Automated permanent tooth detection and numbering on panoramic radiograph using a deep learning approach

Ramadhan Hardani Putra, DDS, MDSc, PhD,[a] Eha Renwi Astuti, Dr, DDS, MDSc,[a]
Dina Karimah Putri, DDS, MSi,[a,b] Monica Widiasri, MKom,[c,d] Putri Alfa Meirani Laksanti, BDS,[e]
Hilda Majidah, BDS,[e] and Nobuhiro Yoda, DDS, PhD[f]

Objective. This study aimed to assess the performance of the deep learning (DL) model for automated tooth numbering in panoramic radiographs.
Study Design. The dataset of 500 panoramic images was selected according to the inclusion criteria and divided into training and testing data with a ratio of 80%:20%. Annotation on the data set was categorized into 32 classes based on the dental nomenclature of the universal numbering system using the LabelImg software. The training and testing process was carried out using You Only Look Once (YOLO) v4, a deep convolution neural network model for multiobject detection. The performance of YOLO v4 was evaluated using a confusion matrix. Furthermore, the detection time of YOLO v4 was compared with a certified radiologist using the Mann-Whitney test.
Results. The accuracy, precision, recall, and F1 scores of YOLO v4 for tooth detection and numbering in the panoramic radiograph were 88.5%, 87.70%, 100%, and 93.44%, respectively. The mean numbering time using YOLO v4 was 20.58 ± 0.29 ms, significantly faster than humans ($P < .0001$).
Conclusions. The DL approach using the YOLO v4 model can be used to assist dentists in daily practice by performing accurate and fast automated tooth detection and numbering on panoramic radiographs. (Oral Surg Oral Med Oral Pathol Oral Radiol 2024;137:537−544)

Panoramic radiography is one of the radiographic techniques in dentistry that produces tomographic images by showing facial structures, including the maxillary and mandibular arches and their supporting structures, in 1 image.[1] Panoramic radiography is used to describe morphology and detect pathology, determine tooth growth and development, detect pathologic abnormalities in the dentomaxillofacial area, and estimate a person's age.[2] Panoramic radiographs are clinically useful to detect jaw fractures, location and position of third molars, dental or bone disease, root remnants, impacted teeth, temporomandibular joint dislocation, and dental anomalies and surrounding tissues. In addition, this examination is often used as an initial evaluation that can provide an overview to determine the need for other projections.[1]

The interpretation of the tooth and its surrounding anatomic structures on panoramic radiographs is an important step in detecting pathologic abnormalities.[3] The first thing in interpreting panoramic radiographs is determining the tooth type and number (nomenclature) based on its anatomy and location. For the oral and maxillofacial radiologist who writes the radiologic report of numerous panoramic radiographs in daily practice, manual tooth numbering on many panoramic radiographs is time-consuming and prone to errors because of the excessive workload.[4] Problems can arise with the error in manual handwriting or mistyping the tooth numbering in the report in regard to right/left and upper/lower descriptions. Errors in tooth numbering on panoramic radiographs might be critical because other dentists will rely on their reports. Further errors can arise because the interpretation is highly dependent on the expertise and skills of a dentist; thus, occasionally, it tends to be subjective.[5]

Statement of Clinical Relevance

The artificial intelligence−based system using a deep learning approach can provide fast and accurate automated tooth numbering on panoramic radiographs, which can be used as a foundation to develop a computer-assisted dental charting system to improve the overall health care system.

[a]Department of Dentomaxillofacial Radiology, Faculty of Dental Medicine, Universitas Airlangga, Surabaya, Indonesia.
[b]Division of Dental Informatics and Radiology, Tohoku University Graduate School of Dentistry, Sendai, Japan.
[c]Department of Informatics, Faculty of Engineering, Universitas Surabaya, Surabaya, Indonesia.
[d]Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.
[e]Undergraduate Study Program, Faculty of Dental Medicine, Universitas Airlangga, Surabaya, Indonesia.
[f]Division of Advanced Prosthetic Dentistry, Tohoku University, Graduate School of Dentistry, Sendai, Japan.
Corresponding author: Eha Renwi Astuti E-mail address: eha-r-a@fkg.unair.ac.id
Received for publication Oct 24, 2022; returned for revision Feb 14, 2023; accepted for publication Jun 1, 2023.

