



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학박사 학위논문

**A Deep Learning-based Computer-
aided Diagnosis Method for
Radiographic Bone Loss and
Periodontitis Stage: A Multi-device
Study**

방사선학적 골 소실량과 치주염
단계의 딥러닝 기반 컴퓨터
보조진단 방법: 다기기 연구

2021 년 2 월

서울대학교 대학원

융합과학부 방사선융합의생명 전공

이 상 정

방사선학적 골 소실량과 치주염 단계의 딥러닝 기반 컴퓨터 보조진단 방법: 다기기 연구

지도교수 이 원 진

이 논문을 공학박사 학위논문으로 제출함

2020 년 11 월

서울대학교 대학원

융합과학부 방사선융합의생명 전공

이 상 정

이상정의 공학박사 학위논문을 인준함

2020 년 12 월

위 원 장	김 태 일	(인)
부위원장	이 원 진	(인)
위 원	허 민 석	(인)
위 원	김 종 효	(인)
위 원	김 광 기	(인)

A Deep Learning-based Computer-aided Diagnosis Method for Radiographic Bone Loss and Periodontitis Stage: A Multi-device Study

by
Sang-Jeong Lee

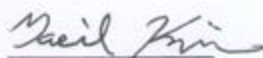
**A Thesis Submitted to the Department of
Biomedical Radiation Sciences in Partial
Fulfilment of the Requirements for the Degree of
Doctor of Philosophy at Seoul National University**

December 2020

Approved by thesis committee:

Chairman

Tae-Il Kim



Vice chairman

Won-Jin Yi



Member

Min-Suk Heo



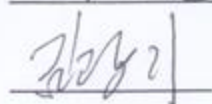
Member

Jong Hyo Kim



Member

Kwang Gi Kim



ABSTRACT

A Deep Learning-based Computer-aided Diagnosis Method for Radiographic Bone Loss and Periodontitis Stage: A Multi-device Study

Sang-Jeong Lee

Department of Biomedical Radiation Sciences

Graduate School of Convergence Science and

Technology

Seoul National University

Periodontal diseases, including gingivitis and periodontitis, are some of the most common diseases that humankind suffers from. The decay of alveolar bone in the oral and maxillofacial region is one of the main symptoms of

periodontal disease. This leads to alveolar bone loss, tooth loss, edentulism, and masticatory dysfunction, which indirectly affects nutrition. In 2017, the American Academy of Periodontology and the European Federation of Periodontology proposed a new definition and classification criteria for periodontitis based on a staging system. Recently, computer-aided diagnosis (CAD) based on deep learning has been used extensively for solving complex problems in radiology.

In my previous study, a deep learning hybrid framework was developed to automatically stage periodontitis on dental panoramic radiographs. This was a hybrid of deep learning architecture for detection and conventional CAD processing to achieve classification. The framework was proposed to automatically quantify the periodontal bone loss and classify periodontitis for each individual tooth into three stages according to the criteria that was proposed at the 2017 World Workshop. In this study, the previously developed framework was improved in order to classify periodontitis into four stages by detecting the number of missing teeth/implants using an additional convolutional neural network (CNN). A multi-device study was performed to verify the generality of the method.

A total of 500 panoramic radiographs (400, 50, and 50 images for device 1, device 2, and device 3, respectively) from multiple devices were collected to train the CNN. For a baseline study, three CNNs, which were commonly used for segmentation tasks and the modified CNN from the Mask Region with CNN (R-CNN) were trained and tested to compare the detection accuracy using

dental panoramic radiographs that were acquired from multiple devices. In addition, a pre-trained weight derived from the previous study was used as an initial weight to train the CNN to detect the periodontal bone level (PBL), cemento-enamel junction level (CEJL), and teeth/implants to achieve a high training efficiency. The CNN, trained with the multi-device images that had sufficient variability, can produce an accurate detection and segmentation for the input images with various aspects.

When detecting the missing teeth on the panoramic radiographs, the values of the precision, recall, F1-score, and mean average precision (AP) were set to 0.88, 0.85, 0.87, and 0.86, respectively, by using CNNv4-tiny. As a result of the qualitative and quantitative evaluation for detecting the PBL, CEJL, and teeth/implants, the Mask R-CNN showed the highest dice similarity coefficients (DSC) of 0.96, 0.92, and 0.94, respectively.

Next, the automatically determined stages from the framework were compared to those that were developed by three oral and maxillofacial radiologists with different levels of experience. The mean absolute difference (MAD) between the periodontitis staging that was performed by the automatic method and that by the radiologists was 0.31 overall for all the teeth in the whole jaw. The classification accuracies for the images from the multiple devices were 0.25, 0.34, and 0.35 for device 1, device 2, and device 3, respectively.

The overall Pearson correlation coefficient (PCC) values between the developed method and the radiologists' diagnoses were 0.73, 0.77, and 0.75 for

the images from device 1, device 2, and device 3, respectively ($p < 0.01$). The final intraclass correlation coefficient (ICC) value between the developed method and the radiologists' diagnoses for all the images was 0.76 ($p < 0.01$). The overall ICC values between the developed method and the radiologists' diagnoses were 0.91, 0.94, and 0.93 for the images from device 1, device 2, and device 3, respectively ($p < 0.01$). The final ICC value between the developed method and the radiologists' diagnoses for all the images was 0.93 ($p < 0.01$).

In the Passing and Bablok analysis, the slopes were 1.176 ($p > 0.05$), 1.100 ($p > 0.05$), and 1.111 ($p > 0.05$) with the intersections of -0.304, -0.199, and -0.371 for the radiologists with ten, five, and three-years of experience, respectively. For the Bland and Altman analysis, the average of the difference between the mean stages that were classified by the automatic method and those diagnosed by the radiologists with ten-years, five-years, and three-years of experience were 0.007 (95 % confidence interval (CI), -0.060 ~ 0.074), -0.022 (95 % CI, -0.098 ~ 0.053), and -0.198 (95 % CI, -0.291 ~ -0.104), respectively.

The developed method for classifying the periodontitis stages that combined the deep learning architecture and conventional CAD approach had a high accuracy, reliability, and generality when automatically diagnosing periodontal bone loss and the staging of periodontitis by the multi-device study. The results demonstrated that when the CNN used the training data sets with increasing variability, the performance also improved in an unseen data set.

Keywords: Deep learning, Computer-aided diagnosis (CAD), Periodontitis,
Radiographic bone loss, Dental panoramic radiograph, Multi-device study

Student Number: 2014-24908

CONTENTS

Abstract	i
Contents	vi
List of figures	viii
List of tables	x
List of abbreviations	xii
Introduction	1
Materials and Methods	5
Overall process for deep learning-based computer- aided diagnosis method	5
Data preparation of dental panoramic radiographs from multiple devices	7
Detection of PBL and CEJL structures and teeth using CNNs	10
Detection of the missing teeth using CNNs	14

Staging periodontitis by the conventional CAD method	17
Evaluation of detection and classification performance	20
Results	22
Detection performance for the anatomical structures	22
Detection performance for the missing teeth	26
Classification performance for the periodontitis stages	30
Classification performance of correlations, regressions, and agreements between the periodontitis stages	36
Discussion	42
References	55
Abstract in Korean	73

LIST OF FIGURES

Figure 1. Overall process for a developed deep learning-based computer-aided diagnosis (CAD) method for radiographic bone loss and periodontitis stage on dental panoramic radiographs	6
Figure 2. Detection results for the periodontal bone level (PBL) (a-c), the cemento-enamel junction level (CEJL) (d-f), and the teeth/implants (g-i) by the Mask R-CNN. The first, second, and third rows are images from the device 1, device 2, and device 3, respectively.	13
Figure 3. The long-axis orientations of the tooth and the implant (a-c), the intersection points of the tooth (implant) long-axis with the periodontal bone level (PBL) and the cemento-enamel junction level (CEJL) (fixture top level), the percentage rate of the radiographic bone loss (RBL) (d-f), and the stages of the periodontitis for each tooth and implant (g-i). The first, second, and third rows are images from the device 1, device 2, and device 3, respectively.	19
Figure 4. The best cases of detection results for the periodontal bone level (PBL) (a-e), the cemento-enamel junction level (CEJL) (f-j), and the teeth/implants (k-o) by the CNNs.	23
Figure 5. The worst cases of detection results for the periodontal bone level (PBL) (a-e), the cemento-enamel junction level (CEJL) (f-j), and the teeth/implants (k-o) by the CNNs. Locations of the false positive and negative errors are indicated in yellow and red by arrows, respectively.	24

Figure 6. The detection results for the missing teeth by the CNNv4-tiny using dental panoramic radiographs acquired from multiple devices. The detected bounding boxes of the locations of the missing teeth were superimposed on the input images for type 1 (a-c) and type 2 (d-f). The first, second, and third rows are images from the device 1, device 2, and device 3, respectively. 27

Figure 7. Precision-recall (PR) curves from automatic detection of the missing teeth with CNNv3-tiny, CNNv3, CNNv4-tiny, and CNNv4. 29

Figure 8. The stages of the periodontitis for each tooth and implant on the dental panoramic radiographs acquired from multiple devices. The automatic diagnosis results by the developed method on the images. The first, second, and third rows are images from the device 1, device 2, and device 3, respectively. 32

Figure 9. Scatter plots with Passing and Bablock regression of mean periodontitis stages obtained using the automatic method and those diagnosed by the radiologists with ten-years (a), five-years (b), and three-years (c) of experience. 40

Figure 10. Bland and Altman plots of mean periodontitis stages obtained using the automatic method and those diagnosed by the radiologists with ten-years (a), five-years (b), and three-years (c) of experience. 41

LIST OF TABLES

Table 1. The descriptions of the dental panoramic X-ray machines used for collecting the patient’s dental panoramic radiographs.	9
Table 2. The detailed strategies for training the CNNs to detect the periodontal bone level (PBL), the cemento-enamel junction level (CEJL), and the teeth/implants.	12
Table 3. The structural configurations and detailed strategies for training the CNNs to detect the missing teeth.	16
Table 4. Dice similarity coefficient (DSC) values for detection performance of the periodontal bone level (PBL), the cemento-enamel junction level (CEJL), and the teeth/implants by the CNNs for the multi-device images.	25
Table 5. Precision, recall, F1-score, and mean average precision (AP), for detecting the missing teeth on the multi-device images by the CNNs.	28
Table 6. The mean absolute differences (MADs) between periodontitis stages obtained using the automatic method and those diagnosed by the radiologists (with ten-years, five-years, and three-years of experience) using the multi-device images.	33
Table 7. The mean absolute differences (MADs) between periodontitis stages for types of teeth obtained using the automatic method and those diagnosed by	

the radiologists (with ten-years, five-years, and three-years of experience) using
the multi-device images. 34

Table 8. The mean absolute differences (MADs) between mean periodontitis
stage of image obtained using the automatic method and those diagnosed by the
radiologists (with ten-years, five-years, and three-years of experience) using the
multi-device images. 35

Table 9. The Pearson correlation coefficients (PCC) between stages obtained
using the automatic method and those diagnosed by the radiologists (with ten-
years, five-years, and three-years of experience) using the multi-device images.
..... 38

Table 10. The intraclass correlation coefficients (ICC) between stages obtained
using the automatic method and those diagnosed by the radiologists (with ten-
years, five-years, and three-years of experience) using the multi-device images
..... 39

LIST OF ABBREVIATIONS

<i>Full Name</i>	<i>Abbreviations</i>
Clinical Attachment Loss	CAL
Radiographic Bone Loss	RBL
Computer Aided Diagnosis	CAD
Convolutional Neural Network	CNN
Institutional Review Board	IRB
Periodontal Bone Level	PBL
Cemento-Enamel Junction Level	CEJL
Oral and Maxillo-Facial	OMF
Regions with Convolutional Neural Network	R-CNN
You Look Only Once	YOLO
Cross-Stage-Partial-connections	CSP
Spatial Pyramid Pooling	SPP

Path Aggregation Network	PAN
Dice Similarity Coefficient	DSC
True Positive	TP
False Positive	FP
False Negative	FN
Average Precision	AP
Precision-Recall	PR
Area Under the Curve	AUC
Mean Absolute Difference	MAD
Pearson Correlation Coefficients	PCC
Intraclass Correlation Coefficients	ICC
Statistical Package for Social Sciences	SPSS
Confidence Interval	CI
Visual Geometry Group	VGG

Support Vector Machine	SVM
Fully Convolutional Network	FCN
Cone-Beam Computed Tomography	CBCT
Region of Interest	ROI

INTRODUCTION

Periodontitis is one of the most common diseases that humankind suffers from, and it is the sixth most prevalent disease in the world. The decay of alveolar bone is one of the main symptoms of periodontitis and this leads to tooth loss, edentulism, and masticatory dysfunction [1].

In 1989, the need for a classification system for periodontitis was felt, to provide a standard for studying the pathogenesis, etiology, and treatment of the disease, and a simple classification criterion was established in 1993 [2]. Subsequently, the classification systems for periodontitis have been continuously revised according to the newly identified scientific and clinical evidence [3]. In 2017, a new classification system for periodontitis based on a staging and grading strategy was proposed at the World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. The stage of periodontitis is determined by its severity, complexity, and other factors, and it is divided into four levels [4].

In general, the clinical attachment loss (CAL) is measured with a periodontal probe, and it is used to evaluate the periodontal health. However, this method is limited in terms of its reliability, and it can be replaced with the radiographic bone loss (RBL) that is measured in radiographic images [4].

