Article

Deep Learning Application in Dental Caries Detection Using Intraoral Photos Taken by Smartphones

Mai Thi Giang Thanh [1,2], Ngo Van Toan [3,4], Vo Truong Nhu Ngoc [2], Nguyen Thu Tra [2,5], Cu Nguyen Giap [6,*] and Duc Minh Nguyen [2,7,*]

1 Hadong Medical College, Hanoi 100000, Vietnam; thanhmtg@yhadong.edu.vn
2 School of Dentistry, Hanoi Medical University, Hanoi 100000, Vietnam; nhungoc@hmu.edu.vn (V.T.N.N.); nguyenthutra@hmu.edu.vn (N.T.T.)
3 Institute for Preventive Medicine and Public Health, Hanoi Medical University, Hanoi 100000, Vietnam; ngovantoan@hmu.edu.vn
4 Hanoi Medical University Hospital, Hanoi 100000, Vietnam
5 Graduate School of Medicine, Nagoya University, Nagoya 466-8560, Japan
6 Department of Informatics, Thuongmai University, Hanoi 100000, Vietnam
7 Division of Research and Treatment for Oral Maxillofacial Congenital Anomalies, Aichi Gakuin University, Nagoya 64-8651, Japan
* Correspondence: cunguyengiap@tmu.edu.vn (C.N.G.); drmduc8@dpc.agu.ac.jp (D.M.N.); Tel.: +84-943-359-958 (C.N.G.); +8180-7893-2739 (D.M.N.)

Abstract: A mobile-phone-based diagnostic tool, which most of the population can easily access, could be a game changer in increasing the number of examinations of people with dental caries. This study aimed to apply a deep learning algorithm in diagnosing the stages of smooth surface caries via smartphone images. Materials and methods: A training dataset consisting of 1902 photos of the smooth surface of teeth taken with an iPhone 7 from 695 people was used. Four deep learning models, consisting of Faster Region-Based Convolutional Neural Networks (Faster R-CNNs), You Only Look Once version 3 (YOLOv3), RetinaNet, and Single-Shot Multi-Box Detector (SSD), were tested to detect initial caries lesions and cavities. The reference standard was the diagnosis of a dentist based on image examination according to the International Caries Classification and Management System (ICCMS) classification. Results: For cavitated caries, YOLOv3 and Faster R-CNN showed the highest sensitivity among the four tested models, at 87.4% and 71.4%, respectively. The sensitivity levels of these two models were only 36.9 % and 26% for visually non-cavitated (VNC). The specificity of the four models reached above 86% for cavitated caries and above 71% for VNC. Conclusion: The clinical application of YOLOv3 and Faster R-CNN models for diagnosing dental caries via smartphone images was promising. The current study provides a preliminary insight into the potential translation of AI from the laboratory to clinical practice.

Keywords: artificial intelligence; deep learning; caries detection; image analysis

1. Introduction

Dental caries is the most common oral health condition [1]. However, a previous Korean study showed only 21% of people in this country go to dental clinics and hospitals for dental examinations [2]. The rate might be significantly lower in low- and middle-income countries where dental examinations are expensive and not covered by insurance [3]. Contrary to the accessible routine checkup, smartphones can be available and affordable in most countries. Thus, a smartphone-based diagnostic tool, which most of the population can easily access, could be a game changer in increasing the number of examinations of people with dental caries.

Deep learning, with two major models—Massive-Training Artificial Neural Networks (MTANNs) and Convolutional Neural Networks (CNNs)—uses network structures consisting of multiple layers for automatically learning and self-learning backpropagation [4].

Deep learning with image input has been explosively growing and promising to become an important platform in medical images. One of its most popular applications in the medical field is classification [5]. Applications of deep learning in dentistry are remarkable in a variety of fields such as teeth-related diseases, dental plaque, and periodontium [6].

In terms of dental caries, currently, different approaches exist for building automatic diagnosis tools, such as the application of common data-mining algorithms on the factors from annual oral checkups [7] or the classification algorithms used two separate steps: image segmentation and classification [7,8]. However, the current prominent approach is building an object detector via deep learning models, such as CNN, deep neural network (DNN), Region-Based CNN (R-CNN), Fast R-CNN, Faster R-CNN, Mask R-CNN, You Only Look Once version 3 (YOLOv3), RetinaNet, and Single-Shot Multi-Box Detector (SSD) [6,9–11]. Research by Ding showed that the YOLOv3 algorithm has a potential capability for caries detection [12]. Kim built a home dental care system using the RetinaNet model and reported that the system allowed users to effectively manage their dental problems by providing needed dental treatment information [13]. Estai conducted a study using the Faster R-CNN for automatic detection of caries on bitewing radiographs. This study demonstrated a promising performance for detecting proximal surface caries on dental bitewings [14]. A study by Moutselos et al. using DNN Mask R-CNN to detect caries on occlusal surfaces showed an accuracy of 0.889 [10]. Another study applied CNN to detect white spots in dental photography, reporting a mean accuracy from 0.81 to 0.84 [11]. Several completed commercial software to detect dental caries are also available, such as Logicon Caries Detector for dental monitoring [15].