2212-4403/$-see front matter
https://doi.org/10.1016/j.oooo.2023.06.003

Currently, various technologies have been developed to help human work, including artificial intelligence (AI). Artificial intelligence is a machine's ability to imitate human intelligence and behavior in carrying out certain tasks. In recent years, AI has experienced rapid development and has become one of the most influential innovations in the world. Many AI applications have helped people's daily lives, such as online search engines, object detection and classification, language recognition, and virtual assistants.[6] The development and application of AI are also emerging in dentistry. Artificial intelligence–based methods can be used to assist dentists in interpreting radiographic results. This method can enable faster data identification and classification and reduce errors from dentist fatigue in interpreting radiographs.[4] Additionally, further development of the AI model could provide automated dental chart filling and treatment planning to reduce the burden of clinical work in daily practice.

Deep learning (DL) is a part of artificial intelligence that can process large amounts of data, such as text, sound, and images.[7] One of the basic types of architecture that runs well in DL is the convolutional neural network (CNN),[8] which has proven capability in image detection and assessing boundary and color characteristics and is also often used to analyze biomedical image results because of its effective ability in image processing.[9] In dentistry, CNN has been proven to detect cephalometric landmarks,[10] segmentation of tooth structure,[11] classifications,[12] detections, and tooth numbering in the dentomaxillofacial radiology field.[13]

Previous studies have tested detection and tooth numbering on radiographs using CNN-based architecture using various methods. Oktay uses the AlexNet architecture with a modified version to detect teeth on panoramic radiographs, achieving an accuracy of >90%.[14] Chen et al. also tested the detection and numbering of teeth using the region-with CNN (R-CNN) on the dataset of 1250 periapical radiographs and obtained a precision value of 91.7%.[15] Tuzoff et al. also conducted a similar study using CNN's VGG-16 with 1.574 panoramic radiographs; the precision value reached 99.45%.[13] Leite et al., detecting and segmenting teeth using a combination of 2 deep-CNNs with 3.576 panoramic radiographs, obtained a precision of 96.9%[4]. Further development on automated tooth numbering on panoramic radiographs is needed before its implementation in clinical practice.

You Only Look Once (YOLO) is a 1-stage object detection algorithm detecting objects in real-time using CNN.[16] Four versions of YOLO have been developed. Bochkovskiy et al. have proven that YOLO v4 is an advanced detector that is faster and more accurate than any other available.[17] In dental radiology, YOLO v4 has been investigated for use in detecting several things, such as periodontal bone loss,[18] location of third molars,[19] and mandibular fractures, with an approximate accuracy of 90%.[20]

Considering its performance for object detection, the YOLO model was used in this study for automated tooth detection and numbering on panoramic radiographs. This study aimed to evaluate the performance of the recent YOLO architecture, namely YOLO v4, by calculating model accuracy, precision, recall, F1 score, and detection time. The results of this study are expected to be an alternative for performing automated tooth numbering with high accuracy and fast detection time to assist dentists in reporting panoramic radiographs.

MATERIALS AND METHODS

The design of this study was observational descriptive to describe the performance of YOLO v4 in tooth numbering on panoramic radiographs. This research obtained ethical approval from the Universitas Airlangga Dental Hospital Ethical Committee, with certificate number 26/UN3.9.3/Etik/PT/2022.

Data set

The data set used in this study is secondary data in the form of digital panoramic radiographs from patients at the Academic Dental Hospital Universitas Airlangga from January 2016 to April 2022. All the image data were obtained using Instrumentarium OP200 D-1 Digital Panoramic System (Instrumentarium Dental) with the following acquisition parameters: 70 kVp, 8 mA, and 12 seconds. The sample size was calculated using the formula $(z^2p[1 - p])/d^2$, with the z score being a confidence level of 95% at 1.96, p estimating the population proportion at 0.5, and d being a margin of error of 0.05. As a result, a minimum of 384.16 samples was determined for this study. A total of 500 pieces of data were divided into training and testing data with a ratio of 80%:20%: 400 panoramic radiographs as training data and 100 panoramic radiographs as test data.

The inclusion criteria for the training data samples used were as follows: Panoramic radiographs with permanent dentition, healthy teeth condition, or caries lesions that did not blur the outline of the dental crown; complete dentition from the first incisors to the second molars in each region (third molars were not required to be present); and good-quality radiographs with focus on the teeth and alveolar bone. The following were excluded: mixed or primary dentition; tooth structure damage that destroys most of the tooth outline, for example, caries, root remnant, or root resorption; tooth restoration that changes the outline of its crown, for example, crown sheath or bridge denture; any teeth with root canal treatments or orthodontic appliances; dental implants; persistent teeth; crowding teeth;

supernumerary teeth; and edentulous samples. The test data criteria used in this study were almost the same as the training data criteria, except that edentulousness and the presence of root canal treatment were included in the inclusion criteria.