Conventional computer-aided diagnosis (CAD) systems can support the clinicians' decision-making process by extracting the significant features from

medical images that are produced by various modalities [5]. However, the conventional CAD approach has limitations such as difficulty in extraction due to the diversity of the disease patterns and it is a time-consuming process [6]. Machine learning based CAD methods can be applied to analyze the retrospective medical data of patients and produce the extracted features with certain disease outcomes [5]. Recent methods that are based on deep learning, which is a subset of machine learning, can automatically extract the relevant features during training, and it uses the whole image directly without the best-feature representation [7-9]. Deep convolutional neural network (CNN), a form of deep learning, is the most commonly used method for segmentation [10, 11], classification [12, 13], and detection [14, 15] of organs and related diseases in medical imaging. For the detection and classification, a variety of studies have been conducted to determine the specific categorical features of the target organs or the intended regions [16].

In the oral and maxillofacial field, CAD approaches that are integrated with CNNs have been used widely to solve complex clinical problems such as detecting landmarks in cephalograms [17], detecting teeth and classification [18-20], diagnosing cavities [21-25], and detecting maxillary sinusitis [26]. These can provide dental professionals with a second opinion that are automatically derived by CAD with deep learning approaches. In regard to periodontal diseases, several related methods that use a large number of images based on the CNNs have been trained to detect RBL on dental panoramic

radiographs without quantification of the loss and classification of the periodontitis stages [27, 28].

In a previous study, a deep learning hybrid framework was developed to automatically classify periodontitis of the individual teeth on the dental panoramic radiographs for the first time. The deep learning architecture was combined with the conventional CAD approach to detect and classify the periodontitis into three stages according to the 2017 World Workshop criteria. However, the previous method was not able to diagnose periodontitis stage four because it could not detect and quantify the locations of the missing teeth. In addition, this method only uses images that are acquired from a single modality, thus, resulting in poor dataset diversity and vendor dependent problems of the CNN. The detection accuracy of the CNN might degrade if the method was directly applied to the images that are collected from different centers or vendors [29]. To overcome this limitation, a CNN that uses transfer learning was developed to improve the detection accuracy of periodontal bone loss on the panoramic radiographs [30].

In this study, a deep learning-based CAD method was developed to classify periodontitis into four stages by using dental panoramic radiographs. This was achieved by implementing an additional CNN to the previously developed method [31]. To segment the anatomical structures on the multi-device images, transfer learning was performed by using the weights derived from the previous study [31] to train the Mask R-CNN. The trained CNN automatically detected and quantified the missing teeth to classify between

periodontitis stages three and four as per the 2017 World Workshop criteria. In addition, a multi-device study was performed to address the generality issue of the developed framework that used the dental panoramic radiographs, acquired with multiple devices from different vendors.

MATERIALS AND METHODS

Overall process for deep learning-based computer-aided diagnosis method

Figure 1 shows the overall process for a deep learning-based CAD method for measuring the RBL and determining the periodontitis stage. All panoramic radiographs that were used in this study were collected retrospectively after removing the identifiable patient information. The study was approved by the Institutional Review Board (IRB) of the Seoul National University Dental Hospital (ERI18001) with a waiver of informed consent. The data collection and all the experiments were performed in accordance to the relevant guidelines and regulations.

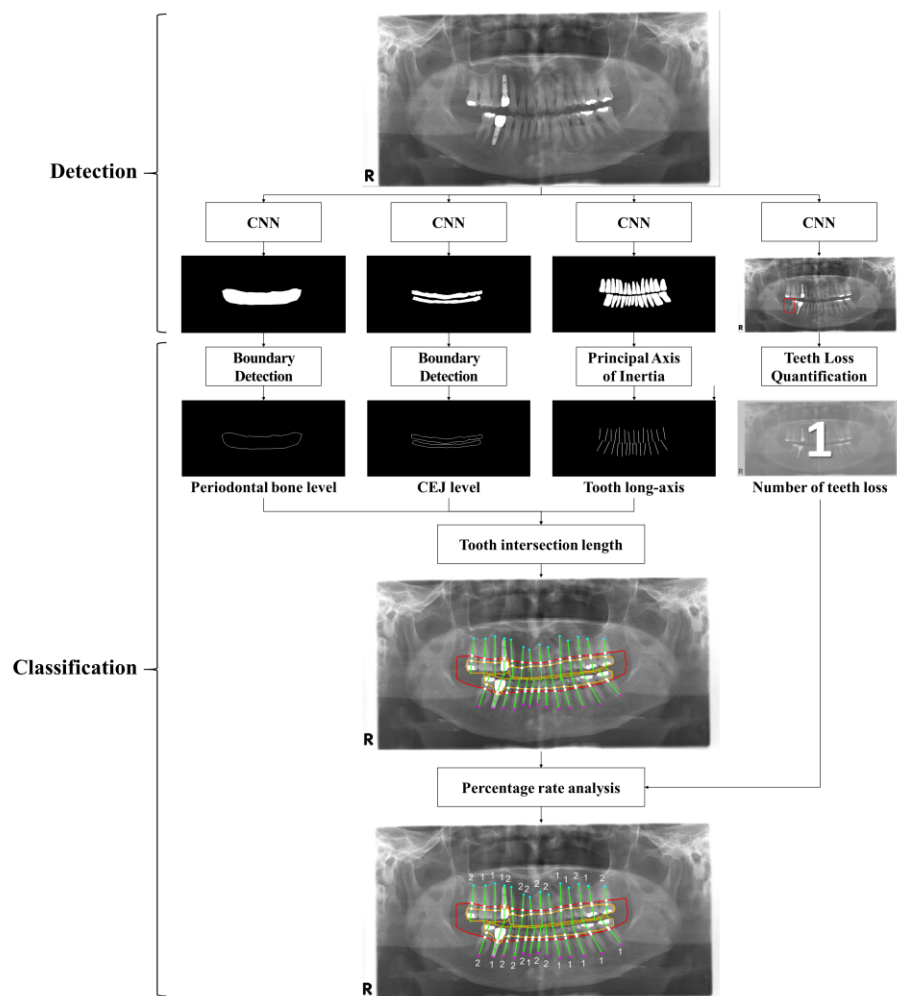


Figure 1. Overall process for a developed deep learning-based computer-aided diagnosis (CAD) method for radiographic bone loss and periodontitis stage on dental panoramic radiographs.

Data preparation of dental panoramic radiographs from multiple devices

The panoramic radiographs of each patient were acquired with three dental panoramic x-ray machines from three devices. A total of 410 panoramic radiographs of the patients were acquired by using device 1, an orthopantomograph OP 100 (Instrumentarium corporation, Tuusula, Finland). A total of 60 panoramic radiographs of the patients for devices 2 and 3 were acquired using PaX-i3D Smart (Vatech, Seoul, Korea) and Point3D Combi (PointNix, Seoul, Korea), respectively. A total of 530 panoramic radiographs, excluding the images of the patients with primary or mixed dentition, were collected to prepare the datasets for training and testing the CNNs and the descriptions are summarized in Table 1. With regard to the panoramic image dataset, 500, 500, and 300 images were used to detect the anatomical structures, such as the periodontal bone level (PBL), cementoenamel junction level (CEJL), and teeth/implants, respectively. In addition, three configurations of the dataset were made for a multi-device study. The first dataset contained all the images from the three devices, and the images in the dataset were also randomly separated into a training set (90 %) and a test set (10 %). In the second dataset, the images from devices 1 and 2 were used as the training set, and images from device 3 were used as the test set. Similarly, for the third dataset, images from devices 1 and 3 were used as the training set, and the images from device 2 were used as the test set. For all the datasets, the number of images that were included in the training set and the test set remained the same.

To detect the location of the missing teeth and to quantify their number, the image of the patient with the missing teeth was selected and classified into two types according to the number of missing teeth. The loss of one non-contiguous tooth was classified as type 1, and loss of two consecutive teeth was classified as type 2. As a result of the selection, 147 and 62 images were used for detecting and quantifying the missing teeth for type 1 and type 2, respectively. The images in the dataset were also randomly separated into a training set (90 %) and a test set (10 %).

The training set was used for CNN detection training, and the testing set was used to evaluate the final trained model. Data augmentation was performed to increase the number of images in the training set. The images were flipped horizontally, rotated, and translated, and their gray values were transformed by contrast-normalization. Therefore, the number of images increased by 64 times in comparison to the original amount. After training the CNNs, thirty panoramic radiographs (ten images for each device), which were not used for the detection, were used to evaluate the classification performance for diagnosing the RBL and the periodontitis stage.

Table 1. The descriptions of the dental panoramic X-ray machines used for collecting the patient's dental panoramic radiographs.

Vendors	Modalities	Number of images	Image size (pixels)	Field of view (cm)
Instrumentarium Corp.	Orthopantomograph OP 100	400	1,976 x 976	30 x 15
Vatech	PaX-i3D Smart	50	2,752 x 1,372	20.20 x 10.5
PointNix	Point3D Combi	50	2,000 x 1,152	31.25 x 10

Detection of PBL and CEJL structures and teeth using CNNs

By using a similar procedure in a previous study [31], the PBL was annotated as one simple structure for the whole jaw on the panoramic radiograph; the CEJL of the teeth (the fixture top level of implants) was annotated as one structure that included the crowns of the teeth and implants at the maxilla and the mandible, respectively. In addition, the boundary of each tooth and implant was independently annotated as one structure. All the annotation processes were performed using the open source labeling software, LabelBox (Labelbox Inc, CA). The oral and maxillofacial (OMF) radiologists also reviewed and corrected the annotated images before training the CNNs.

A baseline study was performed to compare the accuracy with the CNNs that mainly used medical image segmentation technology. To compare the segmentation accuracy, U-Net [32], Dense U-Net [33], SegNet [34], and Mask R-CNN [35], which was used in a previous study, were trained with the same dataset to detect the PBL, CEJL, and teeth/implants on the panoramic radiograph. In addition, Mask R-CNN was trained for detecting the PBL, CEJL, and teeth/implants by using three dataset configurations for the multi-device study. After training all the CNNs, the segmentation accuracy of each CNN was calculated using the images in the test set. All the CNNs were implemented by using Python with the Keras and TensorFlow libraries. The input resolution of the images was modified from the original sizes of the multi-device images to 1024×1024 pixels to improve the training efficiency. The developed CNN was

trained by applying the transfer learning method based on the weights that were calculated in the previous study. Table 2 shows the strategies for training each CNN for the baseline study on GeForce GTX 1080 Ti (NVIDIA, Santa Clara, USA).

After training, the CNNs produced a segmentation mask of the anatomical structures for the input panoramic image. A binary image of the structure that was enclosed by the periodontal bone levels was produced via the binarization of the segmentation mask area and the remaining area. The periodontal bone levels were then detected by extracting the edge of the binary image (Figure 2a-c). The same process was applied for the detection of the CEJL (Figure 2d-f), teeth, and implants from their segmentation mask (Figure 2g-i).

Table 2. The detailed strategies for training the CNNs to detect the periodontal bone level (PBL), the cemento-enamel junction level (CEJL), and the teeth/implants.

Model	U-Net	Dense U-Net	SegNet	Mask R-CNN
Epoch	100	100	100	300
Learning rate	0.001	0.00001	0.001	0.01 (1/10 per 100 epoch)
Batch size	4	4	4	2
Number of GPU	1	1	1	2

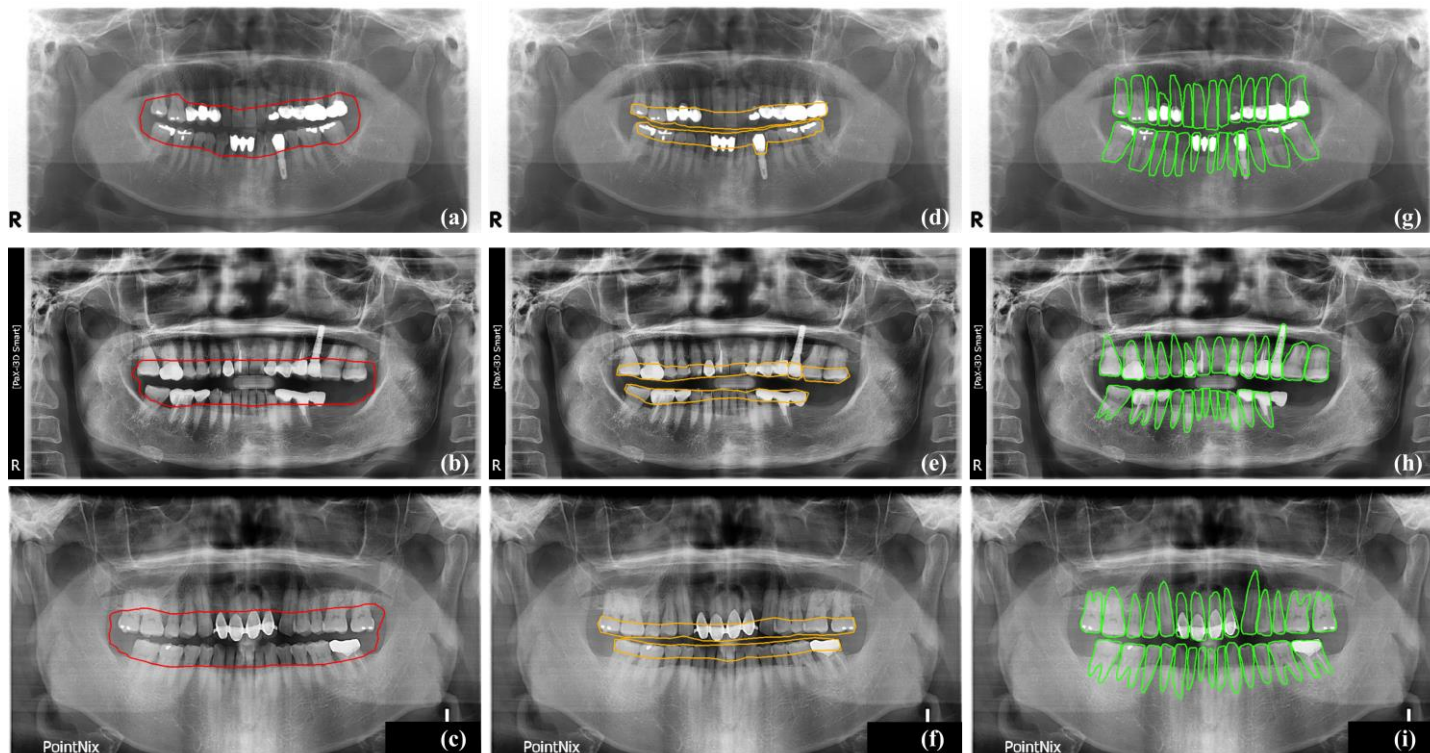


Figure 2. Detection results for the periodontal bone level (PBL) (a-c), the cemento-enamel junction level (CEJL) (d-f), and the teeth/implants (g-i) by the Mask R-CNN. The first, second, and third rows are images from the device 1, device 2, and device 3, respectively.

