Deep Learning for Early Dental Caries Detection in Bitewing Radiographs

Shinae Lee
Gangnam Severance Hospital
Sang-il Oh
SELVAS AI Inc.
Junik Jo
SELVAS AI Inc.
Sumi Kang
Yonsei University
Yooseok Shin
Yonsei University
Jeong-won Park (pjw@yuhs.ac)
Gangnam Severance Hospital

Research Article

Keywords: Deep learning, Dental caries, Bitewing radiographs, Diagnosis, Caries detection, Convolutional neural networks, Artificial intelligence

Posted Date: April 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-395603/v1

Version of Record: A version of this preprint was published at Scientific Reports on August 19th, 2021. See the published version at https://doi.org/10.1038/s41598-021-96368-7.

Deep Learning for Early Dental Caries Detection in Bitewing Radiographs

Shinae Lee[1], Sang-il Oh[2], Junik Jo[2], Sumi Kang[3], Yooseok Shin[3], Jeong-won Park[1, *]

[1] Department of Conservative Dentistry, Gangnam Severance Hospital, College of Dentistry, Yonsei University, Seoul, Korea

[2] SELVAS AI Inc., Seoul 08594, Korea

[3] Department of Conservative Dentistry, College of Dentistry, Yonsei University, Seoul, Korea

[*] Corresponding author at: Department of Conservative Dentistry, Gangnam Severance Hospital, 146-92 Dogok-dong, Gangnam-gu, Seoul 135-720, Korea

E-mail address: pjw@yuhs.ac

Abstract

The early detection of incipient dental caries enables preventive treatment, and bitewing radiography is a good diagnostic tool for posterior incipient caries. In the field of medical imaging, the utilization of deep learning with convolutional neural networks (CNNs) to process various types of images has been actively researched and has shown promising performance. In this study, we developed a CNN model using a U-shaped deep CNN (U-Net) for dental caries detection on bitewing radiographs and investigated whether this model can improve clinicians' performance. In total, 304 bitewing radiographs were used to train the deep learning model and 50 radiographs were used for performance evaluation. The diagnostic performance of the CNN model on the total test dataset was as follows: precision, 63.29%; recall, 65.02%; and F1-score, 64.14%, showing quite accurate performance. When three dentists detected dental caries using the results of the CNN model as reference data, the overall diagnostic performance of all three clinicians significantly improved, as shown by an increased recall ratio (D1, 85.34%; D1', 92.15%; D2, 85.86%; D2', 93.72%; D3, 69.11%; D3', 79.06%). These increases were especially significant in the incipient and moderate caries subgroups. The deep learning model may help clinicians to diagnose dental caries more accurately.

Keywords:

Deep learning, Dental caries, Bitewing radiographs, Diagnosis, Caries detection, Convolutional neural networks, Artificial intelligence

Introduction

The early detection of incipient dental caries can prevent invasive treatment, thereby saving healthcare costs. However, detecting posterior incipient proximal caries with clinical examinations alone is difficult, and bitewing radiography is helpful as the gold standard for diagnosing demineralized proximal caries [1]. The combination of bitewing radiographs and a visual inspection is a routine diagnostic approach for proximal caries detection [2]. Besides radiographs, fiber optic transillumination and fluorescence-based methods, such as DIAGNOdent (KaVo, Charlotte, NC, USA), are other ways to detect dental caries [3]. However, these methods have limitations in detecting posterior incipient proximal caries [4] and incur additional device costs. Bitewing radiograph is still the most reliable and widely used method in clinical situation.

Although radiography is recommended as a diagnostic method, the detection of dental caries using radiographs can be subjective. Major differences exist across observers in terms of whether caries lesions are detected, even using the same radiograph. Factors such as the quality of the radiograph, viewing conditions, the dentist's expectations, variability across examiners (in particular, whether a dentist leans towards or minimizes caries diagnoses), and the length of time per examination cause discrepancies in interrater agreement [5,6]. In a previous study, 34 raters showed considerable variation when examining the same bitewing radiographs, with mean kappa values of 0.30–0.72 for the presence or absence of dental caries and the degree thereof [5,7]. A lack of consistency is a significant problem, especially for the detection of incipient dental caries [8].

In recent years, researchers have actively explored the utilization of deep learning with convolutional neural networks (CNNs) to process various types of medical images, with promising performance. The usage of deep learning for the diagnosis of diseases is increasing, and deep learning has shown precise and expeditious detection with improved clinical outcomes [9]. In dentistry, the use of deep convolutional networks has been investigated since 2015. The U-Net was employed by Ronneberger to analyze dental structure segmentation on bitewing radiographs [10]. Subsequently, multiple deep learning models for diagnosing dental caries or lesion detection on dental X-ray images have been studied [11-13]. Most extant research has been limited to analyses of the detection performance of deep learning models, and some recent papers have compared diagnostic performance between deep learning models and clinicians [8,14]. However, no study has yet investigated the changes that result from using deep learning models in clinical situations, or how clinicians can benefit from deep learning models.