Previous studies were mainly conducted in the laboratory; data on the potential of AI in vivo have been limited [9]. Recently, Duong et al. [16] used photos taken by smartphones of the occlusal surface of molars and premolars to develop an automated caries detection software. Both the training and testing data were dried teeth. Casalegno et al. detected caries lesions in vivo on occlusal and proximal surfaces on posterior teeth. However, images were taken by the near-infrared transillumination device, which is rarely used in clinical practice [17]. A study by J. Kuhnisch used intraoral photos applying deep learning to diagnose tooth decay. It should be noted that photographs were taken by professional cameras (Nikon D300, D7100, and D7200), with a Nikon Micro 105 mm lens [18].

The aim of our study was to develop a deep learning model for dental caries diagnostic that can be used to build a smartphone application using input as smartphone intraoral photos.

2. Materials and Methods

2.1. Photographic Images

Participants were people who came to the School of Dentistry at the Hanoi Medical University, the Dental Department of the Vietnam–Cuba Friendship Hospital, and the Medical Center of the Hadong Medical College, for a dental examination during 2019–2020. Informed consent was obtained from all patients or, in cases of children younger than 18 years old, their parents. Patients with enamel-defected development or filling on the smooth surface were excluded.

All patients had their teeth cleaned to remove dental plaque and stains. Teeth were cleaned with a low-speed handpiece, polishing brushes (IPC, Boston, MA, USA), Nupro prophy paste (Nupro, Dentsply Sirona, Charlotte, NC, USA) then rinsed with water for 10 s and blown dry for 5 s before taking photos. Intraoral images of the smooth surface of teeth were photographed with an iPhone 7 (Apple, Chicago, IL, USA) with 3 views (central, right lateral, and left lateral view) to cover all teeth. The central view focused on buccal surfaces of incisors (tooth A, B) (Figure 1a,c). The lateral view involved buccal surfaces of teeth on one side (teeth C, D, and E) (Figure 1b,d). Equipment supporting photography were lip retractors (Osung MND Co., Ltd., Houston, TX, USA), intraoral mirrors (DME4G, Osung MND Co., Ltd., Houston, TX, USA), lamp-supported mirror handles, and blow-drying (FF-Photo, Osung MND Co., Ltd., Houston, TX, USA). Finally, 1902 intraoral images from

695 people for training and 750 intraoral images from 250 people for testing were included. All images were exported in JPEG format.
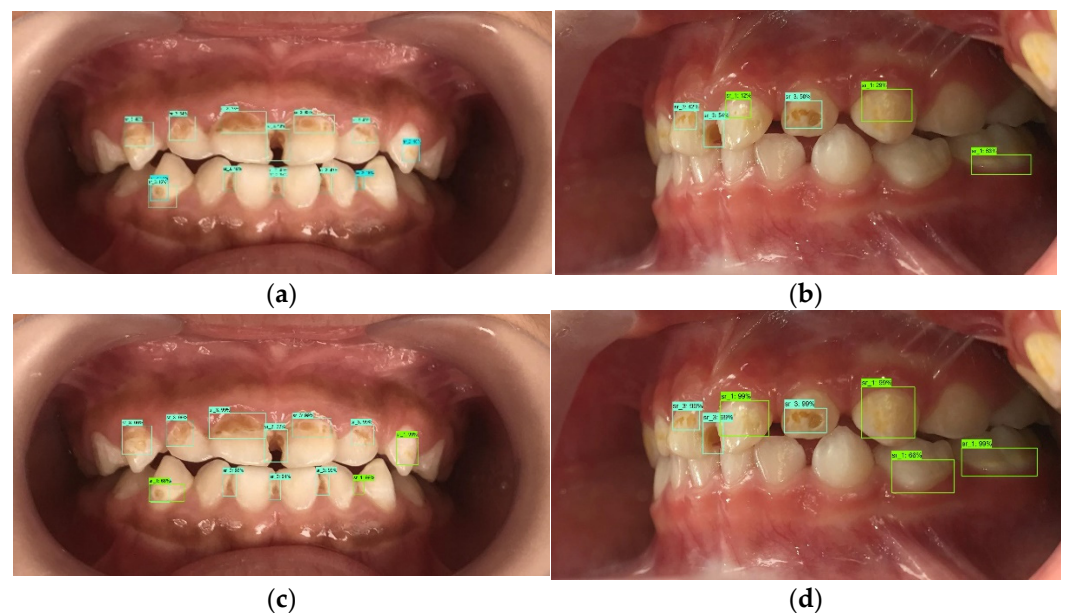

Figure 1. Central view (**a**,**c**) and lateral view (**b**,**d**) photos with non-cavitated (in green boxes) and cavitated lesions (in blue boxes) diagnosed by deep learning models (YOLOv3, Faster R-CNN).