Ground truth

After the data set was collected, the annotation was carried out by labeling the object/tooth on each image in the data set by drawing a bounding box. The annotation was performed and validated by certified dental radiologists, which was considered ground truth in this study. The number of classes used is 32 according to the number of human teeth and with a numeration based on the universal numbering system.[21]

The bounding box was created using the LabelImg software (Tzutalin from Github) for graphic annotations. This annotation process generated a text file for each image. The file contained the annotation format on each line for each object as follows:

$$(\text{object} - \text{id})(x_{\text{center}})(y_{\text{center}})(w)(h)$$

The meaning of the text annotations, namely object-id is a class number (starting from 0 for the first class); x_{center} is the coordinates of the center of the bounding box to the object's width; y_{center} is the coordinates of the center of the bounding box to the object's height; w is the width of the bounding box; and h is the height of the bounding box.

Architecture model

The architecture model used in this research was YOLO v4. The YOLO v4 model has 3 structures: backbone, neck, and head. The backbone structure combines and forms image features on various types of detailed images consisting of cross-stage partial connection-Darknet53. The neck structure mixes and combines image features as an intermediate layer of image features to the prediction layer consisting of spatial pyramid pooling and the path aggregation network. The head structure functions to predict image features, generate bounding boxes, and predict classes consisting of YOLO v3 to detect single or multiple objects in image data. In addition, YOLO v4 uses new techniques for data augmentation, such as CutMix and Mosaic, to train data sets and improve detection capabilities.[17] An overview of the YOLO v4 model can be seen in Figure 1.

Training and testing

After the annotation was completed, the YOLO v4 model was trained with 400 training data points that were annotated to obtain the desired object detector model. Furthermore, testing was carried out on 100 panoramic images that were selected using a simple random sampling method for testing purposes. These data sets were independent of the training data set. In the testing process, information about the time it takes to detect each image was also recorded.

In this study, the training and testing process used Google Collaboratory (Google Colab) media. Google Colab is a cloud-connected machine learning research and education medium from Google. This medium provides adequate central and graphics processing units, critical components in deep learning development. This media can also be used in conjunction with Google Drive as a notebook storage area and for data sets. The input of the training process was panoramic images with 416 × 416 pixels. The learning rate for the training process was

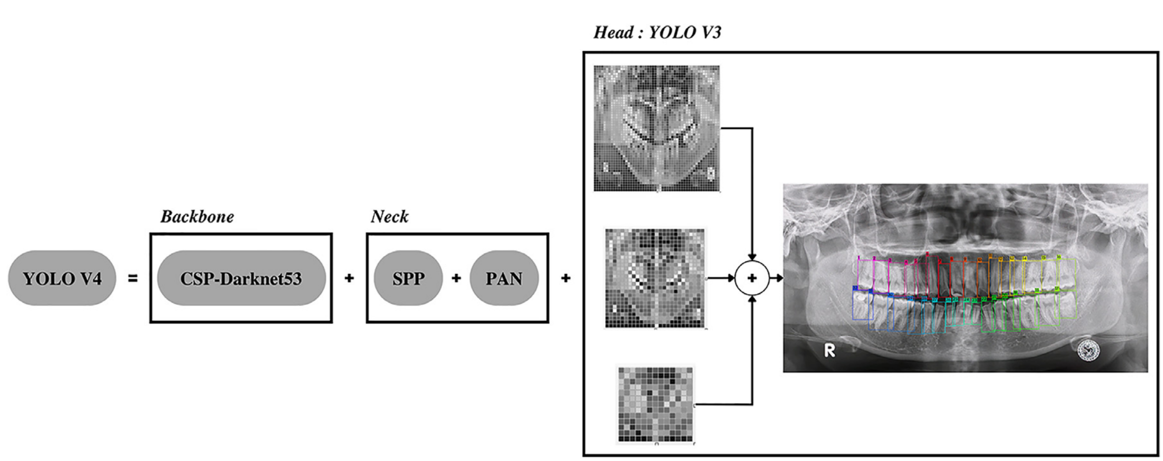

Fig. 1. You Only Look Once v4 model architecture. *YOLO*, You Only Look Once; *CSP*, Cross Stage Partial; *SPP*, Spatial Pyramid Pooling; *PAN*, Path Aggregation Network.

set to 0.001, with batch size and subdivision values of 64 and 16, respectively.