In this study, we developed a U-Net CNN model [10] for dental caries detection on bitewing radiographs through an analysis of dental structure and differences in radiographic density on radiographs without special manipulation, and investigated whether the proposed model can help clinicians diagnose dental caries in actual clinical settings.

Materials and Methods

Data Collection

This study was approved by the Institutional Review Board of Yonsei University Gangnam Severance Hospital and Yonsei University Dental Hospital (IRB No. 3-2019-0062 & No. 2-2019-0031) and all research was carried out in accordance with relevant guidelines and regulations. This study is retrospective study which the data and methods we used are waived from the need of informed consent. The radiographs were randomly selected by two dentists from the archive of the Department of Conservative Dentistry, Yonsei University containing bitewing radiographs taken from January 2017 to December 2018 for caries diagnosis and treatment. Radiographs with low image quality, excessive distortion, or severe overlapping of proximal surfaces due to the anatomical arrangement of particular teeth were excluded, because those features would interfere with a precise caries diagnosis. The bitewing radiographs were taken with the aid of a film-holding device (RINN SCP-ORA; DENTSPLY Rinn, York, PA, USA) using a dental X-ray machine (Kodak RVG 6200 Digital Radiography System with CS 2200; Carestream, Rochester, NY, USA).

The collected data were transferred to a tablet (Samsung Galaxy Note 10.1; Samsung Electronics Co., Suwon, South Korea) as Digital Imaging and Communications in Medicine files, and the two well-trained observers (postgraduate students of the Department of Conservative Dentistry, with a minimum clinical experience of 5 years) examined the dental images sequentially. The observers were allowed to adjust the density or contrast of radiographs as they wished with no time limitation. The observers drew lines for the segmentation of dental structures (caries, enamel, dentin, pulp, metal restorations, tooth-colored restorations, gutta percha) (Fig. 1) on the bitewing radiographs. All types of dental caries (e.g., proximal, occlusal, root and secondary caries) that can be observed on bitewing radiographs were tagged regardless of the severity. Discrepancies in caries tagging were initially resolved by consensus between the two observers, and if the disagreement persisted, it was resolved by another author.

Training the Convolutional Neural Network

The bitewing radiographs were directly used as diagnostic data for CNN without specific pre-processing (e.g. image enhancement and manual setting of the region of interest). The models were trained using each bitewing radiograph with 12-bit depth and its paired binary mask, while only the radiograph itself was fed into the models to detect the target regions for testing. The radiographs were scaled to the size of 572×572 to be used as input for the networks. In total, 304 bitewing radiographs, which were randomly divided into two groups (149 radiographs for $\mathcal{D}_{\mathrm{A}}$, which was used for training on both structure and caries segmentation, and 105 radiographs for $\mathcal{D}_B$, which was used for training on caries segmentation alone, as described below in greater depth) and a group of 50 radiographs with no dental caries ($\mathcal{D}_{\mathrm{C}}$), were used to train the deep learning model, while 50 radiographs ($\mathcal{D}_{\mathrm{D}}$) were used for performance evaluation. To evaluate the generality of the trained models without dataset selection biases, cross-validation was applied to the datasets $\mathcal{D}_{\mathrm{A}}$, $\mathcal{D}_B$, and $\mathcal{D}_{\mathrm{C}}$. Two augmentation processes were applied to the training set to train the model. Image augmentation, including intensity variation, random flipping, rotation, elastic transformation, width scaling, and zooming, was applied first, followed by mask augmentation, consisting of random kernel dilations, and elastic transformation.

To predict the caries and structure regions from each input radiograph, this study used the U-Net architecture to segment the target regions on the pixel level. The U-Net architecture includes a convolutional part and an up-convolutional part. The convolutional part has the typical structure of convolutional neural networks with five convolutional layers. Contrastively, the up-convolutional part performs an up-sampling of the feature map by taking a concatenated output of the previous layer and the opposite convolutional layer as an input. Then, a 1×1

convolution maps the dimensionality of the feature maps to the desired number of classes. Finally, a softmax layer outputs a probability map where each pixel indicates the probability for each class within a range of $[0, 1]$ [ref. Softmax]. The network was optimized by the adaptive moment estimation optimizer with an initial learning rate of 0.00001 [ref. Adam optimizer]. Details on model architecture are provided in the Supplementary Fig. S1 online.