Detection of the missing teeth using the CNNs

Each image was manually labeled by drawing rectangular bounding boxes around the location of the missing teeth with a labeling software for the bounding box detection task named the YOLO mark [36]. The OMF radiologists also reviewed and corrected the annotated images before training the CNNs. Four modified CNNs from the YOLOv3 [37] and YOLOv4 [38] networks named CNNv3, CNNv3-tiny, CNNv4, and CNNv4-tiny were used for detecting and quantifying the missing teeth on the panoramic radiographs. YOLOv4 was designed by replacing YOLOv3's backbone network from DarkNet [39] to DarkNet with Cross-Stage-Partial-connections (CSP) [40], followed by adding spatial pyramid pooling (SPP) [41] and path aggregation network (PAN) [42] modules. Using the same dataset, the detection accuracy of four CNNs with a different number of parameters and layer depths were compared to each other. A hypothesis that the change in the number of parameters and layer depth affects the accuracy of detecting the missing teeth on the panoramic radiographs was demonstrated through this evaluation. Table 3 shows the detailed configurations of the CNNs.

To detect the missing teeth, the CNNs predicted the bounding boxes, confidence, and classes through a logistic regression [43]. The CNNs were trained using the images with different scales that consisted of 320×320 , 416×416 , and 608×608 pixels, which made it possible to predict the locations of the smaller region. The network was trained on a total of 2,000 epochs with a 64-batch size and one- or two-stride sizes. The network outputted bounding

boxes and the confidence (probability) that the bounding box enclosed a lesion for the input panoramic image after training.

Table 3. The structural configurations and detailed strategies for training the CNNs to detect the missing teeth.

Model	CNNv3-tiny	CNNv3	CNNv4-tiny	CNNv4
Backbone	DarkNet	DarkNet	CSPDarkNet	CSPDarkNet
Layers	23	106	37	161
Learning rate	0.001	0.001	0.001	0.001
Epoch	2,000	2,000	2,000	2,000
Batch size	64	64	64	64
Number of GPU	2	2	2	2

Staging periodontitis by the conventional CAD method

The method of automatically diagnosing the stage of periodontitis was the same as the method that was used in my previously published paper [31]. The principal axis of the tooth or the implant was determined by applying the principal axes of inertia to their boundary images [44-46]. The direction of the minor axis for the minimum moment of inertia was defined as the long-axis orientation of the tooth/implant (Figure 3a-c). Then, two intersection points for the tooth/implant long-axis with the PBL and CEJL (the fixture top level for the implant) that was detected by the CNN was used to calculate the two tooth intersection lengths. These are the distances between the intersection points and the root apex point of the tooth for the maxillary and mandibular teeth, respectively.

The percentage rate of the RBL for the tooth/implant was defined as the ratio of the intersection length of the PBL and the CEJL (Figure 3d-f). Based on the percentage rate, the periodontal bone loss of the tooth was automatically classified to identify the stage of periodontitis according to the criteria that was proposed at the 2017 World Workshop on the Classification of Periodontal and Peri-implant Diseases and Conditions [47] (Figure 3g-i). The staging criteria for periodontitis that is based on the RBL of the tooth are as follows. (1) If the RBL was $< 15\%$ (in the coronal third of the root), the periodontitis was classified as stage one. (2) If the RBL was between 15% and 33% (in the coronal third of the root), it was classified as stage two. (3) If the RBL was $> 33\%$ (extending to the middle third of the root and beyond), it was classified

as stage three. (4) If the sum of the number of detected missing teeth and implants was greater than four under the same condition as stage three, then it was classified as stage four [47].

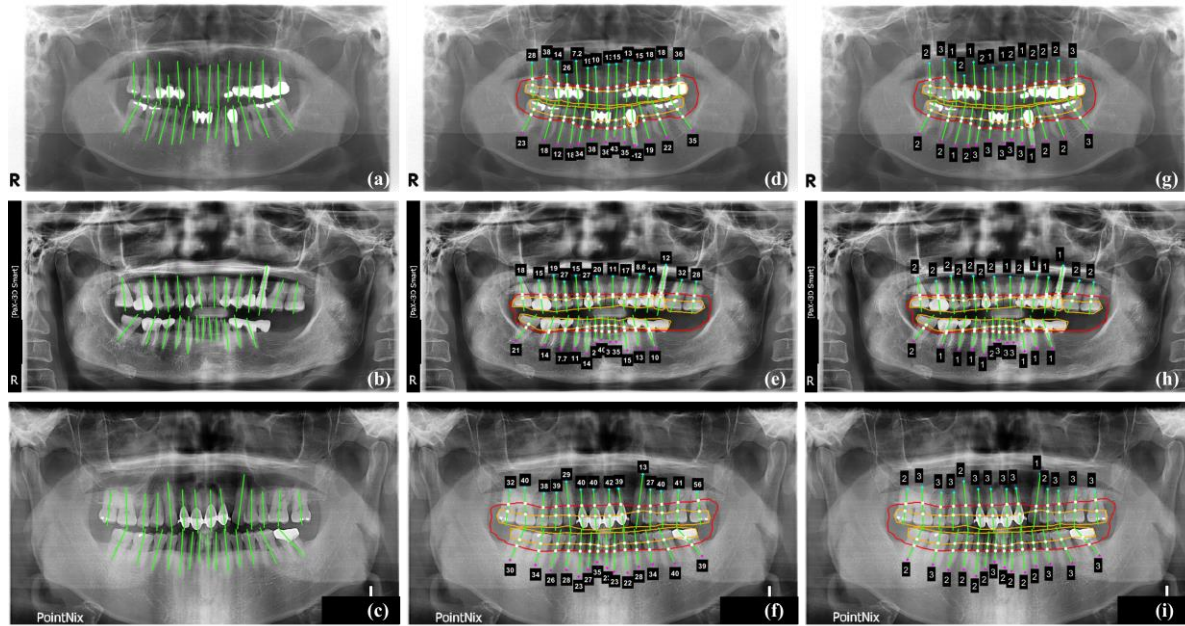


Figure 3. The long-axis orientations of the tooth and the implant (a-c), the intersection points of the tooth (implant) long-axis with the periodontal bone level (PBL) and the cemento-enamel junction level (CEJL) (fixture top level), the percentage rate of the radiographic bone loss (RBL) (d-f), and the stages of the periodontitis for each tooth and implant (g-i). The first, second, and third rows are images from device 1, device 2, and device 3, respectively

Evaluation of the detection and classification performance

To evaluate the detection performance of the PBL, CEJL, and teeth/implants, the dice similarity coefficient (DSC) ($2|S_{gt} \cap S_{det}|/(|S_{gt}| + |S_{det}|)$) was measured between the ground-truth (S_{gt}) and the detected (S_{det}) structures [48-50]. Then, the precision ($TP / (FP + TP)$), recall ($TP / (FN + TP)$), F1-score ($2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$), and mean average precision (AP) values [51, 52] were measured to evaluate the detection performance for the missing teeth. The precision-recall (PR) curve was also computed from the model's detection output by varying the confidence threshold that determined what was considered to be a model-predicted positive detection [51]. The AP was calculated as the average value of the precision across all the recall values.

To evaluate the performance for classifying the periodontitis stages, the automatically determined stages by the developed method were compared to those that were determined by the three OMF radiologists (including a resident with three-years of experience, a fellow with five-years of experience, and a professor with ten-years of experience). The mean absolute difference (MAD) between the stage values that used the developed method and the radiologists' diagnoses was calculated by performing the ANOVA test. In addition, when considering the characteristics of the panoramic radiographic images, such as an inconsistent image magnification and distortion, the teeth were classified by their position and their MAD was calculated. Finally, because alveolar bone loss does not occur independently in a single tooth, the mean stage was

calculated by averaging the automatically determined periodontitis stages for all the teeth in each image. Afterwards, the MAD was calculated by comparing it to the mean stage that was determined by the radiologists.

For the correlation analysis, the Pearson correlation coefficients (PCC) and intraclass correlation coefficients (ICC) between the stage values were calculated to evaluate the correlativity and reliability, respectively using a statistical software program [53, 54] (SPSS, SPSS Inc., Chicago, IL, USA). The Bland and Altman plots [55] were computed to evaluate the degree of agreement between the mean stages that used the developed method and the radiologists' diagnoses. A non-parametric regression analysis was also performed to see whether similar results were obtained between the two methods while applying the Passing–Bablok regression [56].

RESULTS

Detection performance for the anatomical structures

Figures 4 and 5 show the best and worst cases for the detection results, respectively, for the PBL, CEJL, and teeth/implants. These were obtained by the CNNs that used the dental panoramic radiographs acquired through multiple devices. In regard to the images for the worst cases, excluding the Mask R-CNN results, excessively (false positive) or insufficiently (false negative) detected regions were included in the results (Figure 5).

For the baseline study, the DSC values for U-Net, Dense U-Net, SegNet, and Mask R-CNN were 0.86, 0.96, 0.96, and 0.96, respectively, for the detection of the PBL. For the detection of the CEJL, these values were 0.82, 0.91, 0.92, and 0.92, respectively, and, for the detection of the teeth/implants, 0.86, 0.88, 0.88, and 0.94, respectively (Table 4).

For the multi-device study, the DSC values for Mask R-CNN were 0.96, 0.86, and 0.81, for the first, second, and third dataset configurations, respectively, to detect the PBL. For the detection of the CEJL, these values were 0.92, 0.79, and 0.78, respectively, and, for the detection of the teeth/implants, these values were 0.94, 0.79, and 0.75, respectively. The detection accuracies for the second and third dataset configurations did not achieve acceptable levels for use in the automatic diagnosis of the RBL and the periodontitis stage.

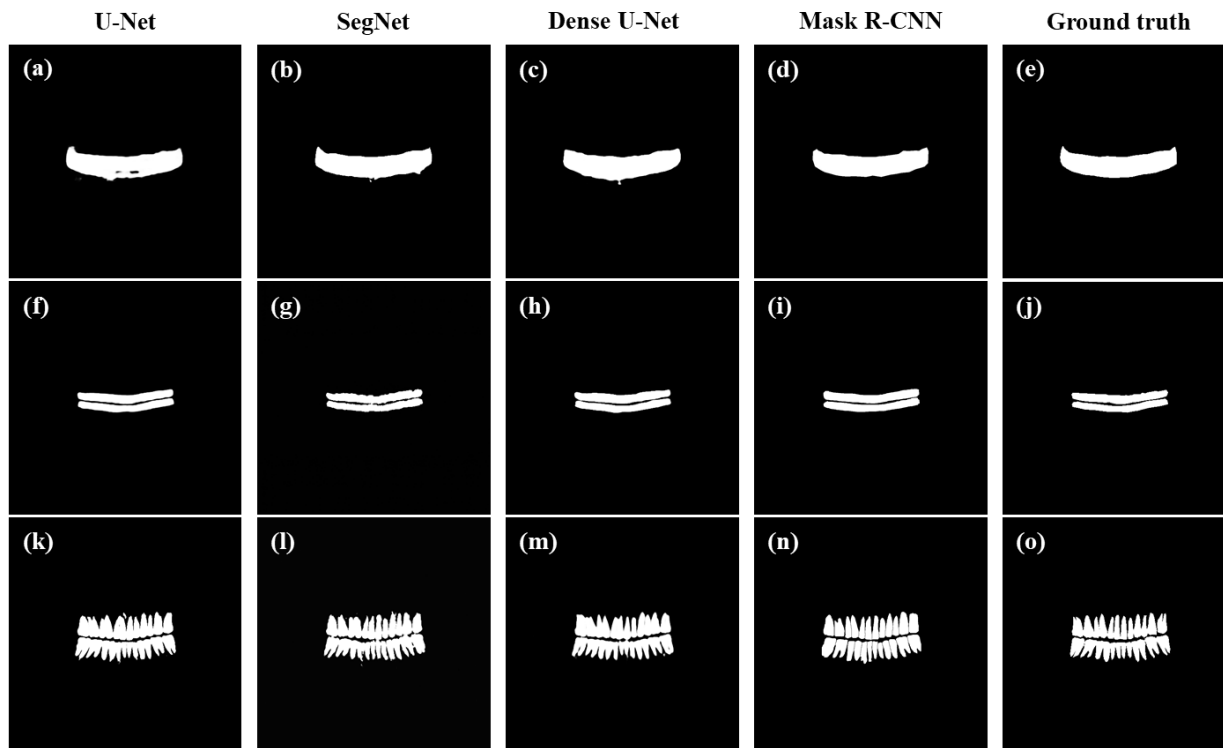


Figure 4. The best cases of detection results for the periodontal bone level (PBL) (a-e), the cemento-enamel junction level (CEJL) (f-j), and the teeth/implants (k-o) by the CNNs.