Data analysis

Data analysis was carried out on 100 test data points by calculating the confusion matrix value first on the results of the numeration test using YOLO v4, which was validated by manual assessment by certified oral and maxillofacial radiologists with 14 years of experience as ground truth. In the confusion matrix, there are 4 categories:

True positive (TP) is if the system detects the tooth object correctly according to ground truth.
False positive (FP) is if the system detects the object of a specific tooth, but misclassify the tooth number based on ground truth.
False negative (FN) is if the system failed to detect the tooth object.
True negative (TN) is if the system does not detect the missing tooth (edentulous area).

The overall accuracy value and each class was calculated using the formula (TP + TN) / (TP + TN + FP + FN). The results of the accuracy value were used to determine how YOLO v4 can perform tooth numbering and detect teeth, which were manually validated. Furthermore, the precision or positive predictive value was calculated as the ratio of a correctly positive value compared with all positive values (TP / TP + FP). Negative prediction value (NPV) was calculated as the ratio of a correctly negative value compared to all negative values (TN / TN + FN). Recall or sensitivity value was calculated as the ratio of a correctly positive value to all results that should be positive (TP / TP + FN). Specificity value was calculated by the ratio of a correctly negative value compared with all results that should be negative (TN / FP + TN). Finally, the F1 score, the harmonic mean of precision and recall, was calculated using the following formula: 2(Precision × Recall) / (Precision + Recall).

In addition to the overall evaluation, the accuracy and Matthews correlation coefficient (MCC) were calculated for each tooth. The MCC evaluates the agreement between the predicted and actual tooth numbers. The MCC score ranges from −1 to +1, where +1 represents a perfect prediction, and −1 represents a completely incorrect prediction. A score of 0 indicates that the model has no better ability to predict the binary class than random guessing. Additionally, the distribution of FP was further analyzed because this parameter is critical and directly affects the performance of the YOLO v4 model used in this study.

Next, the duration of manual tooth numbering on the test data performed by oral and maxillofacial radiologists was measured using a stopwatch. The detection time between YOLO v4 and the radiologist was statistically compared by calculating the mean detection time. The data were not normally distributed based on the normality of the data calculated using the Shapiro-Wilk test. Therefore, the Mann-Whitney test was used to statistically compare the mean detection time between YOLO v4 and human groups.

RESULTS

Based on the results of the testing carried out on 100 panoramic radiograph test data, the number of bounding boxes produced in 32 classes was 3487, as shown in Table I. An example of the correct results of tooth numbering on panoramic radiographs by YOLO v4 can be seen in Figure 2. There were no false negatives, which means that all teeth were detected with the formation of a bounding box in this research. This is an advantage of the YOLO v4 model that has been trained because it can achieve a high accuracy value. Based on the number of TP, TN, FP, and FN produced, the overall accuracy value of the YOLO v4 model in performing tooth numbering on panoramic radiographs was 88.50%. The precision, recall, and F1 scores were calculated as 87.7%, 100%, and 93.44%, respectively. The results of NPV and specificity showed 100% and 36.24%, respectively. The NPV and specificity values were not discussed in detail because these parameters focused on TN value. There were only a few edentulous areas (n = 230) compared with dentate areas (n = 2970) in the testing data, which contributed to the low specificity and thus may not accurately reflect the performance of YOLO v4.

Table I. Confusion matrix result

No. of test data	*Classes*	*Confusion matrix*		*Total bounding box*
100	32	TP	FP	3,487 (100%)
		2.858 (81.96%)	401 (11.50%)	
		FN	TN	
		0 (0%)	228 (6.54%)	

TP, true positive; *FP*, false positive; *FN*, false negative; *TN*, true negative.

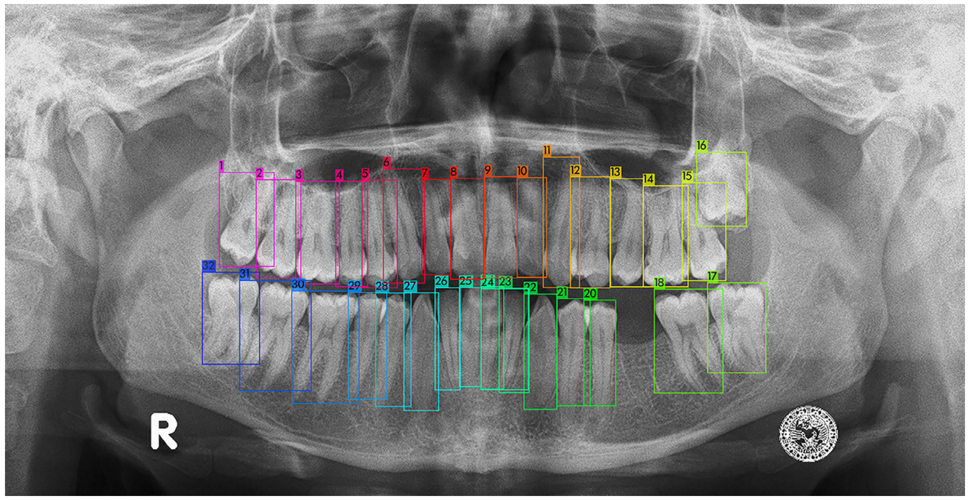

Fig. 2. Result of automated tooth numbering using You Only Look Once v4 on panoramic radiograph.