By simultaneously feeding each input radiograph into the U-SS and the U-CS, a caries probability map and a structure probability map were generated. The program was constructed to visualize detected dental caries as an area on the bitewing radiograph and also to show the degree of dental caries numerically. For the caries probability map, a pixel value of more than 0.55 was classified as a caries pixel. Early in the process of training the model, the threshold value, which shows the highest agreement between the model results and the results diagnosed by experts, was set as 0.55. To reduce the likelihood of false detection, areas of caries detected by the U-CS without overlap with enamel or dentin regions were eliminated. Figure 2 shows an overview of caries detection and false detection refinement.

Performance Evaluation and Comparisons (Inter-dentists, Dentists vs. Model)

To evaluate the performance of dental caries detection, we measured the precision (also known as positive predictive value, %), recall (also known as sensitivity, %), and F1-score (F1 score $= (2(precision * recall)) / (precision + recall)$) according to the overlap ratio (θ), which was set to 0.1, between the agreed-upon and predicted carious regions. We computed the final result by summing the results of cross-validation models on each dataset $\mathcal{D}_{\mathrm{A}}$, $\mathcal{D}_{\mathrm{B}}$, and $\mathcal{D}_{\mathrm{C}}$, whereas $\mathcal{D}_{\mathrm{D}}$ was not included.

To assess whether the developed deep learning model can support the diagnosis of dentists, we used $\mathcal{D}_{\mathrm{D}}$, the evaluation dataset containing 50 radiographs. Specifically, we explored how dentists' diagnoses changed before and after they referred to the predictions of dental caries made by the deep learning model. To do this, the deep learning model, which was previously trained with 304 training images, was used to detect dental caries in 50 radiographs, and three dentists (working in clinics with a clinical experience of 4-6 years) were instructed to tag dental caries independently, without any consultations. After a 3-week interval, the three dentists diagnosed the same radiographs, but with two distinguishable guidance lines, comprising the regions of dental caries predicted by the model and their previous diagnosis. The three dentists were instructed to revise (modify, delete, or add) their previously tagged dental caries by referencing the dental caries results detected by the deep learning model. As a result, each radiograph had seven sets of dental caries regions in total: six from dentists, and one from the model. To evaluate the validity of the model as a diagnostic support system, we analyzed the changes between first and second diagnoses of each dentist by measuring the F1-score. We also compared the concordance of the dental caries diagnosed by the three dentists between the first and second diagnoses.

Comparisons according to the Severity of the Dental Caries

Three dentists reached consensus on dental caries detection in $\mathcal{D}_{\mathrm{D}}$, the evaluation dataset containing 50 radiographs, and divided the agreed-upon dental caries into 3 subgroups according to the severity of caries (incipient, moderate, and advanced). Incipient caries were defined as lesions present in the outer enamel or at the dentinoenamel junction; moderate caries as those in the outer half of the dentin, and advanced caries as those in the inner half of the dentin. If the same area was included, the same dental caries was considered to have been detected regardless of the overlap ratio on the caries lesion level; using this criterion, the precision (%), recall (%), and F1-score were recalculated according to the severity of the agreed-upon dental caries.

Statistical analysis

The diagnostic performance for readers and the U-Net CNN model was calculated in terms of recall (%), precision (%), and the F1-score. To compare recall and precision between readers and the U-Net CNN model, generalized estimating equations (GEEs) were used, while the F1-score was compared using the bootstrapping method (resampling: 1000). All statistical analyses were performed using SAS (version 9.4, SAS Inc., Cary, NC, USA) and R version 3.4.3. The significance level was set at alpha = 0.05.

Results

The diagnostic performance of the final CNN model on the total test dataset ($\mathcal{D}_{A,B,C}$) was as follows: precision, 63.29%; recall, 65.02%; and F1-score, 64.14%.

The deep learning program was quite accurate and showed a stable pattern of dental caries detection performance. All types of dental caries (root caries, secondary dental caries, and gaps under restoration) recognizable on bitewing radiographs, other than proximal dental caries, were detectable. However, the false detection rate of dental caries was somewhat higher when the quality of the radiographs was low, dental overlap was severe, and when the bitewing images included the third molar.