2.2. Reference Standard and Labeling Dataset

Smartphone photos were visually inspected on a laptop (MacBook Air, Apple) by a single experienced dentist (V.T.N.N, >20-year experience) to detect any caries lesions based on the criteria of the International Caries Classification and Management System (ICCMS) [19]. All photos were diagnosed and labeled as follows: Class 0-Sound: "No surface change" (NSC); Class 1-Initial: "Visually Non-Cavitated (VNC), a white spot lesion or brown carious discoloration"; Class 2-Moderate: "Cavitated, a white or brown spot lesion with localized enamel breakdown, or an underlying dentine shadow"; Class 3-Extensive: "Late Cavitated, a distinct cavity in opaque or discolored enamel with visible dentine". In total, 1902 teeth were labeled as class 1, 1598 teeth were labeled as class 2, and 2127 teeth were labeled as class 3 (Figure 2).
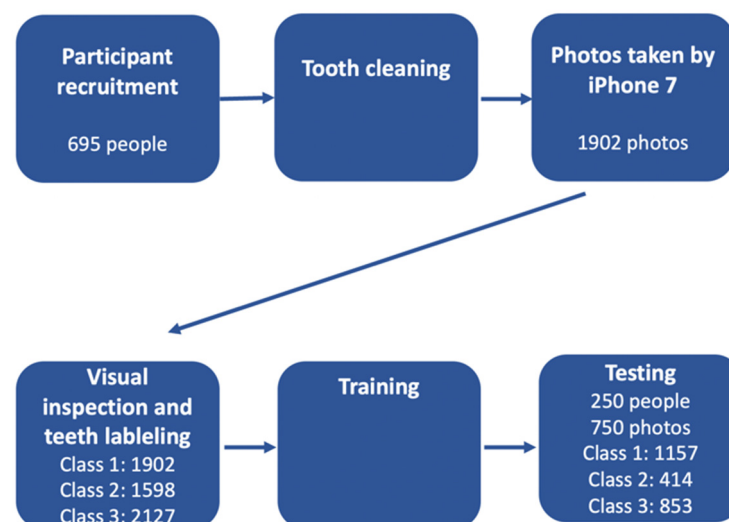

Figure 2. The flow chart of the method.

To test the reliability of the reference standard, a test–retest reliability was performed. An experienced dentist relabeled 30% of photos, randomly selected from the first 500 photos one week after the first diagnosis. Intraclass correlation coefficients (ICCs) were calculated to assess the test and retest the reliability. The ICC values were 0.917 ($p < 0.05$), showing excellent reliability.

2.3. Deep Learning Architecture

Deep learning is a class of artificial neural networks with many advances and has been applied successfully in computer vision including object detection [16]. Different variations of deep learning architecture are developed for object detection problems that are the general model of caries lesion detection. Therefore, in this section, we describe some selected deep learning models for the automatic detection of caries. We also present the data preparation, training process, and analysis method for these techniques.

In terms of technique, different deep learning architectures can be used, such as Fast R-CNN, Faster R-CNN, RetinaNet, YOLOv3, SSD, etc., for automatic detection of caries lesions from intraoral images. They are meta-network architecture, and there is no systematic study comparing these technologies' performance in general. The algorithms designed to perform object detection are commonly based on two approaches: one-stage object detection and two-stage object detection.