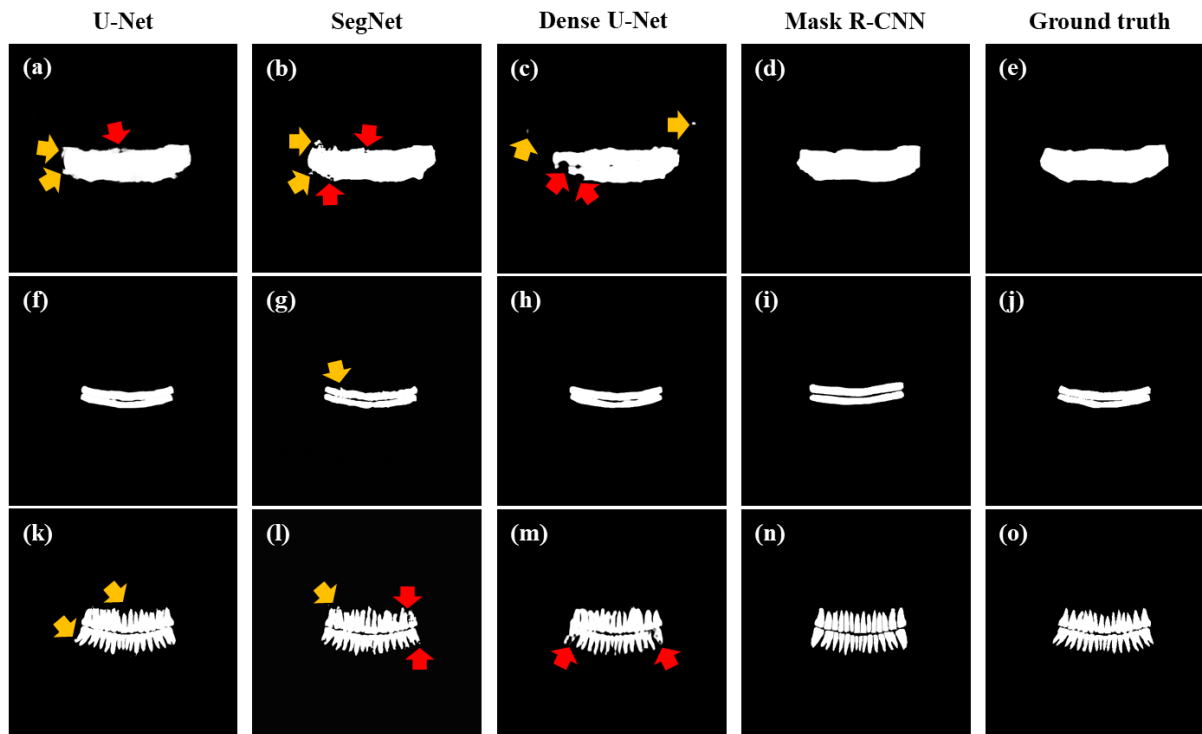


Figure 5. The worst cases of detection results for the periodontal bone level (PBL) (a-e), the cemento-enamel junction level (CEJL) (f-j), and the teeth/implants (k-o) by the CNNs. Locations of the false positive and negative errors are indicated in yellow and red by arrows, respectively.

Table 4. Dice similarity coefficient (DSC) values for detection performance of the periodontal bone level (PBL), the cemento-enamel junction level (CEJL), and the teeth/implants by the CNNs for the multi-device images.

		U-Net	Dense U-Net	SegNet	Mask R-CNN
PBL	Device1	0.88±0.02	0.98±0.02	0.96±0.01	0.95±0.01
	Device2	0.84±0.04	0.94±0.02	0.96±0.01	0.96±0.04
	Device3	0.85±0.03	0.97±0.02	0.95±0.01	0.97±0.01
	Total	0.86±0.03	0.96±0.03	0.96±0.03	0.96±0.01
CEJL	Device1	0.83±0.02	0.92±0.02	0.93±0.01	0.92±0.03
	Device2	0.81±0.02	0.91±0.04	0.91±0.02	0.91±0.10
	Device3	0.82±0.01	0.90±0.02	0.92±0.01	0.92±0.10
	Total	0.82±0.02	0.91±0.03	0.92±0.06	0.92±0.08
Teeth/ Implants	Device1	0.88±0.02	0.90±0.01	0.89±0.05	0.96±0.02
	Device2	0.85±0.01	0.91±0.01	0.87±0.04	0.95±0.01
	Device3	0.84±0.02	0.84±0.02	0.88±0.08	0.91±0.01
	Total	0.86±0.02	0.88±0.03	0.88±0.07	0.94±0.03

Detection performance for missing teeth

Figure 6 shows the detection results for the missing teeth by using CNNv4-tiny, and these were obtained using the dental panoramic radiographs that were acquired with multiple devices from multiple vendors. The bounding boxes were drawn around the positions of the detected missing teeth and they were superimposed on the input images. To quantitatively evaluate the performance of the CNNs, the precision, recall, F1-score, and mean AP values were calculated for the detection of the missing teeth (Table 5). The precision values for CNNv3-tiny, CNNv3, CNNv4-tiny, and CNNv4 were 0.85, 0.85, 0.88, and 0.85, respectively (Table 5). The recall values for CNNv3-tiny, CNNv3, CNNv4-tiny, and CNNv4 were 0.66, 0.83, 0.85, and 0.85, respectively (Table 5). The F1-score values for CNNv3-tiny, CNNv3, CNNv4-tiny, and CNNv4 were 0.74, 0.84, 0.87, and 0.85, respectively (Table 5). The mean of the AP values for CNNv3-tiny, CNNv3, CNNv4-tiny, and CNNv4 were 0.76, 0.83, 0.86, and 0.82, respectively (Table 5). Figure 6 shows the precision-recall (PR) curves for detecting the missing teeth with the CNNs. CNNv4-tiny showed the highest accuracy for detecting the missing teeth on the panoramic radiographs.

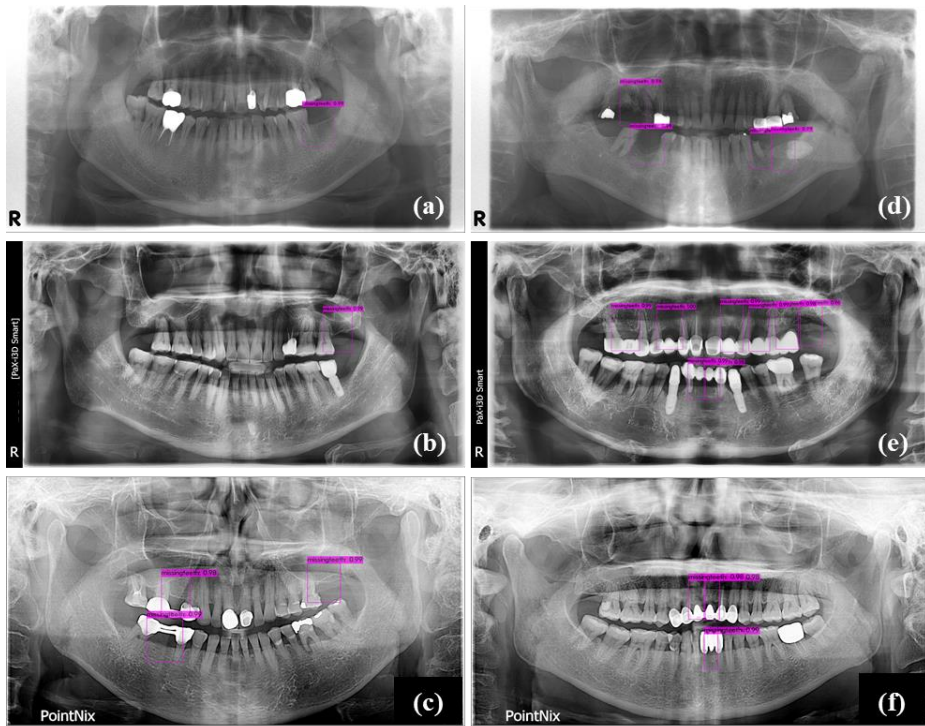


Figure 6. The detection results for the missing teeth by the CNNv4-tiny using dental panoramic radiographs acquired from multiple devices. The detected bounding boxes of the locations of the missing teeth were superimposed on the input images for type 1 (a-c) and type 2 (d-f). The first, second, and third rows are the images from the device 1, device 2, and device 3, respectively.

Table 5. Precision, recall, F1-score, and mean average precision (AP), for detecting the missing teeth on the multi-device images by the CNNs.

	CNNv3-tiny	CNNv3	CNNv4-tiny	CNNv4
Precision	0.85	0.85	0.88	0.85
Recall	0.66	0.83	0.85	0.85
F1-score	0.74	0.84	0.87	0.85
Mean AP	0.76	0.83	0.86	0.82

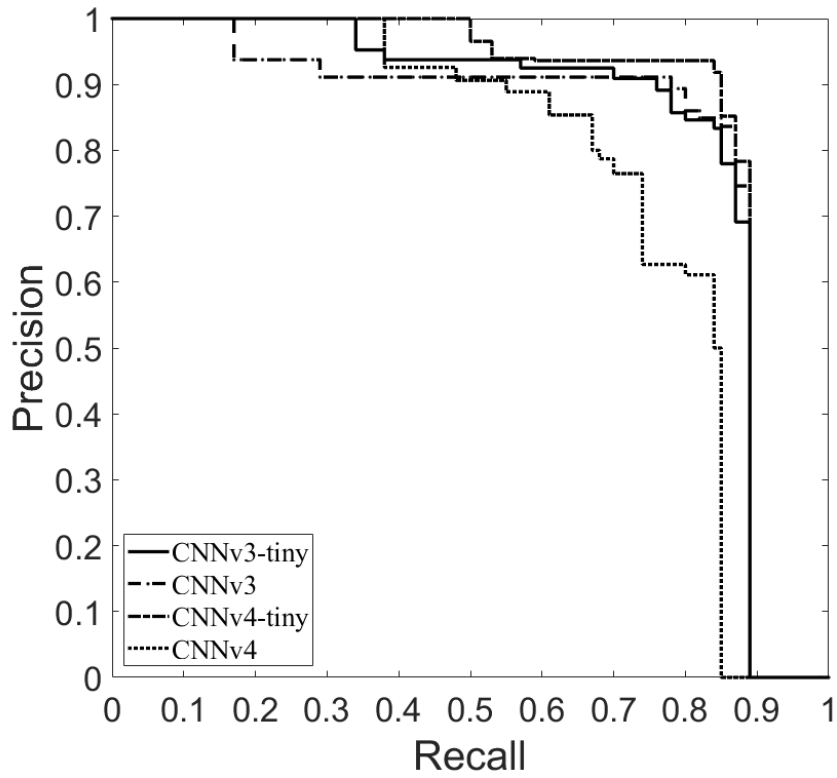


Figure 7. Precision-recall (PR) curves from automatic detection of the missing teeth with CNNv3-tiny, CNNv3, CNNv4-tiny, and CNNv4.

Classification performance for the periodontitis stages

Figure 3 shows the long-axis orientations of the tooth and the implant that was determined by the principal axes of inertia and the intersection points of the tooth's (implant) long-axis with the PBL and CEJL (fixture top level). Figure 3 also shows the percentage rate of the RBL and the completely classified stages of the periodontitis for the teeth/implants according to the criteria that was proposed at the 2017 World Workshop on the Classification of Periodontal and Peri-implant Diseases and Conditions [47]. Figure 8 shows the completely classified stages of the periodontitis for the teeth/implants using the multi-device images. Figures 8e, j, and o show the images of the teeth that are classified as stage four periodontitis because the total number of missing teeth and implants was five or more.

To evaluate the classification performance of the periodontal bone loss, the mean absolute differences (MAD) between the stages that were classified by the automatic method and the radiologists' diagnoses were compared. These MAD values were 0.26, 0.31, and 0.35 for the radiologists with ten-years, five-years, and three-years of experience, respectively, for the teeth of the whole jaw (Table 6). The overall MAD between the stages with the automatic method and the radiologists while using all the images was 0.31. For the images from multiple devices, the MAD values were 0.25, 0.34, and 0.35 for device 1 [31], device 2, and device 3, respectively, for the teeth/implants for the whole jaw (Table 6). Table 7 shows the MAD values with respect to the types of teeth.

These overall MAD values for the incisors, canines, premolars, and molars were 0.37 ± 0.44 , 0.30 ± 0.40 , 0.30 ± 0.45 , and 0.31 ± 0.44 , respectively.

The MAD for the incisors was slightly higher than those for the rest of the types of teeth, but there were no significant statistical differences ($p > 0.05$). Table 8 shows the MAD values of the difference between the average stages for one image. These MAD values were 0.13, 0.16, and 0.25 for the radiologists with ten-years, five-years, and three-years of experience, respectively, for all the images, and 0.10, 0.21, and 0.24 for the device 1, device 2, and device 3 images, respectively. The overall MAD value between the average stages of the images obtained by the automatic method and by all the radiologists was 0.18. The stages that were classified by the automatic method were not significantly different from those that were diagnosed by the radiologists ($p > 0.05$).