The recall, precision, and F1-score according to an overlap ratio (θ) of 0.1 between dentists and the model are shown in Table 1. When the three dentists detected dental caries with support from the deep learning model, their recall, precision, and F1-score increased. When the three observers independently diagnosed dental caries on 50 images, the number of dental caries tagged by each was as follows: D1, 177; D2, 182; and D3, 155. The consistency of caries tagging between observers D1, D2, and D3 was found to be D1-D2, 85.79%; D1-D3, 79.06%; and D2-D3, 79.41% for the F1-score (Table 2). After revision of their diagnoses referencing the dental caries detection results of the deep learning model, the final number of dental caries tagged by the three observers increased as follows: D1, 211; D2, 202; D3, 175. When the clinicians detected dental caries with reference to the results of the CNN model, they detected more caries. The final dental caries tagging consistency of the three observers was found to be D1'-D2', 86.19%; D1'-D3', 81.98%; and D2'-D3', 80.61% for the F1-score. Thus, the discrepancies in interrater agreement decreased. The concordance between the first and second diagnoses of each dentist was as follows: D1-D1', 90.20%; D2-D2', 94.79%; and D3-D3', 92.63%.

When the results were analyzed according to the severity of the dental caries, the clinicians tended to be less accurate in detecting incipient caries lesions (Fig. 3). When the three clinicians referred to the deep learning results as a second opinion, the overall diagnostic performance of all three clinicians significantly improved. These increases were especially significant for incipient and moderate caries, which are easy to miss.

Discussion

The diagnosis and treatment of dental caries, which comprise a basic aspect of dentistry, based on clinical and radiological data depend on clinicians' subjective judgment. Therefore, the existence of significant differences in the standardization and accuracy of dental caries diagnoses even when clinicians observe the same radiographs is a major problem. Likewise, in this study, when each of the three well-trained dentists diagnosed dental caries independently using 50 evaluation images, the dental caries detection rate showed discrepancies. The tagging consistency of the three observers was found to be D1-D2, 85.79%; D1-D3, 79.06%; and D2-D3, 79.41% using the F1 score, indicating that inter-examiner reproducibility showed some deviations. These discrepancies underscore the inherently subjective nature of dental caries diagnosis using radiographs. In this study, to evaluate the usefulness of the deep learning model as a second opinion for the diagnosis of dental caries, the three observers were instructed to revise their initial results regarding the detection of dental caries by referencing the dental caries detection results of the deep learning model. After revision, the final caries tagging consistency of the three observers increased (D1'-D2', 86.19%; D1'-D3', 81.98%; and D2'-D3', 80.61% for the F1-score), showing that the variation between clinicians decreased. The deep learning model can help in the standardization of the inter-examiner detection of dental caries on bitewing radiographs.

In the present study, when three clinicians performed dental caries tagging on the evaluation dataset, the lowest recall ratio was found for incipient caries, followed in ascending order by moderate and advanced caries. This result coincides with the study of Cantu-Garcia (2020) [8]. Clinicians were more likely to miss incipient caries on bitewing radiography and showed lower accuracy than for advanced caries. However, when clinicians referred to the results of the deep learning model when making diagnoses, the recall ratio increased in every group of caries severity. This change was particularly remarkable in the incipient and moderate groups. Therefore, guidance from the deep learning model was meaningful as a way to help clinicians detect caries that could otherwise be mistakenly missed.

Furthermore, after revision, the number of caries tagged by all three observers increased. When the model results were referenced, it was more common for the observer to detect additional dental caries or to increase the extent of caries than it was for the observer to remove the caries or reduce the affected area. Thus, referencing deep learning results as a second opinion was helpful for finding dental caries that had not been discovered initially due to time restraints or mistakes. Since the U-Net CNN model selects candidate carious areas that can be missed by mistake in a busy clinical situation, it can prompt clinicians to look at the relevant areas once more, thereby reducing the likelihood of missing the appropriate timing for treatment as a result of not catching dental caries early.

Thus, it may be helpful for clinicians to reference the dental caries detection results of a deep learning model as a second opinion when using bitewing radiography for caries detection, as a way to improve diagnostic accuracy and to reduce diagnostic variation between observers.

The deep learning model developed in this study showed reliable dental caries detection performance as a result of training with 304 images, which included 763 dental caries. However, a limitation is that the training data set was small. Therefore, two data augmentation processes were performed to compensate for this limitation. Furthermore, only results that equaled or exceeded the threshold value (size of dental caries, prediction probability: 0.55) were considered to be detected dental caries, which was appropriate for reducing false detection results and improving detection performance. Additionally, the model was trained using 50 images with no dental caries (dataset D_C), and a penalty for false detection was applied to the CNN model to reduce false positives. Subsequently, improvements in dental caries detection performance were confirmed. Furthermore, to avoid false positives, caries was detected from only the enamel and dentin areas by combining the results of caries detection (U-CS) and tooth segmentation (U-SS).

In future research, the ongoing addition of training data will be needed. Additionally, to obtain meaningfully high levels of accuracy in a clinical setting, it is necessary to check whether the same dental caries detection performance is shown when using bitewing radiography obtained by devices at multiple institutions, instead of only using bitewing radiography obtained at a single institution. This study used bitewing radiographs obtained at two hospitals (Gangnam Severance and Yonsei University Dental Hospital). The X-ray equipment used at both institutions was identical, and both sets of data showed a similar level of performance in dental caries detection.