Regrading two-stage detectors, they have high localization and recognition accuracy [20]. Previous studies concluded that Faster R-CNN seemed to be the best at detecting small objects due to its power and its stability [17–19]. The dental caries lesions in our study were also small and had low contrast. Therefore, a two-stage detector, Faster R-CNN was selected. Herein, a brief introduction of the Faster R-CNN architecture, its implementation, and the training process are presented. The core blocks of a Faster R-CNN model are depicted generally in Figure 3, and its details can be found in a study by Ren et al. [21]. In Faster R-CNN architecture, convolutional layers are a convolutional neural network, and they work as a feature extraction block. In this study, we tested several CNN networks, including VGG16, Xception, and ResNet50 Inception-Resnet-v2. In the final version, the Inception-Resnet-v2 network was chosen. The architecture of Inception-Resnet-v2 implementation was introduced in detail to the public [22]. Faster R-CNN used a region-based proposal network to identify bounding boxes.
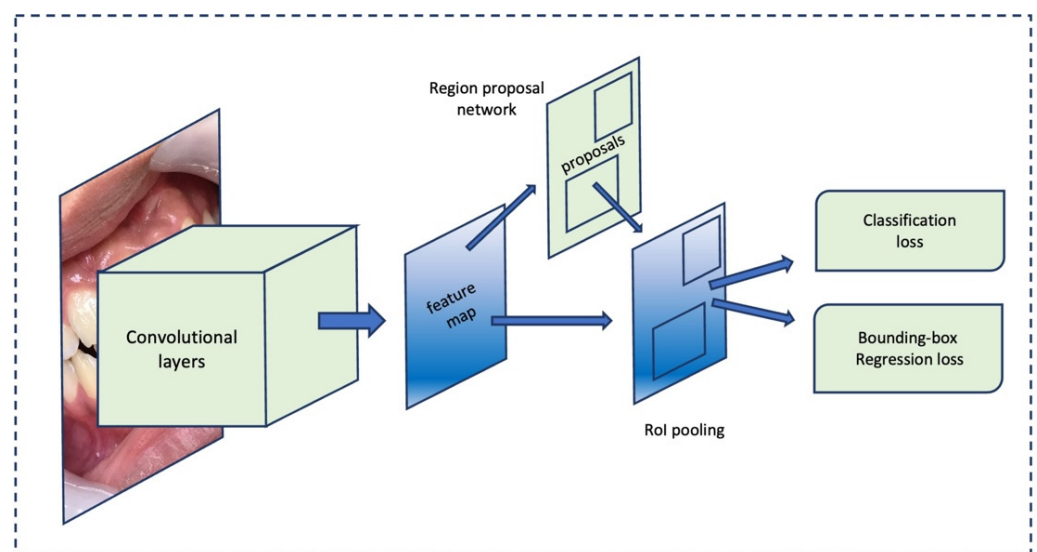

Figure 3. Faster-RCNN architecture.

On the other hand, one-stage detectors obtain high inference speeds. They include the algorithms of the YOLO family, SSD family, and RetinaNet family. The first YOLO version

was introduced by Joseph Redmon et al. in 2015 [23], and an updated version of YOLOv3 was presented in 2018 [24]. In principle, YOLOv3 uses only one single neural network trained by an end-to-end model. The model takes an image as input data, predicts bounding boxes containing objects, and labels each bounding box. Like the previous YOLO versions, YOLOv3 uses the "dimensional clustering proposal" algorithm to identify bounding boxes. The YOLOv3 network architecture is shown in Figure 4, which was presented in the study of MAO et al. [25].
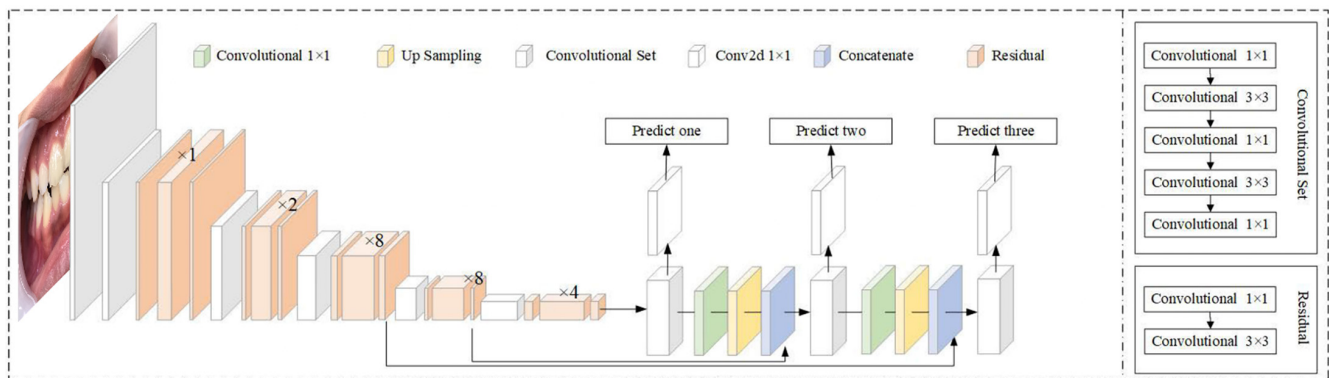

Figure 4. YOLOv3 architecture.

