The Sensitivity and Specificity of YOLO V4 for Tooth Detection on Panoramic Radiographs

Eha Renwi Astuti[1]*, Ramadhan Hardani Putra[1], Dina Karimah Putri[1], Nastiti Faradilla Ramadhani[1], Tengku Natasha Eleena Binti Tengku Ahmad Noor[2,3], Bintang Rahardjo Putra[4], Adhela Maheswari Pikantara Djajadiningrat[4]

1. Departement of Dentomaxillofacial Radiology, Faculty of Dental Medicine, Surabaya, East Java- Indonesia.
2. Malaysian Armed Forces Dental Officer, 609 Armed Forces Dental Clinic, Kem Semenggo, Kuching, Malaysia
3. Membership of Faculty of Dental Surgery, Royal College of Surgeon, Edinburgh University, United Kingdom.
4. Undergraduate Student, Faculty of Dental Medicine, Surabaya, East Java- Indonesia.

Abstract

This study aimed to evaluate the performance of You Only Look Once (YOLO) v4 architecture for tooth detection on panoramic radiographs by calculating the sensitivity and specificity of a trained model.

This observational descriptive study included 400 and 100 panoramic radiograph datasets that were divided into training and test data, respectively. Thirty-two permanent tooth objects were annotated based on the Fédération Dentaire Internationale numbering system. The annotated images were fed into a YOLO v4 model for the training process. Then, the trained model was tested on 100 panoramic images, which had 1,600 teeth and 1,600 edentulous areas. The sensitivity and specificity of YOLO v4 were calculated using a confusion matrix validated manually by a dental radiologist. YOLO v4 produced 1.534 and 1.568 true positive and true negative detections, respectively.

The sensitivity and specificity of YOLO v4 for tooth detection on the panoramic radiographs were 99.42% and 87.06%, respectively. Within the limitations of this study, YOLO v4 demonstrated high sensitivity for tooth detection on panoramic radiographs. Further improvement in specificity should focus on minimizing the number of false positives in tooth detection through dataset improvement and architecture modification.

Clinical article (J Int Dent Med Res 2023; 16(1): 442-446)

Keywords: Tooth detection, panoramic radiographs, YOLO v4, artificial intelligence, medicine.

Received date: 14 January 2023 **Accept date:** 10 February 2023

Introduction

A radiological report of a panoramic radiograph usually consists of interpretation and diagnosis information based on observations of the dental and maxillofacial statuses of the image and serves as a diagnostic aid in clinical diagnosis.[1] The interpretation of the teeth and the surrounding anatomical structures on a panoramic radiograph is an important initial step in detecting pathological abnormalities.[2] The first step in interpreting panoramic radiography is to determine the type of tooth or tooth numeration based on its anatomy and location. Manually numbering teeth on panoramic radiographic images with a large number of images is time consuming and may be prone to errors due to dentist fatigue.[3] In addition, the interpretation of radiographic results is very dependent on the expertise and skills of the dentist.[4]

Artificial intelligence (AI) is the ability of a machine to imitate human intelligence and behavior in carrying out certain tasks. In recent years, AI has experienced rapid development and has become extremely influential globally.[5] The development and application of AI are also emerging in the field of dentistry. The implementation of AI in dental health care can help dentists during daily practice. In the field of dental radiology, an AI system can be used to assist and detect abnormalities in radiographs.

This system is expected to reduce human error and shorten the duration of the radiological report-making process.[3] Deep learning (DL) includes a part of AI that can process large amounts of data, such as text, audio, and

Corresponding author:
Eha Renwi Astuti,
Faculty of Dental Medicine. Prof Dr. Moestopo Street no. 47
Surabaya, East Java- Indonesia.
E-mail: eha-r-a@fkg.unair.ac.id

images. One type of basic architecture that works well with DL is the convolutional neural network (CNN).[6] CNNs have been shown to have excellent abilities in image recognition and the evaluation of border and color features.[7] In dentistry, CNNs are used for cephalometric landmark detection,[8] tooth structure segmentation,[9] classification,[10] tooth detection, and numbering in the field of radiography.[11]

You Only Look Once (YOLO) is an algorithm that uses the CNN concept and is designed to detect objects quickly, accurately, and in real time.[12] YOLO v4 is a state-of-the-art detector that is faster and more accurate than any other detector available.[13, 14] In the literature, YOLO v4 has been widely used in the field of dental radiology to determine periodontal bone loss,[15] detect real contact relations between mandibular third molars and inferior alveolar nerves,[16] detect alveolar bones and mandibular canals,[17] detect permanent tooth bacteria,[18] detect prosthetic restorations,[19] and many other applications.