In this study, the presence of dental caries was evaluated only using radiographs, without visual or clinical examination data. Therefore, this study used the data of agreed-upon dental caries based on the consensus of the observers and has the limitation of lacking gold-standard findings such as histological examination results that would conclusively demonstrate the actual presence of dental caries. It should be kept in mind that bitewing radiography, as used in this study, is vulnerable to false-positive and false-negative diagnoses [15], and therefore may not be sufficient for diagnosing dental caries in images where proximal surfaces overlap or distortion is severe. When an impacted third molar overlaps with the second molar, the CNN model we trained tended to show false-positive responses because of the lack of training data for cases of superposition with other unexpected features; this limitation should be overcome by increasing the size of the training dataset. However, since such areas can be excluded as false-positive errors by clinicians, they are thought not to pose a major problem in clinical settings.

Clinicians should not wholly rely on artificial intelligence-based dental caries detection results, but should instead use them only for reference. Through additional clinical examinations and an assessment of patients' systemic state, overall oral condition, and overall caries risk, clinicians should determine the final diagnosis and treatment plan on their own initiative.

Conclusion

The findings of the present study revealed that referencing the dental caries detection results of a deep learning model as a second opinion may help clinicians to diagnose dental caries more accurately. However, the addition of more training data is needed to achieve more stable and precise results.

Author Contributions

S.L. contributed to data acquisition, analysis, and interpretation, drafted and critically revised the manuscript; S.O. contributed to conception, design, data analysis, and interpretation, drafted the manuscript; J.J. contributed to conception, design, data analysis, and interpretation; S.K., and Y.S. contributed to data acquisition; J.P. contributed to conception, design, data analysis, and interpretation, critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of the work.

Acknowledgments

The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.L.

Data availability

All data are included in this published article.

References

1. Hilton, T. J., Ferracane, J. L. & Broome, J. C. *Summitt's fundamentals of operative dentistry: a contemporary approach* (Quintessence Publishing Company Incorporated, 2013).
2. Pitts, N. J. D. R. The use of bitewing radiographs in the management of dental caries: scientific and practical considerations. **25**, 5-16 (1996).
3. Braga, M. M., Mendes, F. M. & Ekstrand, K. R. J. D. C. Detection activity assessment and diagnosis of dental caries lesions. **54**, 479-493 (2010).
4. Pretty, I. A. & Ellwood, R. P. The caries continuum: opportunities to detect, treat and monitor the re-mineralization of early caries lesions. *J Dent* **41 Suppl 2**, S12-21 (2013).
5. Langlais, R. P., Skoczylas, L. J., Prihoda, T. J., Langland, O. E. & Schiff, T. J. O. S., Oral Medicine, Oral Pathology Interpretation of bitewing radiographs: application of the kappa statistic to determine rater agreements. **64**, 751-756 (1987).
6. Bailit, H. L., Reisine, S. T., Damuth, R. L. & Richards, N. P. The validity of the radiographic method in the pretreatment review of dental claims. *J Public Health Dent* **40**, 26-38 (1980).
7. Bader, J. D. & Shugars, D. A. J. J. o. p. h. d. Variation in dentists' clinical decisions. **55**, 181-188 (1995).
8. Cantu, A. G. *et al.* Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent*, 103425 (2020).
9. Kermany, D. S. *et al.* Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **172**, 1122-1131.e1129 (2018).
10. Ronneberger, O., Fischer, P. & Brox, T. Dental X-ray image segmentation using a U-shaped Deep Convolutional network in *Medical Image Computing and Coputer-Assisted Intervention (MICCAI)*, Vol. 9351 234-243 (ISBI, 2015).

11. Lee, J. H., Kim, D. H., Jeong, S. N. & Choi, S. H. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Periodontal Implant Sci* **48**, 114-123 (2018).
12. Lee, J. H., Kim, D. H., Jeong, S. N. & Choi, S. H. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* **77**, 106-111 (2018).
13. Srivastava, M. M., Kumar, P., Pradhan, L. & Varadarajan, S. Detection of tooth caries in bitewing radiographs using deep learning. *arXiv preprint arXiv:1711.07312* (2017).
14. You, W., Hao, A., Li, S., Wang, Y. & Xia, B. Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments. *BMC Oral Health* **20**, 141 (2020).
15. Douglass, C. W. *et al.* Clinical efficacy of dental radiography in the detection of dental caries and periodontal diseases. *Oral Surg Oral Med Oral Pathol* **62**, 330-339 (1986).