In addition to the overall accuracy value, the details of each class's accuracy and MCC value were also calculated; the results obtained are shown in Table II. From the table, on the maxilla, the highest accuracy value is attributed to tooth number 15, with an accuracy of 97.03%, and the lowest accuracy value is owned by tooth number 13, with an accuracy of 80.15%. In the mandible, tooth numbers 17 and 30 have the highest accuracy value, with an accuracy of 99.01%. The lowest accuracy value is attributed to tooth number 26, with an accuracy of 76.61%. In addition, the mandibular anterior region had the lowest accuracy compared with other regions. A similar pattern was also observed, with the lowest MCC values (0.16 < MCC < 0.22) in the mandibular anterior region.

A total of 401 FPs found in this study affected the accuracy value. The detection result is declared FP if the model successfully forms a bounding box on the tooth but misclassifies the tooth class (Figure 3A). A double bounding box was formed when 1 of the bounding boxes formed incorrectly classified the class of related teeth (Figure 3B). Errors that can occur include situations where the model successfully formed a bounding box on a tooth with multiple annotations

Table II. Accuracy of each class/tooth

Tooth		*Accuracy (%)*	*MCC*	*Tooth*		*Accuracy (%)*	*MCC*
Maxilla	1	95.15	0.87	Mandibula	17	99.01	0.97
	2	97.03	0.74		18	96.12	0.73
	3	92.45	0.50		19	96.12	0.88
	4	82.14	0.37		20	89.19	0.59
	5	80.91	0.58		21	86.36	0.60
	6	80.15	0.18		22	92.52	0.32
	7	87.96	0.49		23	78.33	0.17
	8	88.39	0.41		24	77.05	0.16
	9	90.65	0.29		25	83.05	0.19
	10	93.40	0.45		26	76.61	0.22
	11	80.91	0.19		27	84.75	0.21
	12	84.91	0.60		28	88.99	0.69
	13	87.04	0.39		29	83.48	0.41
	14	94.23	0.65		30	99.01	0.97
	15	96.12	0.44		31	93.27	0.65
	16	92.31	0.82		32	95.15	0.85

MCC, Matthews correlation coefficient; 1, maxillary right third molar; 2, maxillary right second molar; 3, maxillary right first; 4, maxillary right second premolar; 5, maxillary right first premolar; 6, maxillary right canine; 7, maxillary right lateral incisior; 8, maxillary right central incisior; 9, maxillary left central incisior; 10, maxillary left lateral incisior; 11, maxillary left canine; 12, maxillary left first premolar; 13, maxillary left second premolar; 14, maxillary left first; 15, maxillary left second molar; 16, maxillary left third molar; 17, mandibular left third molar; 18, mandibular left second molar; 19, mandibular left first; 20, mandibular left second premolar; 21, mandibular left first premolar; 22, mandibular left canine; 23, mandibular left lateral incisior; 24, mandibular left central incisior; 25, mandibular right central incisior; 26, mandibular right lateral incisior; 27, mandibular right canine; 28, mandibular right first premolar; 29, mandibular right second premolar; 30, mandibular right first; 31, mandibular right second molar; 32, mandibular right third molar.

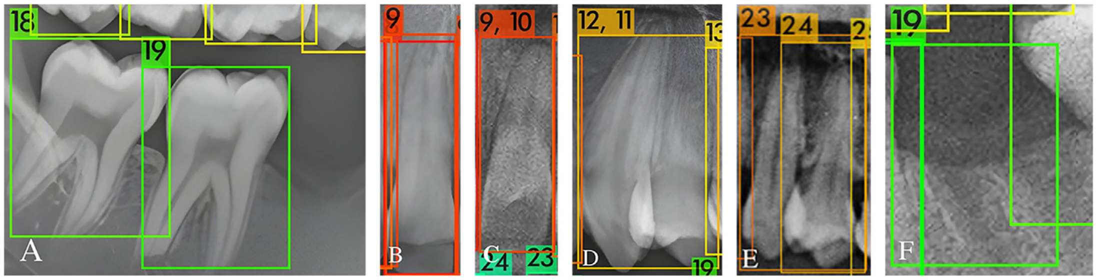

Fig. 3. Examples of false positive images. **(A)** The incorrect annotation on teeth number 47 and 48. **(B)** Double bounding box in tooth number 11. **(C)** A bounding box with a double annotations on tooth number 21. **(D)** A single bounding box on two different teeth. **(E)** Incorrect outline of bounding box on tooth number 23. **(F)** A bounding box on edentulous 36.