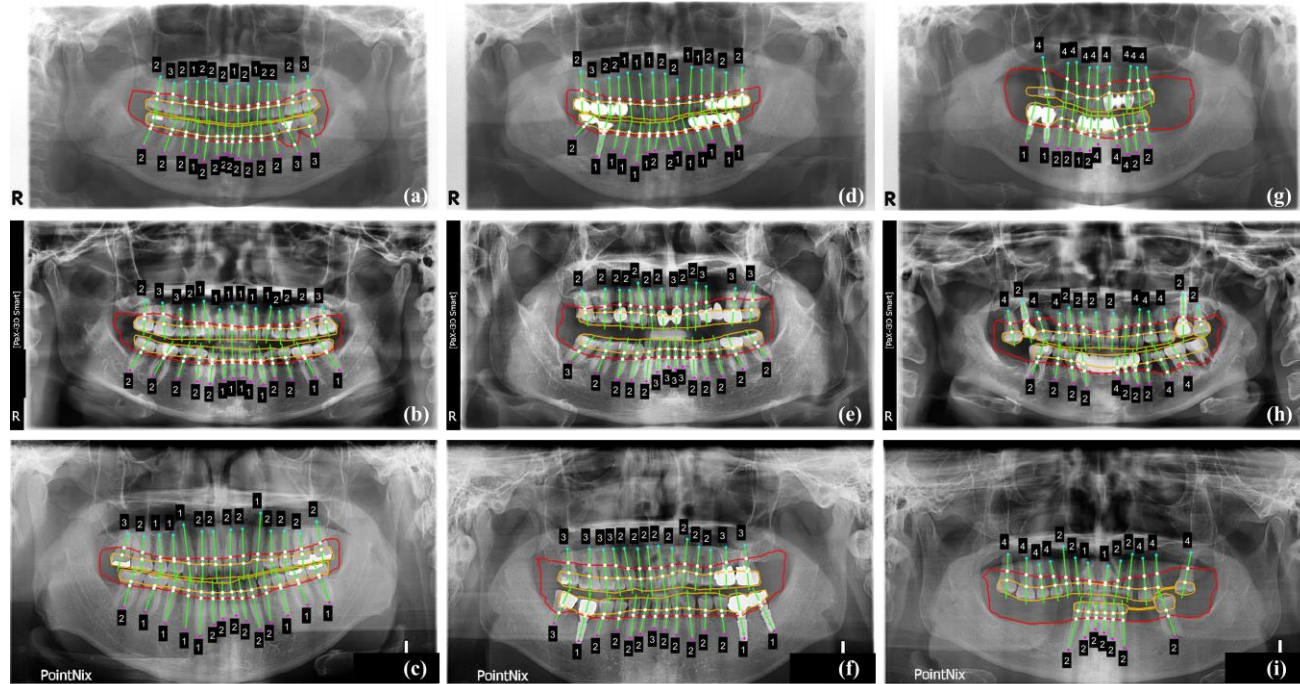


Figure 8. The stages of the periodontitis for each tooth and implant on the dental panoramic radiographs acquired from multiple devices. The automatic diagnosis results by the developed method on the images. The first, second, and third rows are the images from the device 1, device 2, and device 3, respectively.

Table 6. The mean absolute differences (MADs) between periodontitis stages obtained using the automatic method and those diagnosed by the radiologists (with ten-years, five-years, and three-years of experience) using the multi-device images.

Device1 [31]			
	Maxilla	Mandible	Whole jaw
Rad. - 10yrs	0.23±0.19	0.19±0.12	0.21±0.12
Rad. - 5yrs	0.26±0.17	0.25±0.13	0.25±0.11
Rad. - 3yrs	0.29±0.10	0.21±0.11	0.25±0.07
Mean	0.27±0.45	0.23±0.43	0.25±0.44
Device2			
Rad. - 10yrs	0.26±0.13	0.31±0.16	0.30±0.13
Rad. - 5yrs	0.33±0.19	0.26±0.11	0.38±0.20
Rad. - 3yrs	0.39±0.15	0.49±0.24	0.37±0.14
Mean	0.33±0.13	0.35±0.13	0.34±0.10
Device3			
Rad. - 10yrs	0.34±0.13	0.25±0.11	0.28±0.13
Rad. - 5yrs	0.44±0.21	0.31±0.19	0.29±0.15
Rad. - 3yrs	0.35±0.15	0.42±0.27	0.45±0.25
Mean	0.37±0.14	0.33±0.16	0.35±0.09
Overall			
Rad. - 10yrs	0.27±0.16	0.25±0.14	0.26±0.14
Rad. - 5yrs	0.34±0.20	0.27±0.14	0.31±0.18
Rad. - 3yrs	0.34±0.14	0.37±0.24	0.35±0.19
Mean	0.32±0.17	0.30±0.19	0.31±0.18

Rad. = radiologist; yrs = years of experience

Table 7. The mean absolute differences (MADs) between periodontitis stages for types of teeth obtained using the automatic method and those diagnosed by the radiologists (with ten-years, five-years, and three-years of experience) using the multi-device images.

Device1				
	Incisors	Canines	Premolars	Molars
Rad. - 10yrs	0.23±0.30	0.28±0.35	0.18±0.36	0.26±0.37
Rad. - 5yrs	0.29±0.40	0.18±0.31	0.21±0.36	0.39±0.51
Rad. - 3yrs	0.26±0.39	0.20±0.27	0.27±0.47	0.32±0.43
Mean	0.26±0.36	0.22±0.31	0.22±0.40	0.32±0.44
Device2				
Rad. - 10yrs	0.38±0.48	0.20±0.35	0.26±0.48	0.28±0.42
Rad. - 5yrs	0.33±0.48	0.30±0.50	0.21±0.42	0.35±0.45
Rad. - 3yrs	0.41±0.48	0.35±0.46	0.53±0.60	0.41±0.57
Mean	0.37±0.48	0.28±0.44	0.33±0.50	0.35±0.48
Device3				
Rad. - 10yrs	0.45±0.52	0.30±0.40	0.31±0.48	0.38±0.50
Rad. - 5yrs	0.45±0.41	0.38±0.38	0.39±0.52	0.28±0.47
Rad. - 3yrs	0.56±0.52	0.53±0.56	0.38±0.38	0.15±0.23
Mean	0.49±0.48	0.40±0.45	0.36±0.46	0.23±0.37
Overall				
Rad. - 10yrs	0.35±0.43	0.26±0.37	0.25±0.44	0.28±0.41
Rad. - 5yrs	0.35±0.43	0.28±0.40	0.27±0.43	0.34±0.48
Rad. - 3yrs	0.41±0.46	0.36±0.43	0.39±0.48	0.29±0.41
Mean	0.37±0.44	0.30±0.40	0.30±0.45	0.31±0.44

Rad. = radiologist; yrs = years of experience

Table 8. The mean absolute differences (MADs) between mean periodontitis stage of image obtained using the automatic method and those diagnosed by the radiologists (with ten-years, five-years, and three-years of experience) using the multi-device images.

	Device1	Device2	Device3	Overall
Rad. - 10yrs	0.06±0.07	0.15±0.11	0.20±0.13	0.13±0.12
Rad. - 5yrs	0.11±0.09	0.19±0.12	0.19±0.14	0.16±0.12
Rad. - 3yrs	0.12±0.09	0.30±0.18	0.34±0.22	0.25±0.19
Mean	0.10±0.05	0.21±0.09	0.24±0.10	0.18±0.10

Rad. = radiologist; yrs = years of experience

Classification performance of the correlations, regressions, and agreements between the periodontitis stages

Table 9 shows the PCC values of the automatic method and the radiologists' diagnoses. The overall PCC values between the developed method and the radiologists' diagnoses were 0.73, 0.77, and 0.75 for the images from device 1 [31], device 2, and device 3, respectively ($p < 0.01$). The final ICC value between the developed method and the radiologists' diagnoses for all the images was 0.76 ($p < 0.01$). The PCC between the diagnoses of the automatic method and the radiologist with ten-years of experience showed the highest correlation. This PCC showed a strong correlation between the radiologists and the automatic diagnoses.

Table 10 shows the ICC values of the automatic method and the radiologists' diagnoses. The ICC between the diagnoses of the automatic method and the radiologist with ten-years of experience showed the highest correlation. The automatic classification in the developed method had a high reliability with the radiologists' diagnoses. The overall ICC values between the developed method and the radiologists' diagnoses were 0.91, 0.94, and 0.93 for the images from device 1 [31], device 2, and device 3, respectively ($p < 0.01$). The final ICC value between the developed method and the radiologists' diagnoses for all the images was 0.93 ($p < 0.01$). This ICC value indicated an excellent reliability of the automatic diagnoses for the RBL and the periodontitis stage.

Figures 9 and 10 show the results of the Passing and Bablok regression analysis and the Bland and Altman analysis, respectively. For the Passing and Bablok analysis, the closer the slope of the regression line is to one (identity), the higher the agreement between the two methods. In the Passing and Bablok plots, the slopes were 1.176 ($p > 0.05$), 1.100 ($p > 0.05$), and 1.111 ($p > 0.05$) with the intersections of -0.304, -0.199, -0.371 for the radiologists with ten-years, five-years, and three-years of experience, respectively. In the Bland and Altman plots, the average of the difference between the mean stages that were classified by the automatic method and those that were diagnosed by the radiologists with ten-years, five-years, and three-years of experience were 0.007 (95 % confidence interval (CI), -0.060 ~ 0.074), -0.022 (95 % CI, -0.098 ~ 0.053), and -0.198 (95 % CI, -0.291 ~ -0.104), respectively. These values indicated that the mean stages that were classified by the automatic method were not significantly different from those that were diagnosed by the radiologists ($p > 0.05$).

Table 9. The Pearson correlation coefficient (PCC) between stages obtained using the automatic method and those diagnosed by the radiologists (with ten-years, five-years, and three-years of experience) using the multi-device images.

Device1 [31]				
	Automatic method	Rad. - 10yrs	Rad. - 5yrs	Rad. - 3yrs
Automatic method	1	0.76*	0.73*	0.70*
Rad. - 10yrs	0.76*	1	0.72*	0.70*
Rad. - 5yrs	0.73*	0.72*	1	0.70*
Rad. - 3yrs	0.70*	0.70*	0.70*	1
Device2				
Automatic method	1	0.82*	0.77*	0.73*
Rad. - 10yrs	0.82*	1	0.85*	0.81*
Rad. - 5yrs	0.77*	0.85*	1	0.86*
Rad. - 3yrs	0.73*	0.81*	0.86*	1
Device3				
Automatic method	1	0.77*	0.72*	0.77*
Rad. - 10yrs	0.77*	1	0.79*	0.80*
Rad. - 5yrs	0.72*	0.79*	1	0.75*
Rad. - 3yrs	0.77*	0.80*	0.75*	1
Overall				
Automatic method	1	0.80*	0.75*	0.74*
Rad. - 10yrs	0.80*	1	0.81*	0.78*
Rad. - 5yrs	0.75*	0.81*	1	0.79*
Rad. - 3yrs	0.74*	0.78*	0.79*	1

* = significant at $p < 0.01$; Rad. = radiologist; yrs = years of experience

Table 10. The intraclass correlation coefficient (ICC) between stages obtained using the automatic method and those diagnosed by the radiologists (with ten-years, five-years, and three-years of experience) using the multi-device images.

Device1 [31]				
	Automatic method	Rad. - 10yrs	Rad. - 5yrs	Rad. - 3yrs
Automatic method	1	0.86*	0.84*	0.82*
Rad. - 10yrs	0.86*	1	0.84*	0.82*
Rad. - 5yrs	0.84*	0.84*	1	0.82*
Rad. - 3yrs	0.82*	0.82*	0.82*	1
Device2				
Automatic method	1	0.90*	0.87*	0.84*
Rad. - 10yrs	0.90*	1	0.92*	0.90*
Rad. - 5yrs	0.87*	0.92*	1	0.92*
Rad. - 3yrs	0.84*	0.90*	0.92*	1
Device3				
Automatic method	1	0.87*	0.83*	0.87*
Rad. - 10yrs	0.87*	1	0.88*	0.86*
Rad. - 5yrs	0.83*	0.88*	1	0.86*
Rad. - 3yrs	0.87*	0.86*	0.86*	1
Overall				
Automatic method	1	0.89*	0.86*	0.85*
Rad. - 10yrs	0.89*	1	0.89*	0.88*
Rad. - 5yrs	0.86*	0.89*	1	0.88*
Rad. - 3yrs	0.85*	0.88*	0.88*	1

* = significant at $p < 0.01$; Rad. = radiologist; yrs = years of experience

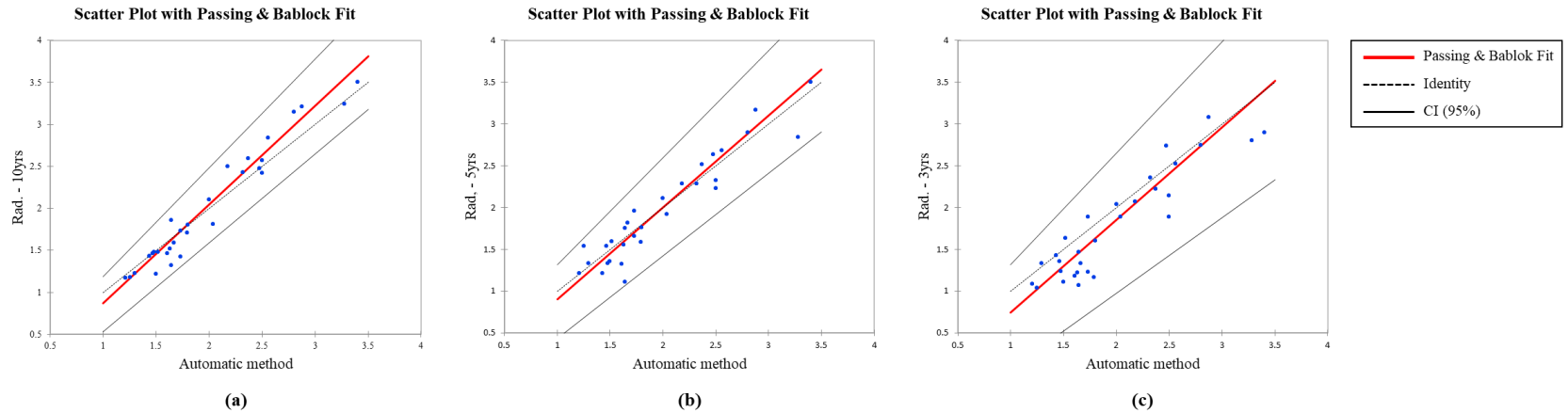


Figure 9. Scatter plots with Passing and Bablock regression of mean periodontitis stages obtained using the automatic method and those diagnosed by the radiologists with ten-years (a), five-years (b), and three-years (c) of experience.