Figure 1. Example of the analysis of dental structures and caries tagging**.** The observers drew lines for the segmentation of dental structures (enamel, dentin, pulp, metal restoration, tooth-color restorations, gutta percha) and dental caries on the bitewing radiographs.

Figure 2. Flowchart of the detection of dental caries in the deep learning model, showing two models: the U-Net for caries segmentation (U-CS), and the U-Net for structure segmentation (U-SS).

Figure 3. Comparison of diagnostic performance (recall) between three dentists (before and after revision) and the deep learning model according to the severity of the agreed-upon dental caries. *The significance level was set at alpha = 0.05 in the *post hoc* analysis.

Table 1. Precision, recall, and F1-score according to overlap ratios between dentists and the CNN model. The overlap ratio (θ) was set to 0.1.

TEST DATASET	PRECISION (%)	RECALL (%)	F1-SCORE (%)
D_{D1}	66.67	82.49	73.74

D_{D2}	68.35	81.87	74.50
D_{D3}	57.71	84.52	68.59
D'_{D1}	82.73	86.26	84.45
D'_{D2}	76.15	82.18	79.05
D'_{D3}	67.56	86.86	76.00

D_{DX}, the dataset of the first diagnoses by three dentists (D1, 2, 3) without model assistance; D'_{DX}, the dataset of revised diagnoses by the three dentists with model assistance.

Table 2. F1-score according to overlap ratios between dentists and the concordance of detection by three dentists of the first and second diagnoses. The overlap ratio (θ) was set to 0.1.

PAIRS	F1-SCORE (%)
D_{D1} - D'_{D1}	**90.20**
D_{D2} - D'_{D2}	**94.79**
D_{D3} - D'_{D3}	**57.71**
D_{D1} - D_{D2}	**85.79**
D_{D2} - D_{D3}	**79.06**
D_{D1} - D_{D3}	**79.41**
D'_{D1} - D'_{D2}	**86.19**
D'_{D2} - D'_{D3}	**81.98**
D'_{D1} - D'_{D3}	**80.61**

D_{DX}, the dataset of the first diagnoses by three dentists (D1, 2, 3) without model assistance; D'_{DX}, the dataset of revised diagnoses by the three dentists with model assistance.

