





## Article

# One-Stage Methods of Computer Vision Object Detection to Classify Carious Lesions from Smartphone Imaging

S. M. Siamus Salahin <sup>1</sup>, M. D. Shefat Ullaa <sup>1</sup>, Saif Ahmed <sup>1</sup>, Nabeel Mohammed <sup>1</sup>, Taseef Hasan Farook <sup>2,\*</sup> and James Dudley <sup>2</sup>

<sup>1</sup> Electrical and Computer Engineering, North South University, Dhaka 1229, Bangladesh; saif.ahmed02@northsouth.edu (S.A.)

<sup>2</sup> Adelaide Dental School, The University of Adelaide, Adelaide 5005, Australia

\* Correspondence: taseef.farook@adelaide.edu.au

**Abstract:** The current study aimed to implement and validate an automation system to detect carious lesions from smartphone images using different one-stage deep learning techniques. 233 images of carious lesions were captured using a smartphone camera system at  $1432 \times 1375$  pixels, then classified and screened according to a visual caries classification index. Following data augmentation, the YOLO v5 model for object detection was used. After training the model with 1452 images at  $640 \times 588$  pixel resolution, which included the ones that were created via image augmentation, a discrimination experiment was performed. Diagnostic indicators such as true positive, true negative, false positive, false negative, and mean average precision were used to analyze object detection performance and segmentation of systems. YOLO v5X and YOLO v5M models achieved superior performance over the other models on the same dataset. YOLO v5M's mAP was 0.727, precision was 0.731, and recall was 0.729, which was higher than other models of YOLO v5, which generated 64% accuracy, with YOLO v5M producing slightly inferior results. Overall mAPs of 0.70, precision of 0.712, and recall of 0.708 were achieved. Object detection through the current YOLO models was able to successfully extract and classify regions of carious lesions from smartphone photographs of in vitro tooth specimens with reasonable accuracy. YOLO v5M was better fit to detect carious microcavitations while YOLO v5X was able to detect carious changes without cavitation. No single model was capable of adequately diagnosing all classifications of carious lesions.

**Keywords:** caries; smartphone; clinical photography; YOLO; object detection



**Citation:** Salahin, S.M.S.; Ullaa, M.D.S.; Ahmed, S.; Mohammed, N.; Farook, T.H.; Dudley, J. One-Stage Methods of Computer Vision Object Detection to Classify Carious Lesions from Smartphone Imaging. *Oral* **2023**, *3*, 176–190. <https://doi.org/10.3390/oral3020016>

Academic Editors: Oana Almasan, Smaranda Buduru and Yingchu Lin

Received: 7 February 2023

Revised: 17 March 2023

Accepted: 24 March 2023

Published: 4 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Dental caries is the bacterial demineralization and destruction of hard tooth tissue. Although non-fatal, the disease is highly prevalent and often remain asymptomatic until significant destruction occurs that leads to physical and financially debilitating consequences for the patient. Therefore, early detection of dental caries as a management strategy is emphasized to achieve the best outcomes in minimally invasive dentistry [1]. Automated detection of such lesions with appropriate early intervention can reduce disease complications and burden costs. The International Caries Detection and Assessment System (ICDAS) has been established as a globally reliable standard for diagnosing caries on an epidemiological scale, with several visual classification systems created for deep learning purposes in image-based cariology [2,3]. In the field of cariology, microphotography is a popular tool to study carious remineralization and oral histopathology [4,5]. The images obtained can also be repurposed for deep learning and computer vision applications.

Integration of deep learning with a computer-aided clinical decision support system (CDSS) has also become very popular with generational leaps in computing power and imaging technology [6]. CDSS systems are capable of analyzing image data to develop prediction models to match relationships between input and output characteristics of clinical photographs and radiomics [7,8].

Applications for target detection have grown recently [9]. Researchers have proposed various image processing techniques and machine learning algorithms for automated object detection from color photographs and radiographic images. Techniques range from the implementation of multiscale fusion deep neural networks for pavement cracks, similar to dental micro-cavitation [10], automated unmanned aerial vehicle (UAV) sensing [11], the application of deep learning to detect potholes on roads similar to visual caries diagnostics [12], and the application of computer vision models such as YOLO in multilayer feature cross-layer fusion networks [13]. However, the role of smartphone microphotography in caries detection within the space of ‘computer assisted object detection’ or computer vision is largely unexplored, with some investigations documenting the trial of closed-source deep learning algorithms for patient-facing tele-dentistry applications [14]. The existing methods of computer vision application from clinical photographs can be grouped into either (a) conventional image processing that performs selective operations to extract important information from images through machine learning-based approaches or (b) staged object detection. One-stage object detection methods have no ‘region proposal’ stage, while two-stage deep learning object detection algorithms employ both ‘region proposal’ and ‘classification regression’ for the purpose of image classification.

One-stage methods are fast, end-to-end models that produce the final result directly using intensive sampling [15]. One-stage methods use ‘sales and aspect ratios’ when extracting features using a convolutional neural network (CNN) and are represented by YOLO (You Only Look Once) and SSD (single shot detector). YOLOv1 is the pioneering work of the one-stage method of target detection, which was first introduced in 2016 [16]. Two-stage methods first implement a selective search before applying a CNN to produce several sparse candidate boxes, which are subsequently classified and regressed. Two-stage methods are represented by a region-based convolutional neural network (R-CNN) and, while very precise, are slower than one-stage methods that are preferred for real-time object detection tasks such as those present in clinical diagnostics [17].

Researchers have proposed various image processing techniques and machine learning algorithms for the automated or semi-supervised detection of dental caries from colored photographs and radiomic imaging. However, previous studies did not explore the possibilities of smartphone microphotography in the deep learning of visual caries assessment. It was also noted that most of the currently published research in cariology utilized standard image processing and machine learning approaches with limited published evidence on the application of one-stage and two-stage deep learning object detection algorithms.

The current study aimed to develop a caries detection system by validating a deep CNN model using different YOLOv5 object detection algorithms to automate the detection of dental caries from smartphone images. The goal of the experiment was to provide a detailed analysis and comparison of the different models to support researchers and practitioners in determining the best deep learning model for a given task of caries diagnostics. Each model was repurposed to classify the visual extent of dental decay in a given image and localize the findings with bounding boxes based on a customized visual analogue of the ICDAS caries classification system. The following three classes were implemented: “Visible change without cavitation”, “Visible change with microcavitation”, and “Visible change with cavitation”. The different models of YOLOv5 were then evaluated using diagnostic parameters such as sensitivity, specificity, accuracy, precision, recall, and mean average precision (mAP).

## 2. Materials and Methods

### 2.1. Reporting Protocols

The study was developed by adhering to the standards for reporting of diagnostic accuracy (STARD) 2015 [18] and minimum information for clinical artificial intelligence modeling (MI-CLAIM) [19] guidelines.

## 2.2. Ethics

The study was deemed ‘negligible risk’ according to the relevant ethics committees and was therefore exempt from ethical review.

## 2.3. Data Acquisition and Annotation

Human anterior teeth with visible smooth surface caries were collected from a source of deidentified specimens. Molars were excluded to reduce variations such as shadow casts, altered translucency, occult pit and fissure lesions, and occlusal surface morphology that might negatively affect the algorithm training process. A 60× standard peripheral, detachable optical zoom lens with light-emitting diode (LED) self-illumination (No. 9595; Yegren Optics Inc., Anyang, China) was attached to the primary camera (12 MP, f/1.8, 26 mm, 1/1.76", 1.8 µm, Dual Pixel Phase Detection Autofocus, optical image stabilization) of a smartphone (S22 5G; Samsung Electronics, Seoul, Korea), and images of cariogenic activity were captured. The images were imported into Adobe Photoshop 2019 as Joint Photographic Experts Group (JPEG) files, and regions of interest (ROI) were focused and cropped into an ellipsoid shape of 20 × 20 mm dimension. The ROI were loaded into the open-source Python project “LabelImg.py” [20], and annotations were carried out by an operator with three years of clinical experience. Labeling was performed according to the global caries classification standard called the ICDAS index, where images were classified into the following categories: “Visible change without cavitation”, “Visible change with microcavitation” and “Visible change with cavitation”.