Several researchers have previously used various AI methods to detect and numerate teeth on radiographs. Zhang used modified CNN architecture to detect and classify teeth on panoramic radiographs with 1,000 periapical radiographs and obtained 98.3% sensitivity for the detection of multiple objects and 96.1% sensitivity for tooth numbering classification.[20] Tuzoff et al. conducted a similar study using VGG-16 CNN with 1,352 pieces of data to obtain 99.4% sensitivity for multiple object detection. Likewise, tooth numbering classification resulted in 99.8% sensitivity and 99.9%specificity.[11] The region convolutional neural network (R-CNN) can also be used for tooth detection and numbering on periapical radiographs. A study used 1,250 items of data for the training process and obtained a sensitivity value for multiple object detection of 98.5% and a sensitivity value for the classification of tooth numbering of 78.2%.[21] Leite et al. developed a combination of two deep CNNs with 3,576 pieces of data for detecting and segmenting teeth with 98.9% sensitivity.[3] Kim et al. also combined CNN architectures with a heuristic algorithm for automatic tooth detection and numbering.

This study showed a sensitivity value of 75.5% and a specificity value of 80.4% for dental detection and obtained a sensitivity value of 84.2% and a specificity value of 75.5% for tooth numbering.[22] Although the results of these studies are very promising, the improvement of automated tooth detection on panoramic radiographs is still needed.

The results of the model can be seen from standard performance measures based on sensitivity and specificity values.[23] However, studies reporting the performance of YOLO for detection based on its specificity are still limited. Sensitivity is the result of comparing the number of correct detections associated with a particular class with the total number of detections associated with that class. Specificity is the result of comparing the number of detections that are not related to a class with all detections that are not related to that class. Thus, this study aimed to focus on the detection ability of YOLO by evaluating its ability to detect teeth correctly, represented by sensitivity, and the ability to classify the edentulous area correctly, represented by specificity.

Materials and methods

The study, which aimed to describe the specificity and sensitivity of YOLO v4 for tooth detection on panoramic radiographs, had an observational descriptive design. This research obtained ethical approval from the Airlangga University Dental Hospital Ethical Committee, with certificate number 31/UN3.9.3/HRECC/PT/2022.

The dataset was divided into training data and testing data with a ratio of 80:20 (400 for training data and 100 for testing data). The test data consisted of a total of 1,600 tooth objects and 1,600 edentulous areas. The purpose of this test data partition was to balance the calculations of sensitivity and specificity through a confusion matrix. In this study, we used Google Colaboratory, which provided cloud computing resources, including Graphic Processing Unit (GPUs), for training and testing the dataset.

The training dataset comprised manually annotated teeth on panoramic images, which had bounding boxes and included the coronal, apical, mesial, and distal boundaries of each tooth. Each tooth represented each class of numbered teeth. There were 32 classes of teeth, which were numbered based on the Fédération Dentaire Internationale tooth-numbering system using the LabelImg software. The annotated data were trained to recognize the tooth numbers using

YOLO v4 architecture to learn the patterns or characteristics of each class until the model reached sufficient prediction for tooth detection.[7]

The testing process was performed on 100 independent panoramic images. The results from the automated tooth detection on the test data were manually validated by an expert and evaluated using a confusion matrix. The results from the bounding box on each tooth were classified as (1) true positive (TP), meaning the system detected the tooth object correctly, which was validated manually; (2) false positive (FP), which meant the system could detect the tooth, but it was wrongly annotated; (3) false negative (FN), meaning the system failed to detect the tooth object; and (4) true negative (TN), which indicated that the system successfully ignored the edentulous area.

The evaluations of the testing data were calculated to obtain the sensitivity and specificity values using the confusion matrix. The sensitivity test compared the correct number of detections in a class with the total number of detections in that class. The specificity test compared the number of detections that were not related to a class with all the detections that were not related to that class. The sensitivity and specificity values for each class were calculated using the following formulas:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Results

The testing process was carried out to determine the quality, capabilities, and weaknesses of the system model that was built. The main goal was to determine whether the built YOLO v4 model met the needs. After the testing process on 100 images, the trained YOLO v4 successfully created a total of 3,344 bounding boxes. Table 1 shows the confusion matrix from the testing process. YOLO v4 demonstrated high numbers of TP and TN values (1,534 and 1,568, respectively). However, a relatively high number of FPs was also present in the results, which showed a shortcoming of YOLOv4. Only nine FNs were found after the validation of the testing data. A higher number of bounding boxes in the test data compared with the ground truth was due to the fact that double bounding boxes were found after manual validation.