SSD, introduced by Liu et al. [26], generates a set of fixed-size bounding boxes on a feature map (also known as the offsets of the bounding box) and the relative scores presenting the label of the object contained. After that, the non-maximum suppression step is applied to combine generated bounding boxes to obtain a final predicted result. Similar to YOLO, the special feature to make the high speed of the SSD is that the model uses only a single neural network. In SSD, the model creates a grid of squares on the feature maps, and each cell is called a feature map cell. From the center of each feature map cell, a set of default boxes is defined to predict the frame that is capable of enclosing objects. The SSD training process has its own matching strategy [26] to refine the probability of the label and bounding box to match the model's ground truth input values (including labels and bounding box offsets). Moreover, the network is combined with many feature maps with different resolutions to detect objects of various sizes and shapes.

RetinaNet is another one-stage object detector, and this neural network architecture focuses on solving the imbalance between foreground and background classes [27]. In its approach, a focal loss function is defined to tackle the imbalance problem, and this function is used to replace the cross-entropy function. The basic blocks of RetinaNet include a Feature Pyramid Network Backbone and two subnetworks that are box-regression subnet and a classification subnet. RetinaNet uses translation-invariant anchor boxes to identify bounding boxes, which is similar to the mechanism of Faster R-CNN.

In this study, all models mentioned above were implemented using the Python programming language and used the Pytorch backend. The training process was carried out using a computer with Intel®Core™i7 CPU-3.00 GHz, 16 Gb RAM, and an 11 Gb memory GPU. The pretraining weights were also applied in the training process to improve processing time and convergence problems.

The training dataset was collected in the Vietnamese community with mobile cameras and was stored in PASCAL VOC format, a common format used for object detection problems [21,28]. In the case of YOLOv3, the training data were converted to a Common Object in Context (COCO) format. The size of original input images varied, depending on the setting of the cell phone camera; therefore, the size of images was automatically scaled to a uniform resolution. For the community application, this uniform resolution was set to 600-by-600 pixels, which is supported by almost available devices. Furthermore, to improve the quality of the training dataset, the Gaussian noise filter was applied to increase image quality.

Image augmentation methods were also applied to enlarge the training dataset. Some image augmentation methods can be commonly applied such as image rotation, shifting, flipping, scaling, cropping, or blurring. However, to minimize unexpected effects on the target disease object, we only applied flipping and rotation position augmentation. The color augmentation was not permitted.

A box was created with a code 1, 2, or 3 for each tooth that was suspected to be decayed. Then, the final output label of each photo was the label that included all diagnoses without detection of lesion location (Figure 1).

2.4. Evaluation

This study utilized common parameters to evaluate architectural deep learning performance using the visual inspection of photos as the reference method. Outputs of the software were conclusions of the presence or absence of carious lesions on the smooth surface of teeth, and the codes of the lesion classification, as Class 0 or 1 or 2 or 3.

TP: true positives, the number of cases that were correctly classified as positive; FP: false positives, the number of cases that were incorrectly classified as positive; FN: false negatives, the number of instances that were incorrectly classified as negative.

Sensitivity (true-positive rate) measures the proportion of positives that are correctly identified as follows:

$$Sensitivity = TP/TP + FN$$

Specificity (true-negative rate) measures the proportion of negatives that are correctly identified as follows:

$$Specificity = TN/TN + FP$$

Accuracy can be represented as the number of classified datasets divided by the total number of data test sets and is defined as

$$Accuracy = TP + TN/TP + TN + FP + FN$$

The precision rate indicates the correct prediction of the number of categories divided by the total number of data falling into that category as follows:

$$Precision = TP/TP + FP$$

2.5. Analysis

The data were analyzed using SPSS version 22 (IBM, Armonk, NY, USA). To conclude a diagnosis of cavitated caries, Class 0 and Class 1 were combinedly classified as "non-cavity" (NC) to compare with "cavitated lesions" (C) (Classes 2 and 3). Meanwhile, to detect early caries, NSC (Class 0) vs. VNC (Class 1) were also analyzed. Sensitivity, specificity, accuracy, recall, and precision were calculated for two classifications.

3. Results

3.1. C vs. NC Classification

Table 1 shows the results for the diagnosis of C vs. NC using machine learning models, compared with a visual inspection of photographs. The sensitivity of the YOLOv3 model was the highest, at 74%, followed by Faster R-CNN, RetinaNet, and SSD models. The accuracy of the Faster R-CNN model was the highest, at 87.4%, and the number of the SSD model was the lowest, at 81%.

Table 1. Model evaluation: results of C vs. NC classification.

Model Deep Learning	Sensitivity %	Specificity %	Accuracy %	Precision %
YOLOv3	74	86.6	83.4	65.3
Faster R-CNN	71.2	92.9	87.4	77.3
RetinaNet	63.2	89.8	83	67.7
SSD	26	99.7	81	97.1

C: cavitated; NSC: no surface change; VNC: visually non-cavitated.

