. Sengupta, Imaging modalities to inform the detection and diagnosis of early caries, *Cochrane Database Syst. Rev.* (3) (2021).
- [50] G. Varoquaux, V. Cheplygina, Machine learning for medical imaging: methodological failures and recommendations for the future, *NPJ. Digit. Med.* 5 (1) (2022).
- [51] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, U. Naseem, Y. Gu, Transfer or self-supervised? Bridging the performance gap in medical imaging, 2024.
- [52] Q. Pu, Z. Xi, S. Yin, Z. Zhao, L. Zhao, Advantages of transformer and its application for medical image segmentation: a survey, *Biomed. Eng. Online* 23 (1) (2024) 14.