Test data (n = 100)		Detection Result	
No. tooth 1,600	No. edentulous 1,600	TP	FP
		1,534	233
		FN	TN
		9	1,568

Table 1. Confusion Matrix.

Then, the values of the confusion matrix were evaluated by calculating the sensitivity and specificity of YOLO v4 for automated tooth detection on the panoramic radiographs. The calculations of the sensitivity and specificity values from the test results of the YOLO v4 model were as follows:

$$\text{Sensitivity}$$

$$= \frac{\text{TP}}{\text{TP} + \text{FN}} \text{X } 100\%$$

$$= \frac{1534}{15134 + 9} \text{X } 100\% = 99.42\%$$

$$\text{Specificity}$$

$$= \frac{\text{TN}}{\text{TN} + \text{FP}} \text{X } 100\%$$

$$= \frac{1568}{1568 + 233} \text{X } 100\% = 87.06\%$$

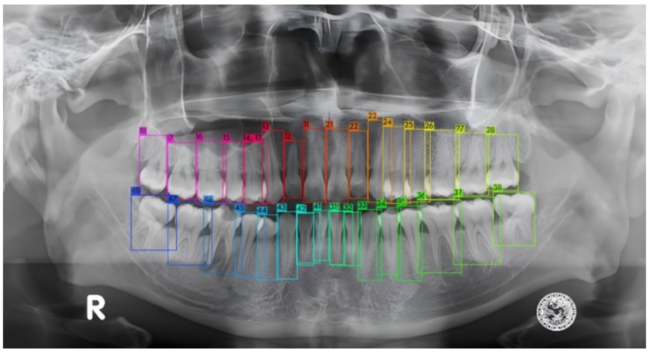

Figure 1. The results from the detection of teeth in a normal tooth arrangement.

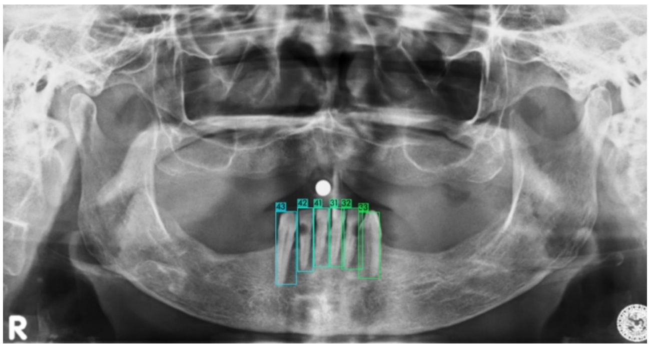

Figure 2. The results from the detection of teeth with an edentulous area.

According to the calculations of sensitivity and specificity, YOLO v4 produced 99.42% and 87.06% for sensitivity and specificity, respectively, for tooth detection on the panoramic radiographs. It can be seen in Figure 1 and Figure 2 that YOLO v4 successfully performed multi-object detection on panoramic radiographs with complete dentition and with an edentulous area. This indicated that YOLO v4 shows superior performance on automatic tooth detection, which is represented by the sensitivity value. However, specificity was lower compared with sensitivity. This indicated that the ability to classify the correct edentulous area was interrupted by the FP value.

Discussion

Measuring the performance of a created model is an important step in deep learning so it can be a consideration for choosing the best model. Measuring the performance of the model can be carried out simply by comparing the actual value with the predicted value. The confusion matrix is a performance measurement in the form of a table that describes the performance of the model on a series of test data whose actual values are known. In this study, sensitivity was calculated to describe the success rate of the YOLO v4 model in detecting objects. The sensitivity value obtained from the test results of 100 images was 99.42%. YOLO v4 has the best sensitivity parameter results because the YOLO architecture has been developed specifically for multiple object detection. In this study, this was proven by the system model successfully detecting all the dental objects (FN = 9 frames). In general, the built YOLO v4 was able to detect dental objects well in a normal tooth arrangement and with an edentulous area.

Such high detection results have also been achieved in previous studies on the detection and numeration of teeth on radiographs using CNNs. Zhang used modified CNN architecture to obtain 98.3% sensitivity,[20] and Chen et al. used R-CNN to generate a sensitivity value of 98.5%.[21] Tuzoff et al. used VGG-16 CNN and reported a sensitivity value of 99.4%.[11] Using a combination of two deep CNNs, Leite et al. achieved 98.9% sensitivity,[3] while Kim et al. obtained 75.5% sensitivity.[22] The performance results of YOLO v4 in our study, especially sensitivity, were not significantly different from other studies. YOLO v4 has the best sensitivity parameter results because the YOLO architecture is specifically designed to detect multiple objects.