3.2. NSC vs. VNC Classification

In terms of detecting NSC and VNC, the sensitivity levels of the four models were significantly decreased. Other parameters such as accuracy and precision were also reduced (Table 2).

Table 2. Model evaluation: results of VNC vs. NSC classification.

Model Deep Learning	Sensitivity %	Specificity %	Accuracy %	Precision %
YOLOv3	36.9	71.4	60.7	49.7
Faster R-CNN	23.4	87.7	67.8	61.5
RetinaNet	26.5	83.3	65.7	61.3
SSD	0	99.7	68.8	0

4. Discussion

Even though intraoral photos taken by smartphones are not usually used for clinical diagnosis of dental caries, the literature showed good accuracy of visual inspection by photographs in detecting dental caries [22]. The doctors' experience is also an important factor in diagnosing caries [28]. In the current study, the reference standard was the diagnosis of a single experienced dentist based on visual inspection by smartphone photos.

Cavitated caries on the tooth surface are a definite sign of dental caries. Differentiating cavitated lesions from sound enamel surfaces on smartphone images is feasible and accurate [29]. In this study, the specificity and sensitivity for cavitated caries detection of the two better-performing models—YOLOv3 and Faster R-CNN—were about 70% and 90%, respectively. These data were lower than previous in vitro studies. For example, a study by Duong et al. in 2021 showed a high result of sensitivity of 88.1% and specificity of 96.6% [16]. Two main factors led to the result. First, our study used intraoral photos taken via smartphones for deep learning training, which means there were several unfavored factors when taking intraoral photos that could interfere with the quality of photos, such as saliva, a lack of light, different camera angles, or presence of soft tissues rather than tooth exclusively. Secondly, the device taking all photos in the current study was a universal smartphone such as iPhone 7. Thus, different types of photos with divergent angles and low quality may lower the diagnostic specificity and sensitivity [30].

Initial caries lesions present as white or yellow-brown spots without destroying any structure or cavitated holes on the enamel surface [19]. Therefore, identification of the initial caries is challenging, even for experienced dentists. A meta-analysis showed the sensitivity and specificity of photographic visual examination for initial caries were only 67% and 79% [31]. In detecting VNC or initial caries, all four models in our study showed relatively low sensitivity, YOLOv3 (36.9%), Faster R-CNN (23.4%), Retina Net (26.5%), and SSD (0%). These deep learning models were trained based on the features that reflected the color, size, and location of predetermined lesions. Since the presence of initial lesions was vague and indistinctive, it might require a large size of data to obtain an accurate result. For example, the Faster R-CNN model misdiagnosed a reflection of light as a white-spot early caries (Figure 5). Our study was conducted with a modest number of clinical images, and a relatively low level of sensitivity in diagnosing initial caries was

not unexpected. Furthermore, dental caries detection from intraoral photos taken using smartphones is a distinctive small-object detection problem. This problem has severely affected the diagnostic results of deep learning algorithms. The limitation of the results of the present study prompts us to make extensive developments for these algorithms in near future. For example, we aim to fine-tune the backbone network of the Faster R-CNN model to deal with indistinct contour problems and adjust the training method for the "regional proposal network" module to improve the ability to locate small objects [32].

Figure 5. Faster R-CNN misdiagnosed a light reflection as a non-cavitated dental caries (green box).

High-resolution, standardized, and single-tooth photographs require professional cameras, expensive macro lenses, and experienced photographers. From the community's application point of view, it is impractical and cannot be obtained by normal people who want to use an application to check their teeth. A recent study by J. Kuhnisch, which used intraoral photos applying deep learning to diagnose tooth decay, achieved more than 90% agreement in the detection of caries. However, photographs are single-tooth photos taken with professional cameras and lenses (Nikon D300, D7100, and D7200, Nikon Micro 105-mm) [18]. Recently, there is one study in which YOLOv3 was used to detect dental caries via smartphone intraoral photos [12]. The study showed the mean average precision value (of 3 times of testing) of the YOLOv3 algorithm was 56.20%, which was lower than our study regarding cavitated caries detection (74%). The indirect comparison should be interpreted with caution since the two studies used two different datasets.

In this study, algorithms YOLOv3 and Faster R-CNN performed better than algorithms RetinaNet and SSD with the data of caries and initial caries lesions diagnosis problems. The results could imply that the theoretical improvements proposed for algorithms RetinaNet and SSD may not be suitable for the distinctive feature of the current data, which contained small and very small objects. Of course, to conclude on the cause of the outperformance of algorithms YOLOv3 and Faster R-CNN will need other in-depth studies on each of the different processing characteristics of the algorithms. However, the research results point to the great potential of YOLOv3 and Faster R-CNN in practical applications.