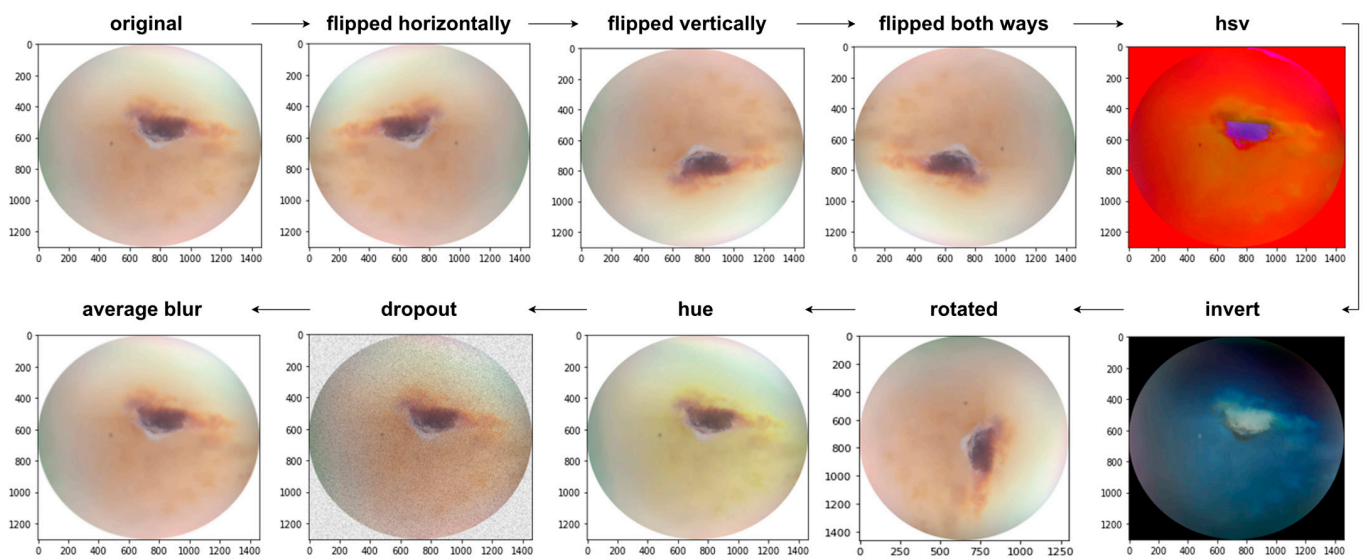
Initially, 20 randomly selected teeth were physically provided to three experts who volunteered to label the teeth according to the classification index provided. Upon conducting an inter-class correlation, Chronbach’s analysis was set at  $\alpha = 0.958$ ,  $r = 0.89 \pm 0.06$ . An operator with at least 3 years of experience, who could produce  $r > 0.8$  against the labels performed by each individual expert within an inter-rater correlation matrix was recruited. In addition, the operator had to demonstrate an intra-rater correlation of  $r > 0.90$  on the same dataset on two individual sessions spaced over one week. The operator recruited following the screening process demonstrated a correlation of 0.90, 0.90, and 0.91, respectively, within the matrix. Intra-rater reliability demonstrated a correlation of  $r = 0.96$  and  $\alpha = 0.98$ .

## 2.4. Training Strategies and Augmentation

The dataset consisted of 233 images of smooth surface caries, where 68 images were discarded for not falling into the classification categories. The dataset was divided into training and validation subsets in the ratio 8:2, ensuring that all images existed in only one of the two subsets. Image augmentation was then performed on the training dataset and outlined in Table 1. Data augmentation refers to the creation of additional data points from the existing dataset to artificially increase the amount of data available for training [21]. Following sequential augmentation (Figure 1) [22], 1452 images were generated for training.

**Table 1.** Image augmentation techniques used.

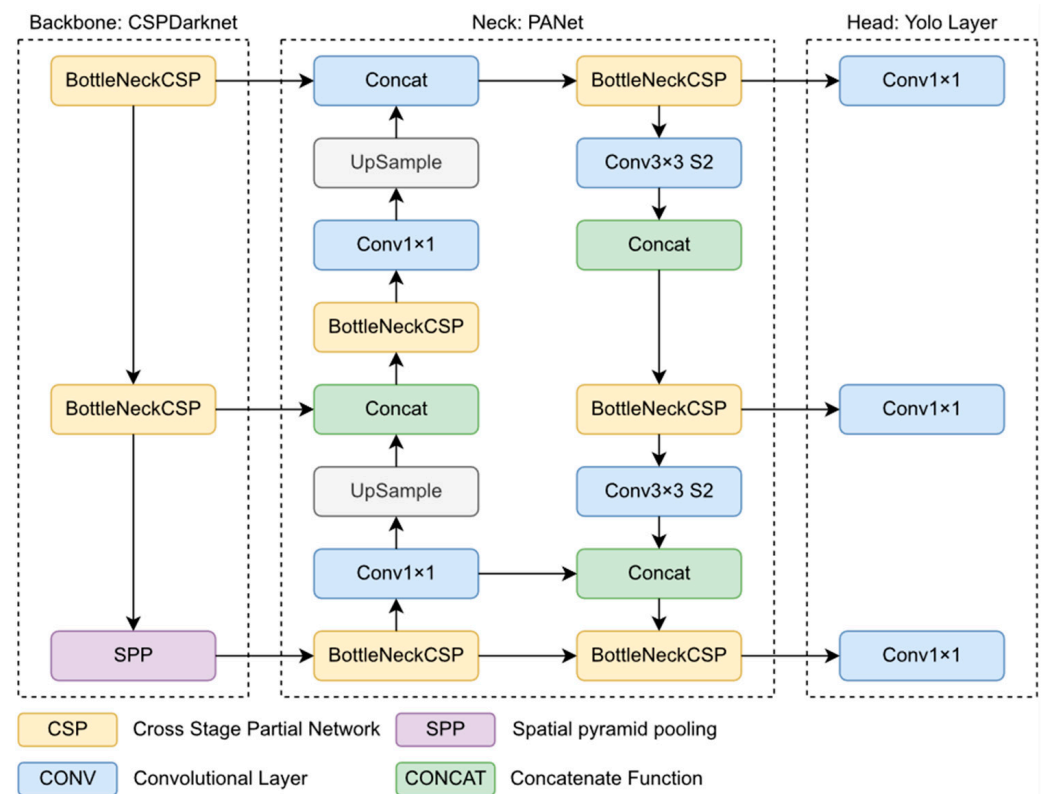
Augmentation Techniques	Description
Flipped horizontally	Reverses the order of the elements in each row
Flipped vertically	Reverses the order of the elements in each column
Flipped both ways	Reverses the order of the elements in both row and column
HSV	Changes the color space from RGB to HSV
Average Blur	Smoothens the image using an average filter.
Dropout	Randomly sets input elements to zero with a given probability
HUE	Raises the hue value
Rotated	Rotated 90 degree clockwise
Invert	Inverts all values in images, i.e., sets a pixel from value v to 255-v



**Figure 1.** Image augmentation techniques implemented.

### 2.5. The Object Detection Model

The structure of YOLOv5 [16] is shown in Figure 2 which comprises of the Backbone, Neck, and Head. The network architecture consisted mainly of three parts: (1) Backbone: Cross Stage Partial (CSP) Darknet; (2) Neck: path aggregation network (PANet); and (3) Head: YOLO Layer. The data input is first channeled to CSPDarknet to extract the features and then fed to PANet for feature fusion. Finally, Yolo Layer outputs detection results (class, score, location, and size).