Figures

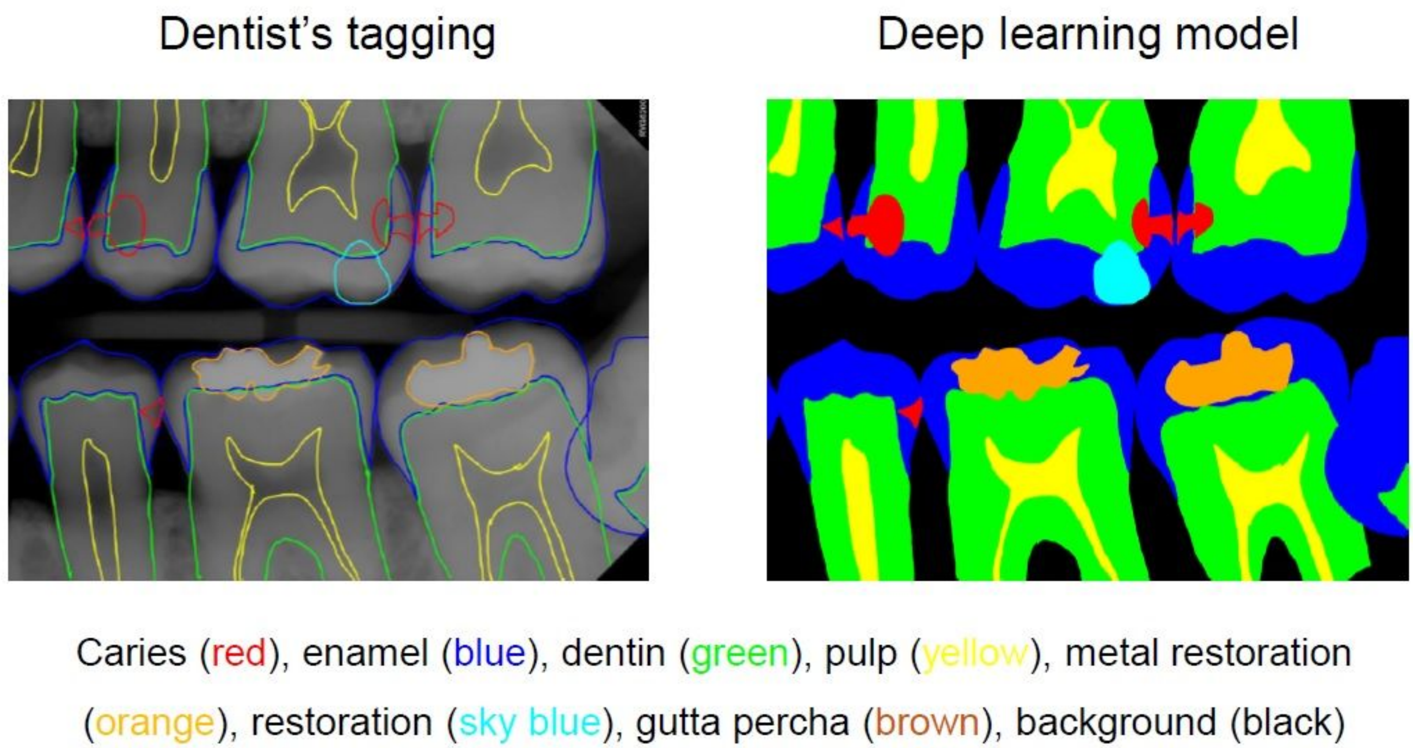


Figure 1

Example of the analysis of dental structures and caries tagging. The observers drew lines for the segmentation of dental structures (enamel, dentin, pulp, metal restoration, tooth-color restorations, gutta percha) and dental caries on the bitewing radiographs.