The results of the current study need to be improved for clinical application, especially in detecting initial caries. Most of the research direction in this field is shifting from the laboratory to clinical application. First, under ideal conditions and facilities, the deep learning model achieves very high accuracy. The next step is trying to maintain acceptable accuracy when testing under unfavored real-world conditions. In our approach, we sought to conduct an initial study with conditions that closely resemble those in the real community and then attempt to improve the accuracy in further stages. Several shortcomings need to be improved in further studies. First, input photos need to be enhanced to obtain a better quality of photos. Even though iPhone 7 was selected due to its popularity and affordable price, the gear camera failed to provide photos with good enough quality. A recent study showed an enhancement of initial photos could significantly improve the performance of the model [12]. This approach should be considered in future studies using smartphone photos as input. Secondly, the number of images for machine learning training needs to increase. Thirdly, the potential algorithms will be modified with the aim of dealing with the small-object detection problem. There might be a long way to go until there is a fast, accurate deep learning algorithm that uses intraoral, unstandardized photos taken from a

universal device to diagnose dental caries. Our study provides researchers and engineers with a view of the performance of four deep learning models in detecting dental caries.

5. Conclusions

In our study, we applied four deep learning models—Faster R-CNN, YOLOv3, RetinaNet, and SSD—to detect non-cavitated caries and cavitated caries through photos taken with a universal smartphone. YOLOv3 and Faster RCNN proved to be promising applications of AI in the real community for detecting cavitated caries. However, the accuracy and sensitivity of four models in detecting initial caries remained lower than expected for practical implementation. The results of this study reveal the possibilities of these models and can be further improved by enhancing input photos, increasing the dataset for training, and applying minor modifications to the deep learning algorithms above.

References

1. James, S.L.; Abate, D.; Abate, K.H.; Abay, S.M.; Abbafati, C.; Abbasi, N.; Abbastabar, H.; Abd-Allah, F.; Abdela, J.; Abdelalim, A. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **2018**, *392*, 1789–1858. [CrossRef]
2. Jung, S.; Jin, B.; Bae, K.; Lee, B.; Han, D. *Strategies and Educational Manuals to Improve National Oral Examination Program*; Ministry of Health and Welfare: Seoul, Korea, 2009; pp. 25–76.
3. Kandelman, D.; Arpin, S.; Baez, R.J.; Baehni, P.C.; Petersen, P.E. Oral health care systems in developing and developed countries. *Periodontology 2000* **2012**, *60*, 98–109. [CrossRef] [PubMed]
4. Choi, J.-W. Assessment of panoramic radiography as a national oral examination tool: Review of the literature. *Imaging Sci. Dent.* **2011**, *41*, 1–6. [CrossRef] [PubMed]
5. Benzian, H.; Hobdell, M.; Mackay, J. Putting teeth into chronic diseases. *Lancet* **2011**, *377*, 464. [CrossRef]
6. Hwang, J.-J.; Jung, Y.-H.; Cho, B.-H.; Heo, M.-S. An overview of deep learning in the field of dentistry. *Imaging Sci. Dent.* **2019**, *49*, 1–7. [CrossRef]
7. Tamaki, Y.; Nomura, Y.; Katsumura, S.; Okada, A.; Yamada, H.; Tsuge, S.; Kadoma, Y.; Hanada, N. Construction of a dental caries prediction model by data mining. *J. Oral Sci.* **2009**, *51*, 61–68. [CrossRef]
8. Patil, S.; Kulkarni, V.; Bhise, A. Algorithmic analysis for dental caries detection using an adaptive neural network architecture. *Heliyon* **2019**, *5*, e01579. [CrossRef]
9. Prados-Privado, M.; García Villalón, J.; Martínez-Martínez, C.H.; Ivorra, C.; Prados-Frutos, J.C. Dental caries diagnosis and detection using neural networks: A systematic review. *J. Clin. Med.* **2020**, *9*, 3579. [CrossRef]
10. Moutselos, K.; Berdouses, E.; Oulis, C.; Maglogiannis, I. Recognizing occlusal caries in dental intraoral images using deep learning. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 1617–1620.
11. Askar, H.; Krois, J.; Rohrer, C.; Mertens, S.; Elhennawy, K.; Ottolenghi, L.; Mazur, M.; Paris, S.; Schwendicke, F. Detecting white spot lesions on dental photography using deep learning: A pilot study. *J. Dent.* **2021**, *107*, 103615. [CrossRef]
12. Ding, B.; Zhang, Z.; Liang, Y.; Wang, W.; Hao, S.; Meng, Z.; Guan, L.; Hu, Y.; Guo, B.; Zhao, R.; et al. Detection of dental caries in oral photographs taken by mobile phones based on the YOLOv3 algorithm. *Ann. Transl. Med.* **2021**, *9*, 1622. [CrossRef]
13. Kim, D.; Choi, J.; Ahn, S.; Park, E. A smart home dental care system: Integration of deep learning, image sensors, and mobile controller. *J. Ambient Intell. Humaniz. Comput.* **2021**, 1–9. [CrossRef] [PubMed]