**Figure 2.** YOLO architecture applied within the current study.

Justifications for the selection of the YOLOv5 architecture are as follows: First, YOLOv5 includes CSPNet [23] into Darknet, constructing CSPDarknet as its backbone. CSPNet resolves the issues with recurring gradient information in large-scale backbones by incorporating the gradient changes into the feature map, thereby decreasing the FLOPS (floating-point operations per second) and the number of parameters of the model, ensuring the inference speed and accuracy while reducing the model size. Such an approach can be considered valuable for caries detection where subsurface cavitation and pit and fissure cavities possess varying hues for the same extent of lesion while reflecting light in different ways.

In a task of caries detection, detection speed and accuracy are essential, and a compact model size also determines its inference efficiency on edge devices. Edge devices carry out processing, filtration, and storage of data passing between networks with limited resources. To improve information flow, the YOLOv5 applied PANet [24] as its neck. With an improved bottom-up path, PANet adopts a new feature pyramid network (FPN) structure, enhancing low-level features' propagation. The feature grid and all feature levels are connected by adaptive feature pooling, which is also used to propagate important information from each feature level to the next subnetwork. In lower layers, PANet improves the use of accurate localization signals, which can increase the object's location accuracy. In order to achieve multi-scale [25] prediction, the head of YOLOv5 generates feature maps in three different sizes ( $18 \times 18$ ,  $36 \times 36$ , and  $72 \times 72$ ), allowing the model to handle small, medium, and large objects. Multi-scale detection confirms that the model can track size changes in tooth decay detection. In other words, the PANet in YOLOv5 was improved to learn and distinguish from smaller features in the image, while the head structure of YOLOv5 allowed objects of different sizes to be tracked.

Figure 3 demonstrates the application of the YOLO system to smartphone images in the current study design. The primary difference was the trade-off between the model's actual file size and inference time when processing commands.

## 2.6. Evaluation Metrics

The mean average precision (mAP) accuracy, sensitivity, specificity, precision, intersection over union (IoU) (Figure 4A), and recall were used to evaluate the performance of the object detection model on carious lesions. (Figure 4B) The mean average precision or mAP, is a comprehensive metric that accounts for the precision, and recall of the predicted bounding boxes. It considers both the localization accuracy and overall performance of the model in terms of detecting all objects in an image [26].

The MI-CLAIM model dictates that all definitions of the evaluation metrics be stated. Precision refers to the proportion of accurately classified positive data (TP) in the deep learning dataset to the total number of correctly classified data. Recall referred to correctness in classifying all positive data.

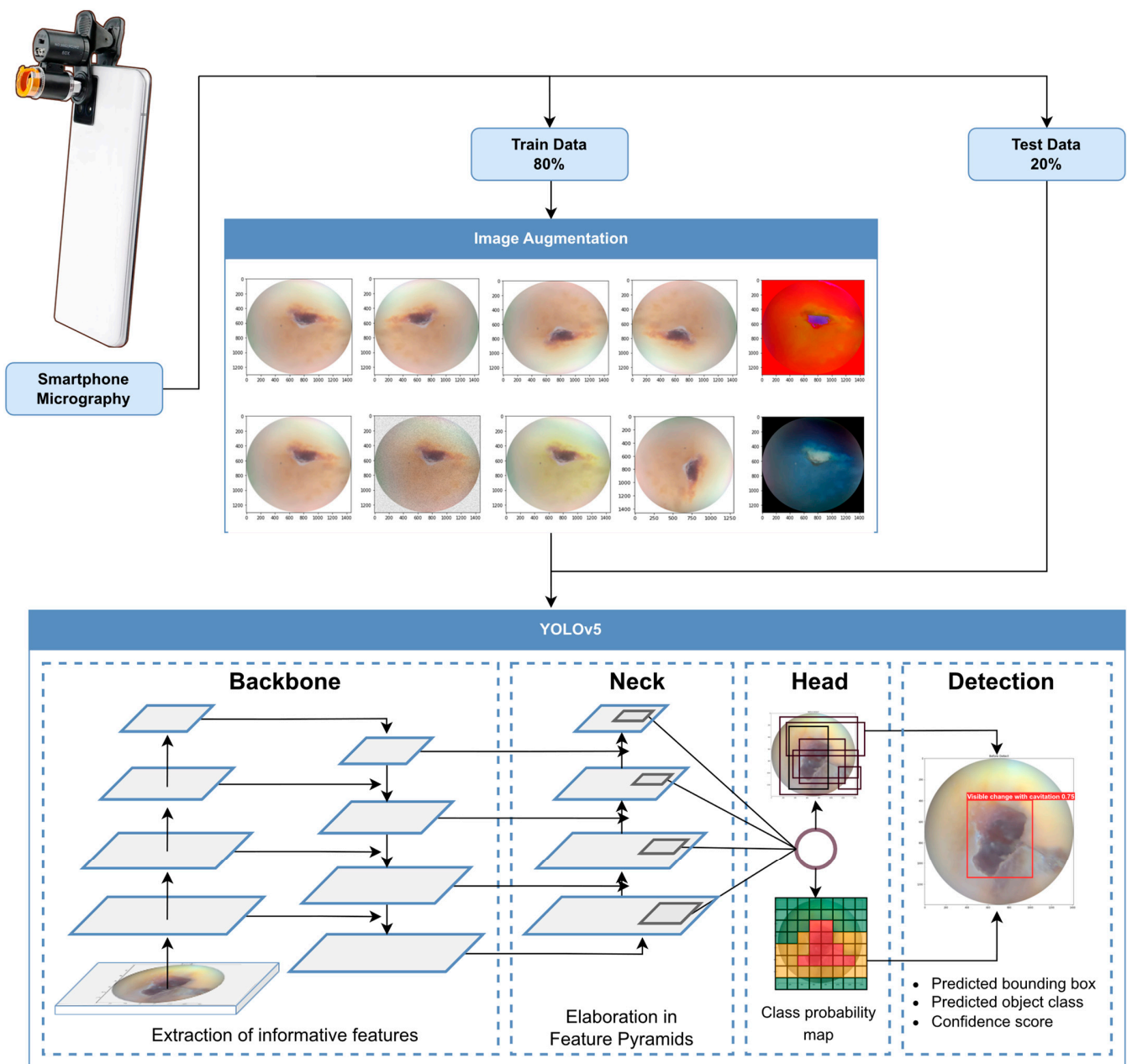
True Positive (TP) referred to the model's accuracy in predicting positive classes. True Negative (TN) referred to the model's accuracy in predicting negative classes. False Positive (FP) was when the model incorrectly predicted a positive class. Finally, False Negative (FN) was when the model incorrectly predicted negative class.

The mean average precision (mAP) function is commonly used to analyze object detection performance of segmentation systems such as YOLOv5, Faster R-CNN, and MobileNet SSD. The mAP generates a score by comparing the detected box to the ground-truth bounding box. The higher the score, the greater the accuracy of the models. It is calculated by finding the average precision (AP) for each class and then averaging over the number of classes.

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (1)$$

The predictions were made by object detection systems using a bounding box and a class label. For each bounding box, the overlap between the predicted bounding box and the ground truth bounding box was calculated using IoU.