(Figure 3C); a single bounding box covered 2 teeth with double annotations (Figure 3D); incorrect outline of the bounding box on a single tooth (Figure 3E); and the presence of a bounding box on the edentulous area (Figure 3F). The percentage distribution of these FP forms can be seen in Table III. The table showed that the most common form of FP was the occurrence of double bounding boxes, which is 70.82%, and the second most was the occurrence of misclassification, which is 12.97%.

In this study, the training process using 400 pieces of training data with 32 classes took 148.78 hours. During the testing process on 100 pieces of data, the average time needed to detect each image was 20.58 ± 0.29 ms. The descriptive statistics of detection time comparison between YOLO v4 and humans are shown in Table IV. The duration of tooth numbering was very fast; it was done in a millisecond. YOLO v4 showed a significantly faster detection time than human manual annotation time ($P < .0001$).

DISCUSSION

Artificial intelligence in dental radiology is currently being developed to obtain fast and accurate performance in the classification, detection, or segmentation process to support an accurate diagnosis. This study deals with automatic tooth detection and numbering on panoramic radiographs as a first step in interpreting radiographic images with the help of artificial intelligence. This automatic tooth numbering can reduce the time required for manual tooth numbering and reduce the risk of errors caused by dentist fatigue. YOLO v4, as the architecture used in this study, is an advanced object detector with the best speed and accuracy of other available detectors.[17] To the best of our knowledge, this study is the first study on the use of YOLO v4 to perform multiobject detection for tooth numbering on panoramic radiographs.

Moreover, this study showed that computational power for developing the DL model could be achieved using cloud-based central processing units, namely Google Collaboratory or Google Colab (https://colab.research.google.com/). However, the limitation of Google Colab is the possibility of interruption to the training process if the browser window is closed or the connection is unstable. Although YOLO v4 has the potential to be used as an object detector in the field of dental radiology, it should be noted that human intervention is still needed to minimize the error from the system. Therefore, the clinician's role is essential when applying AI-based software.

In this study, YOLO v4 showed promising results in performing automated tooth detection and numbering in panoramic radiographs. The overall accuracy, precision, recall, and F1 score were 88.5%, 88.5%, 87.7%, 100%, and 93.44%, respectively. Although the results from this study are not comparable to similar studies because of different tasks and DL architectures, we reviewed the literature with similar purposes using other CNN architectures. As shown in Table V, the performance of YOLO v4 did not differ much from the other studies with comparable data set sizes.[22-24] Interestingly, YOLO v4 has the best recall parameter result because the architecture of YOLO has been developed

Table III. Distribution of FP cases

Case	*Amount*	*Percentage*
Misclassification	52	12.97%
Double bounding box	284	70.82%
Double classification on 1 tooth	48	11.97%
Merged classification of 2 teeth	10	2.49%
Bounding box is not at the margin of the tooth	5	1.25%
Edentulous detected as a tooth	2	0.50%

FP, false positive.

Table IV. Comparison of detection time between YOLO v4 and human

System	*Mean (s)*	*SD (s)*	*Min (s)*	*Max (s)*	*Significance*
YOLO v4	0.02058	0.00290	0,02051	0,02313	$P < .0001$
Specialist	9.928	2.773	4.000	15.73	

YOLO, You Only Look Once.

Table V. Comparison with studies of automated tooth numbering on panoramic radiograph using deep learning approach

Architecture	*Data set*	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1 score (%)*
YOLO v4 in this study	500	88,50	87,70	100	93,44
Faster R-CNN Inception v3[22]	303	77.4	84.5	75.5	79.75
ResNet[23]	895	98.33	97.35	96.81	97.08
DENTECT[24]	1.005	96.4	67.8	56.7	61.76

YOLO, You Only Look Once; *R-CNN*, region-with convolutional neural network.

especially for multiobject detection. The YOLO v4 model also showed as the second-best model according to the F1 score, which is widely considered 1 of the most critical parameters for evaluating the performance of a CNN model. Better F1 scores and other performance metrics compared with the YOLO v4 model were achieved through classification tasks on cropped tooth images from panoramic radiographs optimized using the data augmentation method.[23] These results suggest that YOLO v4 has significant potential for further development as an automated tooth numbering system.