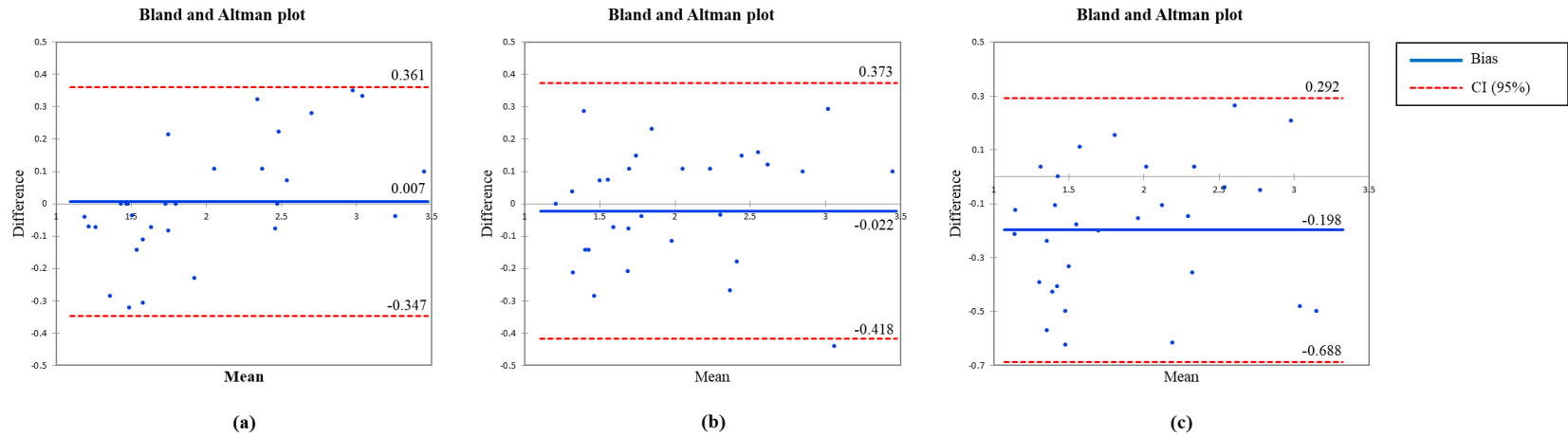


Figure 10. Bland and Altman plots of mean periodontitis stages obtained using the automatic method and those diagnosed by the radiologists with ten-years (a), five-years (b), and three-years (c) of experience.

DISCUSSION

CAD is a computerized method that can provide a second objective opinion to medical staff because of its ability to interpret medical images [57]. Recent CAD approaches were combined with deep learning, a subset of machine learning, to extract the relevant features from the images without using cropped patches [7]. The CNN, which is one of the most established algorithms for deep learning, automatically and adaptively learns the spatial configurations of the image features by using multiple types of layers. [7-9]. During medial image analysis, the CNNs were mainly implemented on a variety of CAD applications to classify, detect, and segment the features.

One key task for radiologists is to make an appropriate differential diagnosis using each patient's medical images. This classification task includes a wide range of applications from determining the presence or absence of a disease to identifying the type of malignancy. One early application was introduced to classify the lung nodules on the chest X-rays by using a simple CNN with two hidden layers [58]. With the development of computing power, deeper and more complex CNN models have been introduced to solve the classification problems in medical imaging analysis. A variety of modified applications that are based on the networks with convolutional layers, such as LeNet [59], ImageNet [60], GoogLeNet [61], visual geometry group (VGG) Net [62], DenseNet [63], ResNet [64], and their derivatives [65-67] were

developed to solve the classification problems for medical images with different modalities.

Automatic detection is performed to identify and localize the regions with diseases, lesions, or anatomical structures in medical imaging analysis [68]. CNNs for the detection tasks are basically similar to those that are used for classification tasks. However, several layers with additional functions, such as a region proposal or regression, were added to the CNNs for detecting diseases or lesions. A variety of models, such as a simple structured CNN with a support vector machine (SVM) [69] to GoogLeNet [61], Inception [70], DetectNet [71], and YOLO [36] were used to detect diseases from the medical images.

CNNs have been used to segment different anatomical structures or lesions in images with a variety of imaging modalities, which includes single-view X-ray, CT, MR, and ultrasound images [72]. Regions with CNN (R-CNN) and its derivatives, such as Fast [73], Faster [74] and Mask R-CNN [35], are representative works for the region-based methods in medical image segmentation. Mask R-CNN is an extension of the Faster R-CNN, in which a branch is added that predicts the object mask in parallel with the existing branch for bounding box recognition [35]. Fully convolutional network (FCN)-based methods are able to learn mapping without extracting the region proposals [75]. The FCNs only had convolutional and pooling layers, which gave them the ability to make predictions on arbitrary-sized inputs [75]. The most popular network architecture for segmentation was U-net [32], which consisted of several convolution layers, followed by deconvolution layers with skip

connections. The derivatives of U-net, such as SegNet [34], V-Net [76], and 3D U-Net [77] have been used for multiple segmentation tasks on medical images.

In dental imaging, various imaging modalities, such as periapical, cephalometric, panoramic radiographs, as well as cone-beam computed tomography (CBCT), have been used to detect and classify diseases with CNN-based CAD approaches. A CNN named VGG-19 was trained to diagnose periodontally-compromised teeth on the periapical images [78]. Some CAD studies have used deep CNN analysis of the panoramic radiographs for the segmentation of teeth on panoramic images [20]. These studies also evaluated the root morphology of the mandibular first molar [25], detected sinusitis by learning the form of the maxillary sinus [26], and diagnosed osteoporosis by evaluating the cortical erosion of the mandibular inferior cortex [79].

In 2017, the American Academy of Periodontology and the European Federation of Periodontology provided a new diagnostic framework for periodontitis based on a multidimensional stage and grade system [47]. The stage related to the severity and extent of periodontitis should be determined via the CAL, or the RBL if CAL is not available [47]. The stages from one to three can be determined from the four stages through CAL or RBL [47]. This grading also allowed for an assessment of the rate of periodontitis progression, which is mainly determined by the primary criteria that consists of direct and indirect evidence of progression [47]. The RBL evaluation is a highly valuable tool in assessing periodontal health [47, 80-82]. Panoramic radiographs and intra-oral periapical radiographs have been mainly used for RBL evaluation.

However, panoramic radiographs are the best modality to screen the lesions of the whole jaw. In addition, a previous study demonstrated a high level of agreement between intra-oral periapical and panoramic radiograph readings of the distances between the CEJL and the PBL as well as the proportional values in relation to the root length [83]. Manual measurement of the RBL for all teeth in a panoramic radiograph is very time-consuming and labor-intensive. In addition, the conventional intra-oral technique is expensive, time-consuming, and disagreeable to the patients, and the radiation doses in intra-oral radiography are not negligible. Therefore, an easy and automated method is essential to evaluate the RBL and to accurately stage the periodontitis on the panoramic radiographs. In the previous studies, the CNN based methods were proposed to detect the RBL on the dental panoramic radiographs without quantifying the RBL and classifying the periodontitis stage [27, 28]. In my previous study, an automatic method for diagnosing periodontal bone loss from dental panoramic radiographs was developed to stage the periodontitis [31] according to the new criteria that was proposed at the 2017 World Workshop [47] for the first time.

In this study, a deep learning method was developed to automatically detect the anatomical structures, such as the PBL, CEJL, and teeth/implants. Then, the conventional CAD method was also developed to classify the periodontal bone loss by quantifying the number of missing teeth.

The purpose of a baseline study is to determine the most suitable network (or model) for solving the target problem. In this study, three baseline

CNNs and Mask R-CNN were implemented to segment the anatomical structures. Regarding the quantitative results for detecting the PBL, CEJL, and teeth/implants from the baseline study, except for the results of U-Net, the DSC values of the CNNs were similar to each other (Table 4). Meanwhile, in the worst case for the qualitative evaluation results of the CNNs, excluding the developed CNN results, excessively or insufficiently detected regions were included in the results for the three remaining CNNs (Figure 5). Consequently, both false positive and negative errors were included in the results of the three CNNs that were used as the baseline. As a result of the baseline study, the Mask R-CNN was finally used for detecting the PBL, CEJL, and teeth/implants.

To improve the generalization ability of the deep CNN for detecting the PBL, CEJL, and teeth/implants, a multi-device study was performed on an additional dataset [84]. A total of 500 panoramic radiographs (400, 50, and 50 images for device 1, device 2, and device 3, respectively) from multiple devices were collected to train the CNN. Because the CNN learns statistical regularity that is specific to the training set, if the data for training the CNN are insufficient, the results show a low accuracy for a new data set [85]. The best solution to overcome this problem is to use a large amount of training images from multiple devices. However, in medical imaging, collecting and annotating a large dataset is not always easy. Transfer learning is the best solution to overcome this problem. Transfer learning is a machine learning method where a network that was developed for a former task is reused as the starting point for a network in a later task [86]. That is, the latter network uses a transferred weight (a pre-

trained weight) that is gained while solving a former problem, as an initial weight for training to solve other related problems [87]. The former learning is called the source and the latter learning is called the target [88-90]. The method of training the network with pre-trained weights can result in a faster training time, and it helps to get a sufficient accuracy, even with a relatively small amount of data [87]. In recent times, transfer learning has included a state-of-the-art multi-convolutional layer neural network, which is known as a CNN [91]. A CNN uses transfer learning to improve the detection accuracy of periodontal bone loss on the panoramic radiographs [28]. In this study, a pre-trained weight that was derived from the previous study [31] was used as an initial weight when training the CNN for detecting the PBL, CEJL, and teeth/implants to achieve a high training efficiency. By training the CNN with transfer learning, the CNN can produce accurate detection outputs for the multi-device images with various aspects.

In a previous study, a CNN named DeNTNet that used 12,179 panoramic radiographs had a DSC of 0.75 for the direct detection of periodontal bone loss areas after segmenting the region of interest (ROI) for detection [28]. In another study, Krois et al. used 2,001 cropped image segments of teeth to train a CNN to detect periodontal bone loss for an individual tooth, and it obtained a F1-score of 0.78. Meanwhile in my previous study, for the detection accuracies with the Mask R-CNN, the values of the dice coefficient were 0.93, 0.91, and 0.91 for the PBL, CEJL, and teeth, respectively, while using 330 panoramic radiographs [31]. In this study, by increasing the number of training

images from 330 to 500, for the same CNN, the values of the dice coefficient improved to 0.96, 0.92, and 0.94 for the PBL, CEJL, and teeth/implants, respectively by using the images of the first dataset configuration. These values were obtained without preprocessing the images, which slightly improved the results. However, for the second and third dataset configurations, the detection accuracies did not achieve acceptable levels for use in the automatic diagnosis of the RBL and the periodontitis stage. Therefore, additional images must be acquired from devices 2 and 3 and they should be used to train the CNNs. In addition, by collecting more images with different devices and training the CNNs, a sufficient detection accuracy can be obtained for use in the automatic diagnosis of the RBL and the periodontitis stage.

Despite a relatively small number of panoramic radiographs to train the CNNs, the developed method can achieve a high detection performance of the images from all the devices. Considering that the PBL is a simple closed-curve structure for the whole jaw according to a previous study [31], this decreases the complexity of the alveolar bone destruction patterns of the tooth. This method allows the detection of the PBL by ignoring a variety of destructive structures (laterals) such as the horizontal bone loss, vertical/angular defects, osseous craters, and furcation involvement [92]. Using a similar process for detecting the CEJLs (fixture top of the implants), the simple closed-curve structures of the CEJLs were detected at the maxilla and the mandible. Through this process, it was possible to reduce the diversity of the anatomical structures of the multi-device images of the patients. In addition, the alveolar bone

(periodontal bone) has a clear boundary in the panoramic radiograph; thus, it is easy to distinguish it from the oral cavity. In other words, the borderline between the alveolar bone and the oral cavity can be clearly distinguished, even in images that are obtained under different conditions in various devices. Similarly, a tooth (implant) and its anatomical components can be clearly distinguished in the panoramic radiograph. From this, the CEJLs and the teeth/implants can easily be distinguished by their structures. As a result, the CNN learned the simplified and clearly distinguishable anatomical structures with a lower variation and it obtained high detection accuracies. In other words, the CNN accurately performed the detection tasks regardless of the different imaging conditions, pre-processing methods, and contrasts. Despite the false negative errors that mainly occurred along both lateral sides of the enclosed oral cavity when detecting the PBL and CEJLs, both sides were irrelevant areas for determining the true PBL and CEJL and diagnosing the RBL.