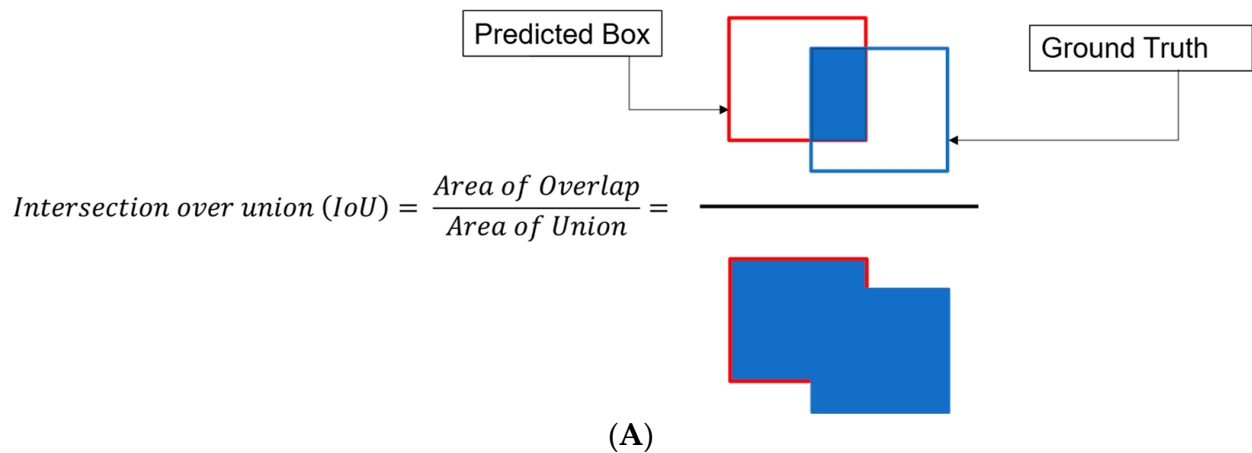
**Figure 3.** Caries detection via YOLO architecture.

The precision and recall were measured utilizing IoU values against a given threshold value. For example, for an IoU threshold value of 0.5, an IoU reading of 0.7 was classified as True Positive while a reading of 0.3 was classified as False Positive (FP).

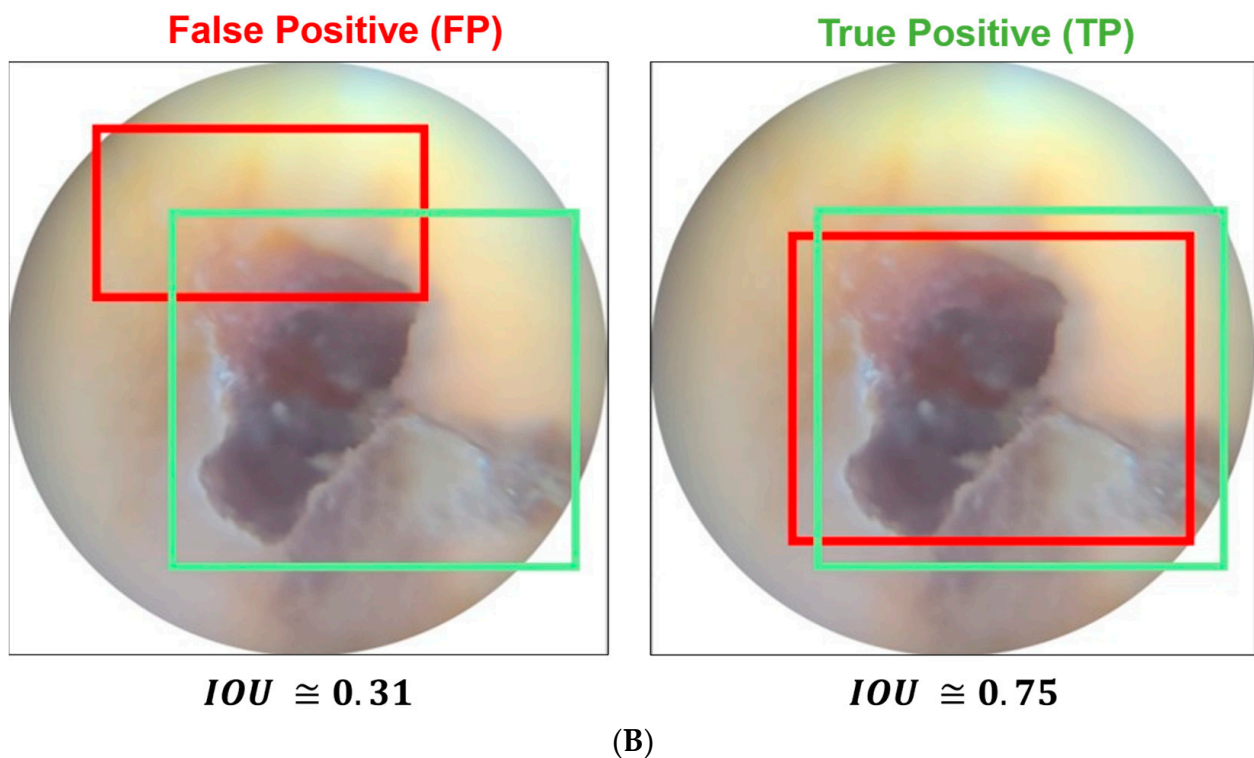
### 2.7. Evaluation Settings

The learning rate was established at 0.01 to speed up model convergence when the loss function produced by the model was lower, and stochastic gradient descent (SGD), the optimizer that reduces the computational load, was chosen for hyperparameter optimization. The learning rate momentum, responsible for creating a learning step size based on the loss function, was set to  $\approx 0.95$  due to the small number of samples in the tooth decay micrography data set. Due to the relatively smaller datasets (165 images primarily after augmentation, after which 1452 images were ready for training), all major pre-trained models of the YOLOv5 family were applied for training to determine which model resulted in the best classification and localization. The hyper-parameter specifications were as

follows: batch size = 10, image resolution =  $640 \times 588$  pixels, 30 epochs, and learning rate =  $1 \times 10^{-2}$ . All codes were structured following the PEP-8 guidelines.



**IF IOU Threshold = 0.5**



**Figure 4.** (A) Application of intersection over union as a metric measurement. (B) Application over union applied for caries detection in precision and recall.

### 3. Results

The current study validated the application of each member of the YOLO v5 family in classifying caries from smartphone microphotography. Tables 2 and 3 display the final model evaluation results. Although YOLOv5M and YOLOv5X achieved almost similar results in terms of accuracy, YOLOv5X achieved a 3.07% higher mAP than YOLOv5M. The other models did not perform well in terms of accuracy, recall, and mAP. Figure 5 shows the precision-recall (PR) curve for each category at different thresholds. The variations in performance of the models were attributed to the differences in filter counts and parameters.

The input images of carious lesions that were processed using YOLO v5X are shown in Figure 6A and the prediction outputs are displayed in Figure 6B. The location of most

carious tooth decay was seen to fit the original site of tooth decay in physical form and was reflected upon by the quantitative diagnostic metrics that were subsequently evaluated (Table 3).

Figure 7 demonstrates the changes in tooth decay detection performance metrics for the five different network structures. (A), (B), and (C) represent recall, precision, and mAP@0.5 (mean average precision at 0.5 intersection over union) curves, respectively, trained up to 30 epochs. The usage of 30 epochs was justified after trial and error, as the YOLO model's performance metrics stopped improving substantially after 25 epochs. YOLO v5M and v5X were the overall highest performers at the end of the training cycles.