14. Estai, M.; Tennant, M.; Gebauer, D.; Brostek, A.; Vignarajan, J.; Mehdizadeh, M.; Saha, S. Evaluation of a deep learning system for automatic detection of proximal surface dental caries on bitewing radiographs. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* 2022; *in press*. [CrossRef] [PubMed]

15. DentalMonitoring Home Page. Available online: https://dental-monitoring.com/ (accessed on 21 May 2022).

16. 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]

17. Casalegno, F.; Newton, T.; Daher, R.; Abdelaziz, M.; Lodi-Rizzini, A.; Schürmann, F.; Krejci, I.; Markram, H. Caries Detection with Near-Infrared Transillumination Using Deep Learning. *J. Dent. Res.* **2019**, *98*, 1227–1233. [CrossRef] [PubMed]

18. Kühnisch, J.; Meyer, O.; Hesenius, M.; Hickel, R.; Gruhn, V. Caries Detection on Intraoral Images Using Artificial Intelligence. *J. Dent. Res.* **2022**, *101*, 158–165. [CrossRef]

19. Pitts, N.B.; Ekstrand, K.R. International Caries Detection and Assessment System (ICDAS) and its International Caries Classification and Management System (ICCMS)—Methods for staging of the caries process and enabling dentists to manage caries. *Community Dent. Oral Epidemiol.* **2013**, *41*, e41–e52. [CrossRef]

20. Lohia, A.; Kadam, K.D.; Joshi, R.R.; Bongale, A.M. Bibliometric Analysis of One-stage and Two-stage Object Detection. *Libr. Philos. Pract.* **2021**, 4910. Available online: https://digitalcommons.unl.edu/libphilprac/4910 (accessed on 28 March 2022).

21. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012.

22. Inês Meurer, M.; Caffery, L.J.; Bradford, N.K.; Smith, A.C. Accuracy of dental images for the diagnosis of dental caries and enamel defects in children and adolescents: A systematic review. *J. Telemed. Telecare* **2015**, *21*, 449–458. [CrossRef]

23. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

24. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

25. Mao, Q.-C.; Sun, H.-M.; Liu, Y.-B.; Jia, R.-S. Mini-YOLOv3: Real-time object detector for embedded applications. *IEEE Access* **2019**, *7*, 133529–133538. [CrossRef]

26. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In *Computer Vision—ECCV 2016. ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science Series; Springer: Cham, Switzerland, 2016; Volume 9905, pp. 21–37.

27. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.

28. Gimenez, T.; Piovesan, C.; Braga, M.M.; Raggio, D.P.; Deery, C.; Ricketts, D.N.; Ekstrand, K.R.; Mendes, F.M. Visual Inspection for Caries Detection: A Systematic Review and Meta-analysis. *J. Dent. Res.* **2015**, *94*, 895–904. [CrossRef] [PubMed]

29. Kohara, E.K.; Abdala, C.G.; Novaes, T.F.; Braga, M.M.; Haddad, A.E.; Mendes, F.M. Is it feasible to use smartphone images to perform telediagnosis of different stages of occlusal caries lesions? *PLoS ONE* **2018**, *13*, e0202116. [CrossRef] [PubMed]

30. Koziarski, M.; Cyganek, B. Impact of Low Resolution on Image Recognition with Deep Neural Networks: An Experimental Study. *Int. J. Appl. Math. Comput. Sci.* **2018**, *28*, 735–744. [CrossRef]

31. Thanh, M.T.G.; Van Toan, N.; Toan, D.T.T.; Thang, N.P.; Dong, N.Q.; Dung, N.T.; Hang, P.T.T.; Anh, L.Q.; Tra, N.T.; Ngoc, V.T.N. Diagnostic Value of Fluorescence Methods, Visual Inspection and Photographic Visual Examination in Initial Caries Lesion: A Systematic Review and Meta-Analysis. *Dent. J.* **2021**, *9*, 30. [CrossRef] [PubMed]

32. Cao, C.; Wang, B.; Zhang, W.; Zeng, X.; Yan, X.; Feng, Z.; Liu, Y.; Wu, Z. An improved faster R-CNN for small object detection. *IEEE Access* **2019**, *7*, 106838–106846. [CrossRef]