The performance analysis of YOLO v4 in detecting each tooth object in a panoramic radiograph was evaluated using accuracy and MCC values. Our results showed that YOLO v4 has limitations in identifying the mandibular anterior region, as the lower incisor teeth in this region have similar tooth anatomic structures, such as a small crown and single root. Additionally, this region is often depicted with overlapping and blurred tooth structures and sometimes overlapping with the cervical spine structures.[25] This can make it challenging for YOLO v4 to accurately detect tooth objects in low-quality panoramic radiographs because of patient positioning errors. Therefore, further studies must focus on optimizing and improving YOLO v4′s performance, particularly in the mandibular anterior region.

In addition to being accurate, YOLO v4 could perform tooth numbering in a very short time. In this study, the average time required for the YOLO v4 model to perform tooth numbering for each image was 20.58 ± 0.29 ms, significantly faster than manual annotation ($P < .0001$). With units of milliseconds, this detection time can be considered a near real-time automated tooth numbering. Using U-Net architecture, Gerhardt et al. also showed fast teeth detection and segmentation in cone beam computed tomography images and obtained a detection time range of about 1.2 to 2.9 seconds.[26] The speed of YOLO v4 in tooth numbering could help medical personnel carry out their duties and reduce undesirable errors.

Besides the advantages of YOLO v4 as a fast and accurate object detector, this model needs high computational power for the training process because it uses DL approach. DL could indeed learn and process extensive data as its advantages, but by forming a large neural network with high computing requirements.[14] This caused the time required for the training process with many classes to be very long. The training time used for 32 classes in this study was 148.78 hours. It should be noted that, in this study, we used the default YOLO v4 architecture. Architectural modification of the YOLO model could shorten the training time and increase performance. Therefore, multidisciplinary collaboration with the informatics field is needed to further optimize the YOLO architecture, especially to improve the performance result.

In this study, the training data used were panoramic radiographs of complete teeth from the first incisors to the second molars to enable YOLO to recognize the ideal shape of each tooth. The result showed that the YOLO model could accurately recognize the tooth shape regardless of the internal appearance of the tooth, including the tooth with filled root canals. In the test data, panoramic radiography with edentulous sites was also included to obtain a TN value. However, it turned out that the system had difficulty classifying the teeth next to the edentulous sites. This result is consistent with previous studies using CNN for tooth detection and numbering in panoramic radiographs.[13] Thus, there were many FPs in the form of double bounding boxes (70.82%) or misclassification (12.97%) in the detection results. As a result, the number of FPs was relatively high (11.54%). Thus, variations in training data also greatly affected the test results on more varied test data. By overcoming this challenge and improving the quality and quantity of data sets, the overall performance of YOLO v4 can be improved by decreasing the number of FP. The data sets certainly affected the results because YOLO was an architecture that uses the concept of DL, which relies on learning data sets.[2]

The limitations of this study were the relatively small number of data sets compared with other similar studies[13] and the less diverse variety of the train data, so the number of FPs found was still quite large. It certainly had a direct impact on the overall performances. Second, the testing data were collected using simple random sampling, resulting in dentate and edentulous area imbalance. Although this study focused on tooth detection, a few edentulous areas can impact the model's specificity that does not reflect the model's performance. The purposive sampling method can give a better understanding of the result, for example, using a balanced number of teeth, root fragments, and/or edentulous areas. However, of course, the training data set should also be adjusted. Moreover, the absence of modifications to the YOLO v4 architecture makes the training process longer and less efficient. Further development of the application of YOLO v4 should also focus on another dentition stage, such as mixed dentition and primary dentition, and detecting pathologies, such as caries, impaction, or other abnormalities detection.

CONCLUSIONS

In conclusion, within this study's limitation, the DL approach's performance using the YOLO v4 architecture showed accurate and fast performance for automated tooth numbering in panoramic radiography. The performance parameter showed accuracy, precision, recall, and the F1 scores of 88.5%, 87.7%, 100%, and 93.44%, respectively. The mean average time needed was only 20.58 ± 0.29 ms, significantly faster than manual annotation. Further studies are needed to improve the performance by increasing the number and variety of data sets together with a multidisciplinary approach. Finally, this system is expected to assist dentists as a computer-assisted diagnostic tool in daily clinical practice to improve the dental health care system.