**Table 2.** Performance evaluation of various yolo models during training-validation phase.

Model	Classification	TP	TN	FP	FN	SN	SP	AC	T.AC
YOLO v5S	Visible change without cavitation	0.41	0.48	0.59	0.52	0.44	0.44	0.44	0.59
	Visible change with microcavitation	0.69	0.28	0.31	0.72	0.48	0.47	0.48	
	Visible change with cavitation	0.75	1	0.25	0	1	0.80	0.87	
YOLO v5M	Visible change without cavitation	0.55	0.38	0.45	0.62	0.47	0.45	0.46	0.65
	Visible change with microcavitation	0.65	0.36	0.35	0.64	0.64	0.64	0.50	
	Visible change with cavitation	1	1	0	0	1	1	1	
YOLO v5L	Visible change without cavitation	0.23	0.88	0.77	0.12	0.65	0.53	0.55	0.54
	Visible change with microcavitation	0.69	0	0.31	1	0.40	0	0.34	
	Visible change with cavitation	0.50	1	0.50	0	1	0.66	0.75	
YOLO v5X	Visible change without cavitation	0.41	0.81	0.59	0.19	0.68	0.57	0.61	0.64
	Visible change with microcavitation	0.62	0	0.38	1	0.38	0	0.31	
	Visible change with cavitation	1	1	0	0	1	1	1	
YOLO v5N	Visible change without cavitation	0.32	0.71	0.68	0.29	0.52	0.51	0.51	0.63
	Visible change with microcavitation	0.75	0.24	0.25	0.76	0.49	0.48	0.49	
	Visible change with cavitation	1	0.78	0	0.22	0.81	0.78	0.89	

**Table 3.** Performance evaluation of various yolo models during the testing phase.

Model	Classification	Precision	Recall	mAP@0.5
YOLO v5S	Visible change without cavitation	0.453	0.455	0.303
	Visible change with microcavitation	0.606	0.688	0.75
	Visible change with cavitation	0.797	0.984	0.895
	Overall	0.619	0.709	0.649
YOLO v5M	Visible change without cavitation	0.687	0.5	0.531
	Visible change with microcavitation	0.56	0.625	0.588
	Visible change with cavitation	0.887	1	0.995
	Overall	0.712	0.708	0.705



Table 3. Cont.

Model	Classification	Precision	Recall	mAP@0.5
YOLO v5L	Visible change without cavitation	0.598	0.542	0.465
	Visible change with microcavitation	0.667	0.75	0.712
	Visible change with cavitation	0.663	0.75	0.87
	Overall	0.643	0.681	0.682
YOLO v5X	Visible change without cavitation	0.611	0.5	0.528
	Visible change with microcavitation	0.677	0.688	0.657
	Visible change with cavitation	0.904	1	0.995
	Overall	0.731	0.729	0.727
YOLOv5N	Visible change without cavitation	0.545	0.273	0.367
	Visible change with microcavitation	0.698	0.723	0.716
	Visible change with cavitation	0.659	1	0.845
	Overall	0.634	0.665	0.643

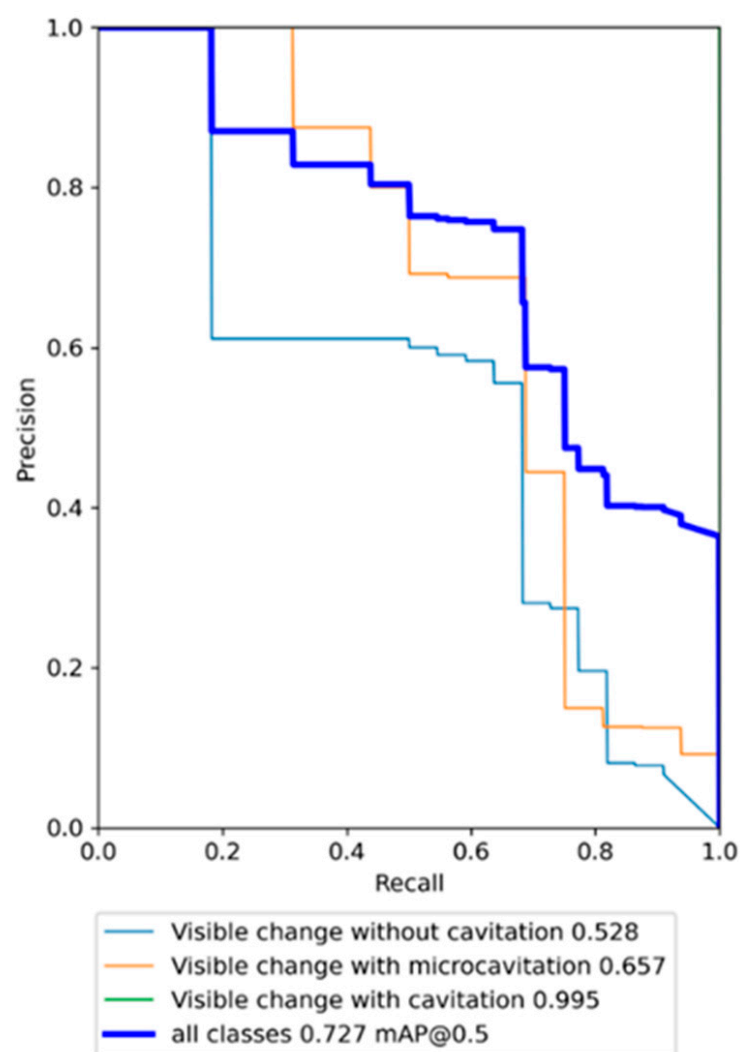
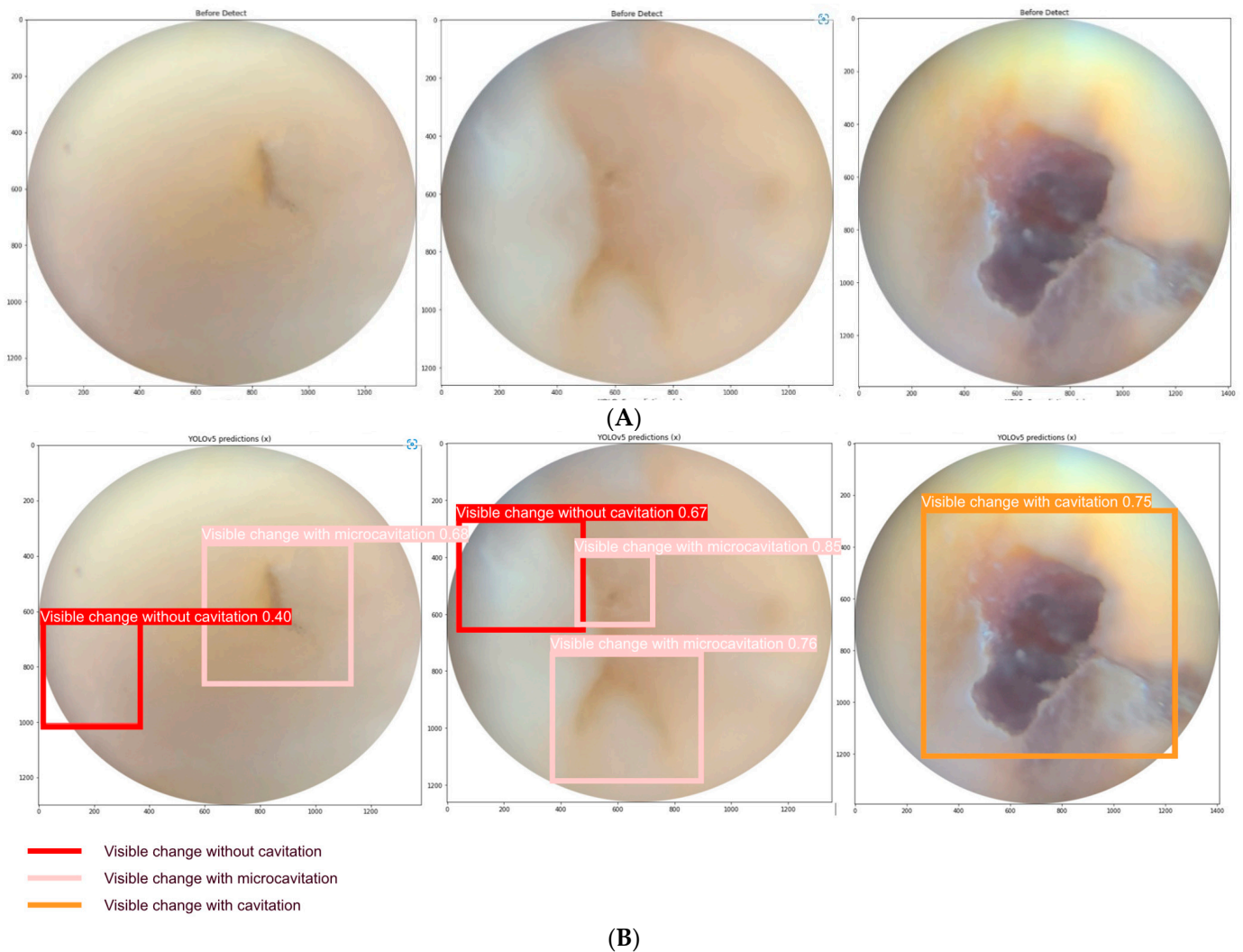
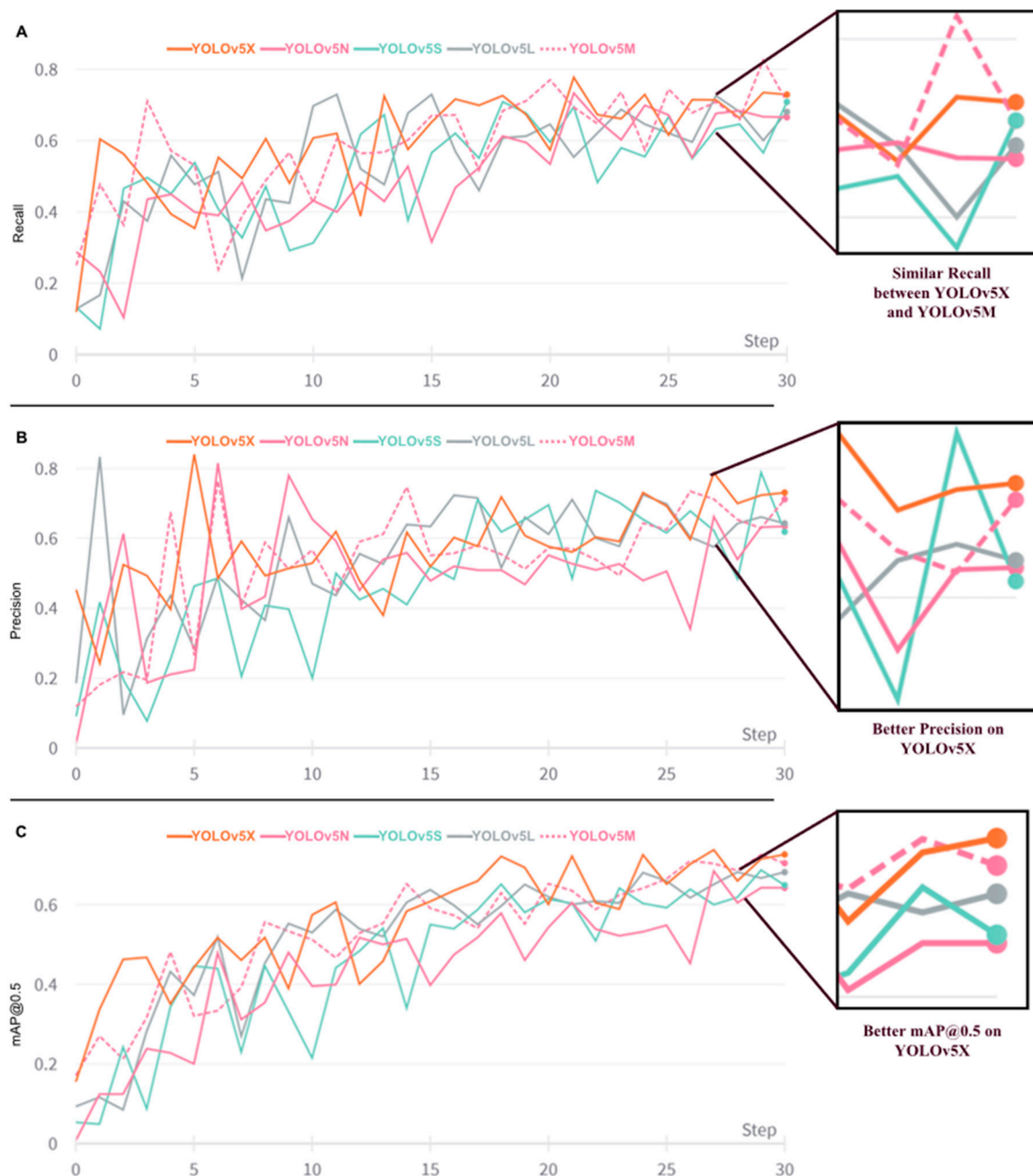