The former method in my previous study automatically classified the periodontal bone loss of the individual tooth in order to identify the periodontitis stage based on the percentage rate analysis through conventional CAD processing [31]. The percentage rate of the RBL for each tooth (implant) was computed by calculating the ratio of the intersection length of the PBL and the CEJL for each tooth. The tooth intersection lengths are defined by the distances from the root apex point of the tooth to the two intersection points of the tooth's (implant) long-axis with the periodontal bone and the CEJL (the fixture top level for the implant). The long-axis orientation of the tooth, or the

implant, was determined by applying the principal axes of inertia to their boundary images. According to the periodontitis stage definition that was proposed at the 2017 World Workshop, periodontitis is classified into four stages. However, the previous method was unable to diagnose periodontitis stage four because it could not detect and quantify the locations of the missing teeth. Presently, no studies have been able to detect and quantify the missing teeth on the panoramic radiographs. In this study, four CNNs that were modified from the YOLOv3 [37] and YOLOv4 [38] networks were implemented to detect and quantify the missing teeth on the panoramic radiographs. CNNv4-tiny, which has 37 layers, achieved a mean AP value of 0.86, which was the best performance on the test set. That is, for simple tasks like performing a single-class detection, it might be more efficient to use a network with fewer parameters and layers in terms of speed and accuracy [93, 94].

The MAD between the periodontitis staging that was performed by the automatic method and the radiologists was 0.31 overall for all the teeth in the jaw. The classification accuracies for the images from multiple devices were 0.25, 0.34, and 0.35 for device 1 [31], device 2, and device 3, respectively. The imbalanced dataset was used to train the developed CNN for detecting the PBL, CEJL, and teeth/implants. To overcome the dataset imbalance problem, the developed CNN was trained using transfer learning, and the CNN achieved a high detection accuracy (Table 4). However, the errors that are caused by the imbalance of the amount of data could not be completely solved, and this must

be overcome by performing further research. In addition, the classification accuracy for the incisors and molars was lower than the canines and premolars ($p < 0.05$). This is because the incisor roots were blurred by their overlapping with the vertebrae (Table 7).

The correlation analysis was performed to evaluate the diagnostic correlativity and reliability of the developed method. The PCC values between the automatic method and the radiologists were 0.73, 0.77, and 0.75 for device 1 [31], device 2, and device 3, respectively, overall for the whole jaw (Table 9). The ICC values were 0.91, 0.94, and 0.93 for device 1 [31], device 2, and device 3, respectively, overall for the whole jaw (Table 10). These correlation values were highest between the automatic method and the radiologist with the most experience. Therefore, the automatic method for staging periodontitis had a high accuracy and an excellent diagnostic reliability. In other words, it was demonstrated that the method of training the developed CNN with transfer learning contributed to the improvement of the accuracy and reliability of the automatic diagnosis.

The regression and difference analyses were also performed to evaluate the agreement between the mean periodontitis stages of an image that is classified by the automatic method and diagnosed by the radiologists. Passing and Bablok proposed a regression model to compare the methods that fit the regression line from two data sets [56]. The result of the regression can be interpreted to evaluate the agreement of the two methods. The result consists of a scatter diagram and a regression line that enables the visual inspection of the

agreement. The equation ($y = a + bx$) of the regression line means a constant (regression line's intercept (a)) and a proportional (regression line's slope (b)) difference with confidence intervals (95 % CI) [95].

Figure 9 shows the results of the Passing and Bablock regression analysis. The slope coefficients of the regression lines were 1.176, 1.100, and 1.111, with the intersections of -0.304, -0.199, -0.371 for the radiologists with ten-years, five-years, and three-years of experience, respectively. The agreement with the radiologist with five-years of experience was slightly higher than that with the radiologist with ten-years of experience according to the coefficients and intersections. However, the deviations between the upper and lower bounds of the slope coefficient were 0.237, 0.294, 0.429, for the radiologists with ten-years, five-years, and three-years of experience, respectively. This means that the deviations were the lowest between the mean stages that were determined by the automatic method and by the most experienced radiologist.

Bland and Altman developed a scattered plot to compute the difference between two quantitative measurements by constructing the limits of agreement [55, 96]. The difference between the paired results was computed and plotted against the mean of the paired measurements. If the difference was close to zero, this indicates that the methods are systematically producing similar results [96]. In the Bland and Altman plots (Figure 10), the average of the difference between the mean stages that are classified by the automatic method and those diagnosed by the radiologists with ten-years, five-years, and three-years of

experience were 0.007, -0.022, and -0.198, respectively. Therefore, the results indicate that the mean stages that were determined by the developed method showed a high level of agreement with those that were diagnosed by the radiologists. In addition, the automatically determined mean periodontitis stages showed the highest level of agreement with the most experienced radiologist.

The developed method used the percentage rate of the periodontal bone loss to automatically and classify the periodontitis into four stages of the whole jaw according to the renewed criteria that was proposed at the 2017 World Workshop [47]. The method achieved a high detection performance of the PBL (or the CEJL) by simplifying the complexity of the bone destruction patterns. There was also a high classification performance for automatic periodontitis staging by using the conventional CAD approach. The framework that combined the deep learning architecture and conventional CAD approaches demonstrated a high accuracy, reliability, and agreement in the automatic and complete diagnosis of staging the periodontitis. Consequently, this approach will help dental professionals to definitively diagnose and treat periodontitis. However, this method cannot provide the absolute value of the periodontal bone loss because of the inconsistent image magnification and distortion in dental panoramic radiography. In addition, according to the criteria that was proposed in the 2017 workshop, the stage of periodontitis was diagnosed by considering the severity and complexity [4]. However, the developed method classified the periodontitis stage by considering only the severity factors. Finally, the

classification accuracy for the incisors and molars was lower than the other types of teeth. In future investigations, the method must be improved to diagnose the stage of periodontitis while considering both the severity and complexity factors. In addition, the deep learning-based method for the elimination of the cervical vertebrae will be developed to improve the classification accuracy for the incisors.

In conclusion, the developed method can help dental professionals to diagnose and monitor periodontitis systematically and precisely on panoramic radiographs. This may substantially improve the performance of dental professionals when diagnosing and treating periodontitis.

REFERENCES

1. Tonetti, M.S., et al., *Impact of the global burden of periodontal diseases on health, nutrition and wellbeing of mankind: A call for global action*. Journal of Clinical Periodontology, 2017. **44**(5): p. 456-462.
2. Armitage, G.C., *Development of a classification system for periodontal diseases and conditions*. Annals Periodontology, 1999. **4**(1): p. 1-6.
3. Caton, J.G., et al., *A new classification scheme for periodontal and peri-implant diseases and conditions - Introduction and key changes from the 1999 classification*. Journal of Periodontology, 2018. **89**: p. S1-S8.
4. Tonetti, M.S., H. Greenwell, and K.S. Kornman, *Staging and grading of periodontitis: Framework and proposal of a new classification and case definition*. Journal of Clinical Periodontology, 2018. **45 Suppl 20**: p. S149-S161.

5. Chan, H.P., L.M. Hadjiiski, and R.K. Samala, *Computer-aided diagnosis in the era of deep learning*. Medical Physics, 2020. **47**(5): p. E218-E227.
6. Kwon, O., et al., *Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network*. Dentomaxillofacial Radiology, 2020. **49**(8): 20200185.
7. Yamashita, R., et al., *Convolutional neural networks: an overview and application in radiology*. Insights Imaging, 2018. **9**(4): p. 611-629.
8. Yonekura, A., et al., *Automatic disease stage classification of glioblastoma multiforme histopathological images using deep convolutional neural network*. Biomedical Engineering Letters, 2018. **8**(3): p. 321-327.
9. Billah, M. and S. Waheed, *Gastrointestinal polyp detection in endoscopic images using an improved feature extraction method*. Biomedical Engineering Letters, 2018. **8**(1): p. 69-75.
10. Kallenberg, M., et al., *Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk Scoring*.

- IEEE Transactions on Medical Imaging, 2016. **35**(5): p. 1322-1331.
11. Zhao, X.M., et al., *A deep learning model integrating FCNNs and CRFs for brain tumor segmentation*. Medical Image Analysis, 2018. **43**: p. 98-111.
 12. Esteva, A., et al., *Dermatologist-level classification of skin cancer with deep neural networks*. Nature, 2017. **542**(7639): p. 115-118.
 13. Gao, X.H.W., R. Hui, and Z.M. Tian, *Classification of CT brain images based on deep learning networks*. Computer Methods and Programs in Biomedicine, 2017. **138**: p. 49-56.
 14. Teramoto, A., et al., *Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique*. Medical Physics, 2016. **43**(6): p. 2821-2827.
 15. Hannun, A.Y., et al., *Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network*. Nature Medicine, 2019. **25**(3): p. 65-69.

16. Krizhevsky, A., I. Sutskever, and G.E. Hinton, *ImageNet classification with deep convolutional neural networks*. Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012. **1**: p. 1097–1105.
17. Lee, H., M. Park, and J. Kim, *Cephalometric landmark detection in dental x-ray images using convolutional neural networks*. Proceedings of the Medical Imaging 2017: Computer-aided Diagnosis, 2017. **10134**: 101341W.
18. Ronneberger, O., P. Fischer, and T. Brox, *Dental X-ray image segmentation using a U-shaped Deep Convolutional network*. Proceedings of the International Symposium on Biomedical Imaging, 2015.
19. Miki, Y., et al., *Classification of teeth in cone-beam CT using deep convolutional neural network*. Computers in Biology and Medicine, 2017. **80**: p. 24-29.
20. Jader, G., et al., *Deep instance segmentation of teeth in panoramic X-ray images*. Proceedings of the 31st Conference on Graphics, Patterns and Images (SIBGRAPI), 2018: p. 400-407.

21. Ben, A.R., E. Ridha, and Z. Mourad, *Detection and classification of dental caries in x-ray images using deep neural networks*. Proceedings of the 11th Conference on Software Engineering and Advanced Applications, 2016: p. 223-227.
22. Srivastava, M.M., et al. *Detection of Tooth caries in Bitewing Radiographs using Deep Learning*. arXiv preprint arXiv:1711.07312, 2017.
23. Choi, J., H. Eun, and C. Kim, *Boosting Proximal Dental Caries Detection via Combination of Variational Methods and Convolutional Neural Network*. Journal of Signal Processing Systems for Signal Image and Video Technology, 2018. **90**(1): p. 87-97.
24. Lee, J.H., et al., *Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm*. Journal of Dentistry, 2018. **77**: p. 106-111.
25. Hiraiwa, T., et al., *A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography*. Dentomaxillofacial Radiology, 2019. **48**(3): 20180218.

26. Murata, M., et al., *Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography*. Oral Radiology, 2018. **35**: p. 301-307.
27. Krois, J., et al., *Deep Learning for the Radiographic Detection of periodontal Bone Loss*. Scientific Reports, 2019. **9**(1): 8495.
28. Kim, J., et al., *DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs*. Scientific Reports, 2019. **9**(1): 17615.
29. Chen, C., et al., *Improving the Generalizability of Convolutional Neural Network-Based Segmentation on CMR Images*. Frontiers in Cardiovascular Medicine, 2020. **7**: 105.
30. Chang, H.J., et al., *Deep Learning Hybrid Method to Automatically Diagnose Periodontal Bone Loss and Stage Periodontitis*. Scientific Reports, 2020. **10**(1): 7531.
31. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention, 2015.

- 32. Zhang, Y., et al. *Image super-resolution using very deep residual channel attention networks*. Proceedings of the European Conference on Computer Vision (ECCV), 2018.
- 33. Badrinarayanan, V., A. Kendall, and R. Cipolla, *Segnet: A deep convolutional encoder-decoder architecture for image segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017. **39**(12): p. 2481-2495.
- 34. He, K.M., et al., *Mask R-CNN*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020. **42**(2): p. 386-397.
- 35. Redmon, J., et al. *You only look once: Unified, real-time object detection*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- 36. Redmon, J. and A. Farhadi, *Yolov3: An incremental improvement*. arXiv preprint arXiv:.02767, 2018.
- 37. Bochkovskiy, A., C.-Y. Wang, and H.-Y.M. Liao, *YOLOv4: Optimal Speed and Accuracy of Object Detection*. arXiv preprint arXiv:.10934, 2020.
- 38. Redmon, J., *Darknet: Open source neural networks in c*. 2013.

39. Wang, C.-Y., et al. *CSPNet: A new backbone that can enhance learning capability of cnn*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020.
40. He, K.M., et al., *Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015. **37**(9): p. 1904-1916.
41. Liu, S., et al. *Path aggregation network for instance segmentation*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
42. Lin, T.-Y., et al., *Feature pyramid networks for object detection*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
43. Kim, D.S., et al., *Principal direction of inertia for 3D trajectories from patient-specific TMJ movement*. Computers in Biology and Medicine, 2013. **43**(3): p. 169-175.
44. Yi, W.J., et al., *Comparison of trabecular bone anisotropies based on fractal dimensions and mean intercept length*

- determined by principal axes of inertia*. Medical & Biological Engineering & Computing, 2007. **45**(4): p. 357-364.
45. Yi, W.J., et al., *Direct measurement of trabecular bone anisotropy using directional fractal dimension and principal axes of inertia*. Oral Surgery, Oral Medicine, Oral Pathology, and Oral Radiology, 2007. **104**(1): p. 110-116.
46. Tonetti, M.S., H. Greenwell, and K.S. Kornman, *Staging and grading of periodontitis: Framework and proposal of a new classification and case definition*. Journal of Periodontology, 2018. **89**: p. S159-S172.
47. Polak, M., H. Zhang, and M.H. Pi, *An evaluation metric for image segmentation of multiple objects*. Image and Vision Computing, 2009. **27**(8): p. 1223-1227.
48. Seyedhosseini, M. and T. Tasdizen, *Semantic Image Segmentation with Contextual Hierarchical Models*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016. **38**(5): p. 951-964.