ACKNOWLEDGMENT

We thank Professor Wei Cheong Ngeow from the University of Malaya for editing this manuscript.

DECLARATION OF INTEREST

None.

REFERENCES

1. White SC, Pharoah MJ. *Oral Radiology (Principle and Interpretation)*. New York, NY: Elsevier Health Sciences; 2014.
2. Putra RH, Doi C, Yoda N, Astuti ER, Sasaki K. Current applications and development of artificial intelligence for digital dental radiography. *Dentomaxillofac Radiol*. 2022;51:20210197.
3. Baksi BG, Alpöz E, Soğur E, Mert A. Perception of anatomical structures in digitally filtered and conventional panoramic radiographs: a clinical evaluation. *Dentomaxillofac Radiol*. 2010;39:424-430.
4. Leite AF, van Gerven A, Willems H, et al. Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs. *Clin Oral Investig*. 2021;25:2257-2267.
5. Lin PL, Huang PY, Huang PW, Hsu HC, Chen CC. Teeth segmentation of dental periapical radiographs based on local singularity analysis. *Comput Methods Programs Biomed*. 2014;113:433-445.
6. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2:230-243.
7. Yu D, Hinton G, Morgan N, Chien JT, Sagayama S. Introduction to the special section on deep learning for speech and language processing. *IEEE Trans Audio Speech Language Process*. 2011;20:4-6.
8. Singh SP, Kumar A, Darbari H, Singh L, Rastogi A, Jain S. Machine translation using deep learning: an overview. Jaipur, India: IEEE; 2017:162-167.
9. Cao C, Liu F, Tan H, et al. Deep learning and its applications in biomedicine. *Genom, Proteom Bioinform*. 2018;16:17-32.
10. Lee H, Park M, Kim J. Cephalometric landmark detection in dental x-ray images using convolutional neural networks. In: *Proceeding of Medical Imaging 2017: Computer Aided Diagnosis*, Florida, USA: SPIE; 2017:494-499.
11. Wang Y, Zheng J. Real-time face detection based on YOLO. In: *ICKII*. 2018:221-224.
12. Miki Y, Muramatsu C, Hayashi T, et al. Classification of teeth in cone-beam CT using deep convolutional neural network. *Comput Biol Med*. 2017;80:24-29.
13. van Tuzoff D, Tuzova LN, Bornstein MM, et al. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofac Radiol*. 2019;48:20180051.
14. Oktay AB. Tooth detection with convolutional neural networks. Trabzon, Turkey: IEEE; 2017:1-4.
15. Chen H, Zhang K, Lyu P, et al. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Sci Rep*. 2019;9:1-11.
16. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2016:779-788.
17. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: optimal speed and accuracy of object detection. Available at: https://arxiv.org/abs/2004.10934. Accessed July 1, 2022.
18. Jiang L, Chen D, Cao Z, et al. A two-stage deep learning architecture for radiographic assessment of periodontal bone loss. *BMC Oral Health*. 2021;22:1-9.
19. Zhu T, Chen D, Wu F, Zhu F, Zhu H. Artificial intelligence model to detect real contact relationship between mandibular third molars and inferior alveolar nerve based on panoramic radiographs. *Diagnostics*. 2021;11:1664.
20. Son DM, Yoon YA, Kwon HJ, An CH, Lee SH. Automatic detection of mandibular fractures in panoramic radiographs using deep learning. *Diagnostics*. 2021;11:933.
21. Peck S, Peck L. Tooth numbering progress. *Angle Orhod*. 1996;66:83-84.
22. Kim C, Kim D, Jeong HG, Yoon SJ, Youm S. Automatic tooth detection and numbering using a combination of a CNN and heuristic algorithm. *Appl Sci*. 2020;10:5624.
23. Lin SY, Chang HY. tooth numbering and condition recognition on dental panoramic radiograph images using CNNs. *IEEE Access*. 2021;9:166008-166026.
24. Yüksel AE, Gültekin S, Simsar E, et al. Dental enumeration and multiple treatment detection on panoramic X-rays using deep learning. *Sci Rep*. 2021;11:1-10.
25. Peretz B, Gotler M, Kaffe I. Common errors in digital panoramic radiographs of patients with mixed dentition and patients with permanent dentition. *Int J Dent*. 2012;2012:584138.
26. Gerhardt M, Fontenele RC, Leite AF, et al. Automated detection and labelling of teeth and small edentulous regions on cone-beam computed tomography using convolutional neural networks. *J Dent*. 2022;122:104139.