Figure 5. Precision-recall graph.

YOLOv5X and YOLOv5M achieved similar results for recall (0.708 for YOLOv5M and 0.729 for YOLOv5X, respectively) and precision (0.712 for YOLOv5M and 0.731 for YOLOv5X, respectively) across the tri-classification system. YOLOv5X generated an mAP of 0.727, while YOLOv5M generated an mAP of 0.705.



**Figure 6.** (A) Input images of carious lesions. (B) Output performance for carious lesions.



**Figure 7.** Performance metrics of the YOLO models during training and validation plotted against number of epochs at the horizontal axis. (A): Recall, (B): Precision, (C): mAP@0.5.

#### 4. Discussion

One of the main focuses of the current study was to explore the feasibility of the CDSS in carious lesion detection, where YOLOv5X and v5M were deemed the superior models for the current dental dataset. Interestingly, no single model was seen to be capable of diagnosing all three visual classifications of caries defects. While YOLO v5M excelled at diagnosing microcavitations, YOLO v5X performed better at diagnosing the absence of cavitations. Both diagnostic classifications are important in minimally invasive operative procedures, where best practice advises against the preparation of intact crown surfaces with carious progression. This is in favor of arresting caries with spontaneous remineralization without compromising the integrity of the remaining tooth structure. Microcavitations are minute details on the carious tooth surfaces that often require clinical magnification

and instrument-driven chairside evaluations to detect and confirm. Each iteration or family member of YOLO v5 has its own unique properties. YOLO v5S (S = Small) and v5M (M = Medium) being more adept at diagnosing the intricate details of microcavitation as opposed to YOLO v5L (L = Large) and v5X (extra-large) emphasizes that a model with a larger bank of pre-trained architecture may not be necessary or efficient as Medium might be more adequate at handling most case-specific tasks, similar to previous reports [27]. In the instance of real-time caries detection, the implementation of a light-weight variant of YOLO may translate to faster inference times with a subsequently reduced cost of application when scaled across the entire patient pool in a clinic. Such a claim must however be considered with the previously established knowledge that inter-device variations are quite common in clinical practice [28].

When comparing multiple object detection algorithms within one environment, a cost analysis of real-time implementation is warranted. A good number of in vitro simulations, such as the one performed in the current study “Google Colaboratory”, are executed over cloud computing, where inference times can vary heavily with network speeds and bandwidth limitations. Based on the findings of the current report and outcomes from previous investigations, a prediction can be made on how the YOLO v5 model on COCO datasets may behave in a clinic with a local network [29]. From the findings, an execution of YOLO v5S and v5M would require 14 to 41mb of storage at an inference time between 2.2 and 2.9 ms. The requirements increase exponentially when v5X at 168mb is required for a task that reports an inference time of 6.0 ms [29]. In simplified terms, while these methods can be implemented within a dental loupe with smart lenses, an optimized method of switching across the models needs to be developed that can adapt to real-time changeovers in dental diagnostic needs. Such an approach is necessary to prevent diagnostic latencies that may exceed 50 ms, a range noticeable by the human eye-head coordination, when switching to the model with greater accuracy for the specific lesion being analyzed in real-time [30].