49. Zou, K.H., et al., *Statistical validation of image segmentation quality based on a spatial overlap index*. Academic Radiology, 2004. **11**(2): p. 178-189.
50. Szegedy, C., A. Toshev, and D. Erhan, *Deep neural networks for object detection*. Proceedings of the Advances in Neural Information Processing Systems, 2013.
51. Girshick, R., et al., *Rich feature hierarchies for accurate object detection and semantic segmentation*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014.
52. Hunt, R.J., *Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability*. Journal of Dental Research, 1986. **65**(2): p. 128-130.
53. Muller, R. and P. Buttner, *A critical discussion of intraclass correlation coefficients*. Statistics in Medicine, 1994. **13**(23-24): p. 2465-2476.
54. Altman, D.G. and J.M. Bland, *Measurement in medicine: the analysis of method comparison studies*. Journal of the Royal Statistical Society, 1983. **32**(3): p. 307-317.

55. Passing, H. and W. Bablok, *A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I.* Clinical Chemistry and Laboratory Medicine (CCLM), 1983. **21**(11): p. 709-720.
56. Cheng, J.Z., et al., *Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans.* Scientific Reports, 2016. **6**: 24454.
57. Lo, S.B., et al., *Artificial convolution neural network techniques and applications for lung nodule detection.* IEEE Transactions on Medical Imaging, 1995. **14**(4): p. 711-718.
58. LeCun, Y., et al., *Gradient-based learning applied to document recognition.* Proceedings of the IEEE, 1998. **86**(11): p. 2278-2324.
59. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks.*

Proceedings of the Advances in Neural Information Processing Systems, 2012.

60. Szegedy, C., et al. *Going deeper with convolutions*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
61. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556, 2014.
62. Huang, G., et al. *Densely connected convolutional networks*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
63. He, K., et al. *Deep residual learning for image recognition*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
64. Rajkomar, A., et al., *High-Throughput Classification of Radiographs Using Deep Convolutional Neural Networks*. Journal of Digital Imaging, 2017. **30**(1): p. 95-101.

65. Rajpurkar, P., et al., *Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning*. arXiv preprint arXiv:1711.05225, 2017.
66. Shen, W., et al., *Multi-scale Convolutional Neural Networks for Lung Nodule Classification*. Information Processing in Medical Imaging, 2015. **24**: p. 588-599.
67. Ker, J., et al., *Deep Learning Applications in Medical Image Analysis*. IEEE Access, 2018. **6**: p. 9375-9389.
68. Ciompi, F., et al., *Bag-of-Frequencies: A Descriptor of Pulmonary Nodules in Computed Tomography Images*. IEEE Transactions on Medical Imaging, 2015. **34**(4): p. 962-973.
69. Szegedy, C., et al. *Rethinking the inception architecture for computer vision*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
70. Tao, A., J. Barker, and S. Sarathy, *Detectnet: Deep neural network for object detection in digits*. Parallel Forall, 2016. **4**.

- 71. Hesamian, M.H., et al., *Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges*. Journal of Digital Imaging, 2019. **32**(4): p. 582-596.
- 72. Girshick, R. *Fast r-cnn*. Proceedings of the IEEE International Conference on Computer Vision, 2015.
- 73. Ren, S., et al. *Faster r-cnn: Towards real-time object detection with region proposal networks*. Proceedings of the Advances in Neural Information Processing Systems, 2015.
- 74. Long, J., E. Shelhamer, and T. Darrell. *Fully convolutional networks for semantic segmentation*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
- 75. Milletari, F., N. Navab, and S.-A. Ahmadi. *V-net: Fully convolutional neural networks for volumetric medical image segmentation*. Proceedings of the 4th International Conference on 3D Vision (3DV), 2016.
- 76. Çiçek, Ö., et al. *3D U-Net: learning dense volumetric segmentation from sparse annotation*. Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, 2016.

77. Lee, J.H., et al., *Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm*. Journal of Periodontal and Implant Science, 2018. **48**(2): p. 114-123.
78. Lee, J.S., et al., *Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study*. Dentomaxillofacial Radiology, 2019. **48**(1): 20170344.
79. Albandar, J.M. and D.K. Abbas, *Radiographic Quantification of Alveolar Bone Level Changes - Comparison of 3 Currently Used Methods*. Journal of Clinical Periodontology, 1986. **13**(9): p. 810-813.
80. Bjorn, H. and K. Holmberg, *Radiographic determination of periodontal bone destruction in epidemiological research*. Odontol Revy, 1966. **17**(3): p. 232-250.
81. Lang, N.P. and R.W. Hill, *Radiographs in Periodontics*. Journal of Clinical Periodontology, 1977. **4**(1): p. 16-28.
82. Persson, R.E., et al., *Comparison between panoramic and intra-oral radiographs for the assessment of alveolar bone levels in a*

- periodontal maintenance population*. Journal of Clinical Periodontology, 2003. **30**(9): p. 833-839.
83. Tao, Q., et al., *Deep Learning-based Method for Fully Automatic Quantification of Left Ventricle Function from Cine MR Images: A Multivendor, Multicenter Study*. Radiology, 2019. **290**(1): p. 81-88.
 84. Shorten, C. and T.M. Khoshgoftaar, *A survey on Image Data Augmentation for Deep Learning*. Journal of Big Data, 2019. **6**(1): 60.
 85. Tan, C., et al. *A survey on deep transfer learning*. Proceedings of the International Conference on Artificial Neural Networks, 2018.
 86. Bozinovski, S., *Reminder of the First Paper on Transfer Learning in Neural Networks, 1976*. Informatica, 2020. **44**(3).
 87. Pan, S.J. and Q. Yang, *A survey on transfer learning*. IEEE Transactions on Knowledge and Data Engineering, 2009. **22**(10): p. 1345-1359.

88. Pratt, L.Y., *Discriminability-based transfer between neural networks*. Proceedings of the Advances in Neural Information Processing Systems, 1992.
89. Weiss, K., T.M. Khoshgoftaar, and D. Wang, *A survey of transfer learning*. Journal of Big data, 2016. **3**(1): 9.
90. Goodfellow, I., et al., *Deep learning*. Vol. 1. 2016: MIT press Cambridge.
91. Ozcan, G. and A.E. Sekerci, *Classification of alveolar bone destruction patterns on maxillary molars by using cone-beam computed tomography*. Nigerian Journal of Clinical Practice, 2017. **20**(8): p. 1010-1019.
92. Zhao, Z.-Q., et al., *Object detection with deep learning: A review*. IEEE Transactions on Neural Networks and Learning Systems, 2019. **30**(11): p. 3212-3232.
93. Jiao, L., et al., *A survey of deep learning-based object detection*. IEEE Access, 2019. **7**: p. 128837-128868.
94. Bilic-Zulle, L., *Comparison of methods: Passing and Bablok regression*. Biochemia Medica, 2011. **21**(1): p. 49-52.

95. Giavarina, D., *Understanding bland altman analysis*. Biochemia Medica, 2015. **25**(2): p. 141-151.

국 문 초 록

방사선학적 골 소실량과 치주염

단계의 딥러닝 기반 컴퓨터

보조진단 방법: 다기기 연구

이상정

융합과학부 방사선융합의생명전공

서울대학교 융합과학기술대학원

치주염과 치은염을 포함한 치주질환은 인류가 겪고 있는 가장 흔한 질환 중 하나이다. 구강 및 악안면 부위 치조골의 침하는 치주질환의 주요 증상이며, 이는 골 손실, 치아 손실, 치주염을

유발할 수 있으며, 이를 방치할 경우 저작 기능 장애로 인한 영양실조의 원인이 될 수 있다. 2017 년 미국치주학회(American Academy of Periodontology)와 유럽치주학회(European Federation of Periodontology)는 공동 워크숍을 통해 치주염에 대한 새로운 정의와 단계 분류 및 진단에 관련된 기준을 발표하였다. 최근, 딥러닝을 기반으로 한 컴퓨터 보조진단 기술(Computer-aided Diagnoses, CAD)이 의료방사선영상 분야에서 복잡한 문제를 해결하는 데 광범위하게 사용되고 있다.

선행 연구에서 저자는 파노라마방사선영상에서 치주염을 자동으로 진단하기 위한 딥러닝 하이브리드 프레임워크를 개발하였다. 이는 해부학적 구조물 분할을 위한 딥러닝 신경망 기술과 치주염의 단계 분류를 위한 컴퓨터 보조진단 기술을 융합하여 단일 프레임워크에서 치주염을 자동으로 분류, 진단하는 방법이다. 이를 통해 각 치아에서 방사선적 치조골 소실량을 자동으로 정량화하고, 2017 년 워크숍에서 제안된 기준에 따라 치주염을 3 단계로 분류하였다.

본 연구에서는 선행 개발된 방법을 개선하여 상실 치아와 식립된 임플란트의 수를 검출, 정량화하여 치주염을 4 단계로 분류하는 방법을 개발하였다. 또한 개발된 방법의 일반화 정도를 평가하기 위해 서로 다른 기기를 통해 촬영된 영상을 이용한

다기기 연구를 수행하였다. 3 개의 기기를 이용하여 총 500 매의 파노라마방사선영상을 수집하여 CNN 학습을 위한 데이터셋을 구축하였다. 수집된 영상 데이터셋을 이용하여, 기존 연구에서 의료영상 분할에 일반적으로 사용되는 3 개의 CNN 모델과 Mask R-CNN 을 학습시킨 후, 해부학적 구조물 분할 정확도 비교 평가를 실시하였다. 또한 CNN 의 높은 학습 효율성 확보와 및 다기기 영상에 대한 추가 학습을 위해 선행 연구에서 도출된 사전 훈련 가중치(pre-trained weight)를 이용한 CNN 의 전이학습을 실시하였다.

CNNv4-tiny 를 이용하여 상설 치아를 검출한 결과, 0.88, 0.85, 0.87, 0.86, 0.85 의 precision, recall, F1-score, mAP 정확도를 보였다. 해부학적 구조물 분할 결과, Mask R-CNN 을 기반으로 수정된 CNN 은 치조골 수준에 대해 0.96, 백악법랑경계 수준에 대해 0.92, 치아에 대해 0.94 의 분할정확도(DSC)를 보였다.

이어 개발된 방법을 이용하여 학습에 사용되지 않은 30 매(기기 별 10 매)에서 자동으로 결정된 치주염의 단계와 서로 다른 임상경험을 가진 3 명의 영상치의학 전문의가 진단한 단계 간 비교 평가를 수행하였다. 평가 결과, 모든 치아에 대해 자동으로 결정된 치주염 단계와 전문의들이 진단한 단계 간 0.31 의 오차(MAD)를 보였다. 또한 기기 1, 2, 3 의 영상에 대해 각각 0.25,

0.34, 0.35 의 오차를 보였다. 개발된 방법을 이용한 결과와 방사선 전문의의 진단 사이의 PCC 값은 기기 1, 2, 3 의 영상에 대해 각각 0.73, 0.77, 0.75 로 계산되었다 ($p < 0.01$). 전체 영상에 대한 최종 ICC 값은 0.76 ($p < 0.01$)로 계산되었다. 또한 개발된 방법과 방사선 전문의의 진단 사이의 ICC 값은 기기 1, 2, 3 의 영상에 대해 각각 0.91, 0.94, 0.93 으로 계산되었다 ($p < 0.01$). 마지막으로 최종 ICC 값은 0.93 으로 계산되었다 ($p < 0.01$).

Passing 및 Bablok 분석의 경우 회귀직선의 기울기와 x 축 절편은 교수, 임상강사, 전공의에 대해 각각 1.176 ($p > 0.05$), 1.100 ($p > 0.05$), 1.111 ($p > 0.05$)와 -0.304 , -0.199 , -0.371 로 나타났다. Bland 와 Altman 분석의 경우 자동으로 결정된 영상 별 평균 단계와 영상치의학 전공 치과 의사의 진단 결과 간 교수, 임상강사, 전공의에 대해 0.007 (95 % 신뢰 구간 (CI), $-0.060 \sim 0.074$), 각각 -0.022 (95 % CI, $-0.098 \sim 0.053$), -0.198 (95 % CI, $-0.291 \sim -0.104$)로 계산되었다.

결론적으로, 본 논문에서 개발된 딥러닝 하이브리드 프레임워크는 딥러닝 신경망 기술과 컴퓨터 보조 진단 기술을 융합하여 환자의 파노라마 방사선 영상에서 치주염을 4 단계로 분류하였다. 본 방법은 높은 해부학적 구조물 및 상실 치아 검출 정확도를 보였으며, 자동으로 결정된 치주염 단계는 임상 의의 진단

결과와 높은 일치율과 상관성을 보여주었다. 또한 다기기 연구를
통해 개발된 방법의 높은 정확성과 일반화 정도를 검증하였다.

주요어 : 딥러닝, 컴퓨터 보조진단, 파노라마방사선영상, 방사선학적
골 소실량, 치주염, 다기관 연구

학 번 : 2014-24908