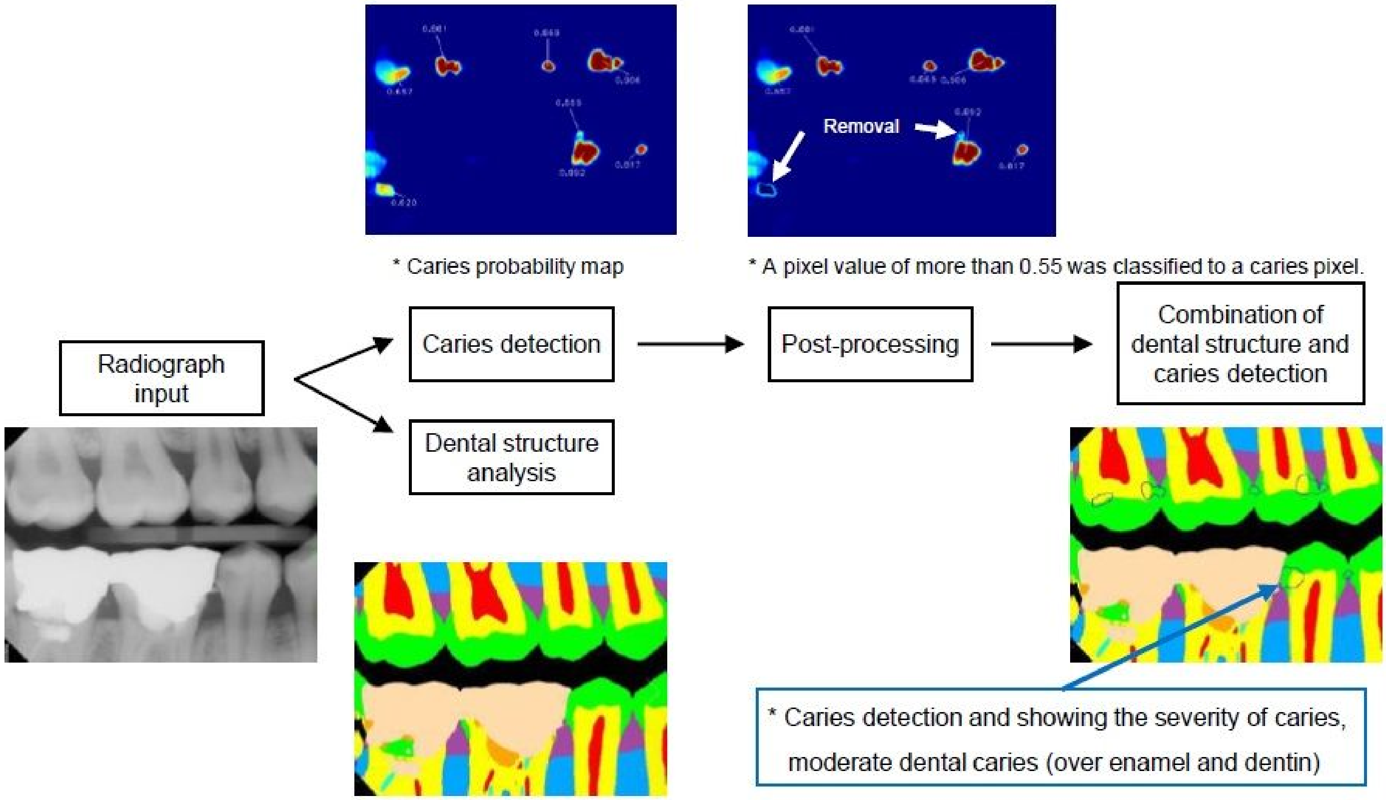
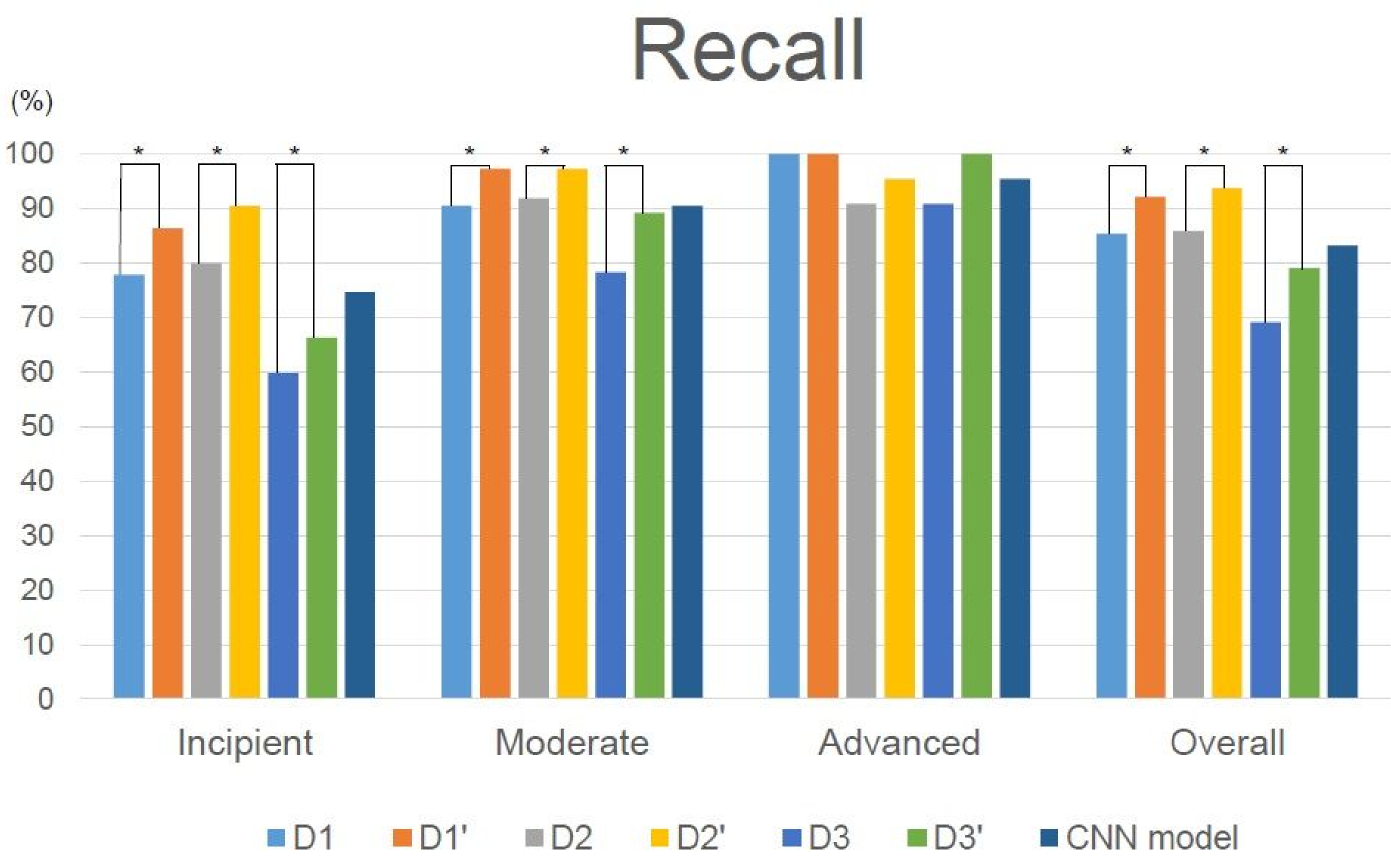


Figure 2

Flowchart of the detection of dental caries in the deep learning model, showing two models: the U-Net for caries segmentation (U-CS), and the U-Net for structure segmentation (U-SS).



DX, before revision; DX', after revision.
Abbreviation: CNN, convolutional neural network.

Figure 3

Comparison of diagnostic performance (recall) between three dentists (before and after revision) and the deep learning model according to the severity of the agreed-upon dental caries. *The significance level was set at alpha = 0.05 in the post hoc analysis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryfigureS1.docx