Prior to the development of deep learning models for object detection, algorithms based on image processing were frequently used for image segmentation and detection. Several research studies have been proposed to automatically detect tooth decay using conventional image processing and machine learning techniques. A study by Duong et al. [2] used a two-step detection scheme using support vector machine (SVM) for dental caries detection. The researchers of the study concluded that the proposed SVM system required further improvement and verification as the data was only captured from the smartphone images. Another study [31] showed that 20% of the ROI was mistakenly diagnosed as dental caries using this technique, indicating that diagnosis via radiography alone without an objective assessment is inaccurate. Furthermore, radiography in itself is not as readily available in the form of smartphone micrography yet, with collimator-attachable apparatuses strictly controlled for the radiation hazards they pose.

A previous investigation [32] used deep CNNs to detect caries lesions on near-infrared light transillumination (NILT) images obtained from both in vitro and in vivo datasets to assess the models’ generalizability. However, the authors concluded that using in vitro setups to generate NILT imagery and subsequently training CNNs with the data resulted in lower accuracy. Another study [33] used four deep learning models, consisting of faster region-based convolutional neural networks (faster R-CNNs), YOLOv3, RetinaNet, and a single-shot multi-box detector (SSD), to detect initial caries lesions and cavities. While the study had three classes, only two classes were compared at a time in a binary format. Similar to the current report, the authors used common parameters such as sensitivity, specificity, accuracy, and precision to evaluate architectural deep learning performance and achieved favorable outcomes in the “cavitated lesions” (C) vs. “non-cavity” (NC) classification. However, their outcomes were substantially challenged by the “Visually Non-Cavitated” (VNC) vs. “No surface change” (NSC) classification.

Intersection-over-union (IoU) was a crucial parameter to select samples and models in previous research [34] and also plays an important role as a metric in evaluating how

accurately a model has applied its bounding boxes over a detected object [35]. Data augmentation played a key role in upscaling the small sample size in the current study to aid computational evaluation [36]. However, excessive augmentation may lead to data bias that is undetectable using *in vitro* simulations but may render the trained models ineffective when applied to real-life clinical scenarios [37]. To mitigate the problem, the current study limited its data augmentation practices to a reasonably minimum amount according to best practice outlines from current medical literature [38].

Aside from the limited size of the dataset, larger variations of carious lesions could have been helpful. In the current study, caries classification was based on a visual estimation, with true depth not being established. A histological ground reference and multi-center testing of trained algorithms was not performed similar to that in previous reports and serve as limitations in the inception of the workflow [8]. The deidentification process during data acquisition also meant that the age of the dentition could not be identified, which could have benefited the generalizability of the study. Furthermore, images of large pits and fissures were avoided, as were shadow casts, altered translucency, occult lesions with subsequently mineralized pit and fissure lesions, and variations in occlusal surface morphology, which were considered unhelpful in the algorithm comparison process. The current study could be expanded in future research with multi-label classifications [39], cost-sensitive learning [40], and curriculum learning features. Future studies would include a shift from the supervised learning provided to train the current model and develop a system that can perform caries detection in unsupervised environments, essentially predicting lesions from novel clinical photographs, possibly by separating the networks utilized for image segmentation and object detection. The promise of implementation onto novel photographs can open frontiers in 3D object detection in other aspects of restorative dentistry [41].

Oral health resources are unbalanced globally, with people from many regions having limited access to dental professionals. Moreover, traditional clinical and radiographic evaluations can add economic burdens for individuals of low income, which might also prevent them from attending regular clinical visits [42]. Hence, there is an increased need for intelligent systems that can aid in detecting underlying dental cavities at a low cost among large populations and prioritize the queue of patients in subsidized or not-for-profit practices. Such a system, when integrated in the form of a smartphone or camera application, can benefit clinicians by quickly screening patients based on the existing state of caries progression and streamlining remote consultations and referrals through automated progress monitoring, with the outcomes of the current work able to strengthen the neural backbone of existing smartphone-based tele-dentistry applications being researched [14]. Object detection methods such as the ones documented in the current literature can aid in the effort to geographically standardize the quality of dental healthcare.

## 5. Conclusions

The current *in vitro* study demonstrated that one-stage method of object detection was able to detect carious lesions from smartphone photography with varying outcomes in classification accuracy. Of the models tested, the YOLO v5X performed better in diagnosing carious lesions with microcavitation, while the YOLO v5M fared better in diagnosing non-cavitated carious lesions. Within the limitations of the current concept, no single object detection model was seen to be capable of detecting the progression of a carious lesion, and a combined approach may be required to implement a real-time diagnostic model in clinical practice.

**Author Contributions:** S.M.S.S.: conceptualization, methodology, software, formal analysis, manuscript writing, M.D.S.U.: conceptualization, methodology, software, formal analysis, writing, S.A.: conceptualization, methodology, software, formal analysis, reviewing, N.M.: Formal analysis, investigation, data curation, T.H.F.: validation, resources, formal analysis, writing, reviewing, supervision, J.D.: Validation, supervision, project administration. All authors have read and agreed to the published version of the manuscript.



**Funding:** This research received no external funding.

**Data Availability Statement:** Data and codes are available in a publicly accessible repository: [https://github.com/sifatk69/tooth\\_decay\\_object\\_detection](https://github.com/sifatk69/tooth_decay_object_detection) accessed on 6 February 2023.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Neuhaus, K.W.; Ellwood, R.; Lussi, A.; Pitts, N.B. Traditional lesion detection aids. *Detect. Assess. Diagn. Monit. Caries* **2009**, *21*, 42–51.
2. Duong, D.L.; Kabir, M.H.; Kuo, R.F. Automated caries detection with smartphone color photography using machine learning. *Health Inform. J.* **2021**, *27*, 14604582211007530. [CrossRef] [PubMed]
3. Berdouses, E.D.; Koutsouri, G.D.; Tripoliti, E.E.; Matsopoulos, G.K.; Oulis, C.J.; Fotiadis, D.I. A computer-aided automated methodology for the detection and classification of occlusal caries from photographic color images. *Comput. Biol. Med.* **2015**, *62*, 119–135. [CrossRef] [PubMed]
4. Meharry, M.R.; Dawson, D.; Wefel, J.S.; Harless, J.D.; Kummet, C.M.; Xiao, X. The effect of surface defects in early caries assessment using quantitative light-induced fluorescence (QLF) and micro-digital-photography (MDP). *J. Dent.* **2012**, *40*, 955–961. [CrossRef]
5. Morrison, A.S.; Gardner, J.M. Smart phone microscopic photography: A novel tool for physicians and trainees. *Arch. Pathol. Lab. Med.* **2014**, *138*, 1002. [CrossRef]
6. van Ginneken, B. Fifty years of computer analysis in chest imaging: Rule-based, machine learning, deep learning. *Radiol. Phys. Technol.* **2017**, *10*, 23–32. [CrossRef]
7. Mookiah, M.R.K.; Acharya, U.R.; Chua, C.K.; Lim, C.M.; Ng, E.Y.K.; Laude, A. Computer-aided diagnosis of diabetic retinopathy: A review. *Comput. Biol. Med.* **2013**, *43*, 2136–2155. [CrossRef]
8. Farook, T.H.; Dudley, J. Automation and deep (machine) learning in temporomandibular joint disorder radiomics. A systematic review. *J. Oral. Rehabil.* **2023**. [CrossRef]
9. Rao, M.A.; Lamani, D.; Bhandarkar, R.; Manjunath, T.C. Automated detection of diabetic retinopathy through image feature extraction. In Proceedings of the 2014 International Conference on Advances in Electronics Computers and Communications, Bangalore, India, 10–11 October 2014; pp. 1–6.
10. Liu, Z.; Gu, X.; Chen, J.; Wang, D.; Chen, Y.; Wang, L. Automatic recognition of pavement cracks from combined GPR B-scan and C-scan images using multiscale feature fusion deep neural networks. *Autom. Constr.* **2023**, *146*, 104698. [CrossRef]
11. Liu, W.; Quijano, K.; Crawford, M.M. YOLOv5-Tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8085–8094. [CrossRef]
12. Wang, D.; Liu, Z.; Gu, X.; Wu, W.; Chen, Y.; Wang, L. Automatic detection of pothole distress in asphalt pavement using improved convolutional neural networks. *Remote Sens.* **2022**, *14*, 3892. [CrossRef]
13. Qu, Z.; Gao, L.Y.; Wang, S.Y.; Yin, H.N.; Yi, T.M. An improved YOLOv5 method for large objects detection with multi-scale feature cross-layer fusion network. *Image Vis. Comput.* **2022**, *125*, 104518. [CrossRef]
14. Al-Jallad, N.; Ly-Mapes, O.; Hao, P.; Ruan, J.; Ramesh, A.; Luo, J.; Wu, T.T.; Dye, T.; Rashwan, N.; Ren, J.; et al. Artificial intelligence-powered smartphone application, AICaries, improves at-home dental caries screening in children: Moderated and unmoderated usability test. *PLoS Digit. Health* **2022**, *1*, e0000046. [CrossRef] [PubMed]
15. Gandhi, M.; Dhanasekaran, R. Diagnosis of diabetic retinopathy using morphological process and SVM classifier. In Proceedings of the 2013 International Conference on Communication and Signal Processing, Melmaruvathur, India, 3–5 April 2013; pp. 873–877.
16. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
17. Du, N.; Li, Y. Automated identification of diabetic retinopathy stages using support vector machine. In Proceedings of the IEEE 32nd Chinese Control Conference, Xi'an, China, 26–28 July 2013; pp. 3882–3886.
18. Cohen, J.F.; Korevaar, D.A.; Altman, D.G.; Bruns, D.E.; Gatsonis, C.A.; Hooft, L.; Irwig, L.; Levine, D.; Reitsma, J.B.; De Vet, H.C.; et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. *BMJ Open* **2016**, *6*, e012799. [CrossRef]
19. Norgeot, B.; Quer, G.; Beaulieu-Jones, B.K.; Torkamani, A.; Dias, R.; Gianfrancesco, M.; Arnaout, R.; Kohane, I.S.; Saria, S.; Topol, E.; et al. Minimum information about clinical artificial intelligence modeling: The MI-CLAIM checklist. *Nat. Med.* **2020**, *26*, 1320–1324. [CrossRef]
20. Yakovlev, A.; Lisovychenko, O. An approach for image annotation automatization for artificial intelligence models learning. *Адаптивні Системи Автоматичного Управління* **2020**, *1*, 32–40. [CrossRef]
21. Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinoujscie, Poland, 9–12 May 2018; pp. 117–122.
22. Casado-García, Á.; Domínguez, C.; García-Domínguez, M.; Heras, J.; Inés, A.; Mata, E.; Pascual, V. CLoDSA: A tool for augmentation in classification, localization, detection, semantic segmentation and instance segmentation tasks. *BMC Bioinform.* **2019**, *20*, 323. [CrossRef]

23. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
24. Wang, K.; Liew, J.H.; Zou, Y.; Zhou, D.; Feng, J. Panet: Few-shot image semantic segmentation with prototype alignment. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27–28 October 2019; pp. 9197–9206.
25. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
26. Padilla, R.; Netto, S.L.; Da Silva, E.A.B. A survey on performance metrics for object-detection algorithms. In Proceedings of the 2020 international Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 1–3 July 2020; pp. 237–242.
27. Ali, L.; Alnajjar, F.; Parambil, M.M.A.; Younes, M.I.; Abdelhalim, Z.I.; Aljassmi, H. Development of YOLOv5-Based Real-Time Smart Monitoring System for Increasing Lab Safety Awareness in Educational Institutions. *Sensors* **2022**, *22*, 8820. [\[CrossRef\]](#)
28. Farook, T.H.; Rashid, F.; Alam, M.K.; Dudley, J. Variables influencing the device-dependent approaches in digitally analysing jaw movement—A systematic review. *Clin. Oral. Investig.* **2022**, *27*, 489–504. [\[CrossRef\]](#)
29. Dlužnevskij, D.; Stefanovic, P.; Ramanauskaite, S. Investigation of YOLOv5 efficiency in iPhone supported systems. *Balt. J. Mod. Comput.* **2021**, *9*, 333–344. [\[CrossRef\]](#)
30. Goossens, H.H.; Opstal, A.V. Human eye-head coordination in two dimensions under different sensorimotor conditions. *Exp. Brain Res.* **1997**, *114*, 542–560. [\[CrossRef\]](#)
31. Musri, N.; Christie, B.; Ichwan, S.J.A.; Cahyanto, A. Deep learning convolutional neural network algorithms for the early detection and diagnosis of dental caries on periapical radiographs: A systematic review. *Imaging Sci. Dent.* **2021**, *51*, 237. [\[CrossRef\]](#)
32. Holtkamp, A.; Elhennawy, K.; Cejudo Grano de Oro, J.E.; Krois, J.; Paris, S.; Schwendicke, F. Generalizability of deep learning models for caries detection in near-infrared light transillumination images. *J. Clin. Med.* **2021**, *10*, 961. [\[CrossRef\]](#)
33. Thanh, M.T.G.; Van Toan, N.; Ngoc, V.T.N.; Tra, N.T.; Giap, C.N.; Nguyen, D.M. Deep learning application in dental caries detection using intraoral photos taken by smartphones. *Appl. Sci.* **2022**, *12*, 5504. [\[CrossRef\]](#)
34. Zhao, Q.; Chen, B.; Xu, H.; Ma, Y.; Li, X.; Feng, B.; Yan, C.; Dai, F. Unbiased IoU for Spherical Image Object Detection. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 508–515.
35. Wu, S.; Li, X.; Wang, X. IoU-aware single-stage object detector for accurate localization. *Image Vis. Comput.* **2020**, *97*, 103911. [\[CrossRef\]](#)
36. van Dyk, D.A.; Meng, X.-L. The art of data augmentation. *J. Comput. Graph. Stat.* **2001**, *10*, 1–50. [\[CrossRef\]](#)
37. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [\[CrossRef\]](#)
38. Chlap, P.; Min, H.; Vandenberg, N.; Dowling, J.; Holloway, L.; Haworth, A. A review of medical image data augmentation techniques for deep learning applications. *J. Med. Imaging Radiat. Oncol.* **2021**, *65*, 545–563. [\[CrossRef\]](#)
39. Tawiah, C.A.; Sheng, V.S. A study on multi-label classification. In *Industrial Conference on Data Mining*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 137–150.
40. Ling, C.X.; Sheng, V.S. Cost-sensitive learning and the class imbalance problem. *Encycl. Mach. Learn.* **2008**, *2011*, 231–235.
41. Farook, T.H.; Ahmed, S.; Jamayet, N.B.; Rashid, F.; Barman, A.; Sidhu, P.; Patil, P.; Lisan, A.M.; Eusufzai, S.Z.; Dudley, J.; et al. Computer-aided design and 3-dimensional artificial/convolutional neural network for digital partial dental crown synthesis and validation. *Sci. Rep.* **2023**, *13*, 1561. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Petersen, P.E.; Bourgeois, D.; Ogawa, H.; Estupinan-Day, S.; Ndiaye, C. The global burden of oral diseases and risks to oral health. *Bull. World Health Organ.* **2005**, *83*, 661–669. [\[PubMed\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.