The specificity value of the results from the automatic tooth detection on panoramic radiography using YOLO v4 and 500 datasets was 87.06%. This means that the ability of the YOLOv4 model to correctly ignore edentulous areas on panoramic radiographs was 87.06%. In this study, the specificity value was lower than the results of previous similar studies that used CNNs for tooth detection. Tuzoff et al. reported a specificity value of 99.9% for tooth numbering classification using VGG-16.[11] Kim et al. demonstrated automatic tooth detection and numbering using a combination of a CNN and a heuristic algorithm with 303 panoramic radiographs and obtained 75.5% specificity for tooth numbering.[22] Improving specificity is needed by developing YOLO models with sufficient training data in terms of quantity and dataset variation.

The limitations of this research were the relatively small number of datasets and the less diverse variations of the data train; therefore, there were many FPs. This had a direct impact on the specificity value, as the specificity value was lower than in previous studies. We used the standard YOLO v4 architecture, but architecture modifications could have improved the performance. In this study, the detection was carried out only on permanent teeth. For further studies, primary teeth or permanent teeth can be used. These two phases are more difficult because of the presence of tooth germs, so it is necessary to examine whether YOLO can still detect teeth in these phases.

Conclusions

Within the limitations of this study, YOLO v4 demonstrated high sensitivity for tooth detection on panoramic radiographs. Although the specificity value of YOLO v4 was 87.06%, it can be improved by focusing on minimizing the number of FPs in tooth detection through dataset improvement and architecture modification. The application of YOLO v4 can also be expanded for primary teeth or mixed dentition on panoramic radiographs.

Declaration of Interest

The authors report no conflict of interest.

References

1. Kurniati N, Firman RN, Pramanik F, Soewondo W, SS Inne. Analysis of the Position and Density of Condyle Head, Gonial Angle, and Ramus Height in Terms of Panoramic Radiographs. J Int Dent Med Res 2021; 14(3): 1079-1086
2. Baksi BG, Alpöz E, Soğur E, Mert A. Perception of anatomical structures in digitally filtered and conventional panoramic radiographs: A clinical evaluation. Dentomaxillofacial Radiology 2010 Oct;39(7):424–30.
3. Leite AF, Gerven A van, Willems H, Beznik T, Lahoud P, Gaêta-Araujo H, et al. Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs. Clin Oral Investig 2021 Apr 1;25(4):2257–67.
4. 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 Feb;113(2):433–45.
5. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: Past, present and future. Vol. 2, Stroke and Vascular Neurology. BMJ Publishing Group; 2017: p. 230–43.
6. Singh SP, KA, DH, SL, RA, & JS. International Conference on Computer, Communications and Electronics (Comptelix). IEEE; 2017.
7. Cao C, Liu F, Tan H, Song D, Shu W, Li W, et al. Deep Learning and Its Applications in Biomedicine. Vol. 16, Genomics, Proteomics and Bioinformatics. Beijing Genomics Institute; 2018: p. 17–32.
8. Lee H, Park M, Kim J. Cephalometric landmark detection in dental x-ray images using convolutional neural networks. In: Medical Imaging 2017: Computer-Aided Diagnosis. SPIE; 2017: p. 101341W.
9. Wang CW, Huang CT, Lee JH, Li CH, Chang SW, Siao MJ, et al. A benchmark for comparison of dental radiography analysis algorithms. Med Image Anal. 2016 Jul 1;31:63–76.
10. Miki Y, Muramatsu C, Hayashi T, Zhou X, Hara T, Katsumata A, et al. Classification of teeth in cone-beam CT using deep convolutional neural network. Comput Biol Med. 2017 Jan 1;80:24–9.
11. Tuzoff DV, Tuzova LN, Bornstein MM, et al. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofac Radiol*. 2019;48(4):20180051. doi:10.1259/dmfr.20180051
12. 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. IEEE Computer Society; 2016: p. 779–88.
13. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal Speed and Accuracy of Object Detection. 2020 Apr 22; Available from: http://arxiv.org/abs/2004.10934
14. Aini Q, Lutfiani N, Kusumah H, Zahran MS. Deteksi Dan Pengenalan Objek Dengan Model Machine Learning: Model Yolo. 2021; (6):1
15. Jiang L, Chen D, Cao Z, Wu F, Zhu H, Zhu F, et al. A Two-Stage Deep Learning Architecture for Radiographic Assessment of Periodontal Bone Loss. 2021; Available from: https://doi.org/10.21203/rs.3.rs-885130/v1
16. 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(9):1
17. Widiasri M, Arifin AZ, Suciati N, Fatichah C, Astuti ER, Indraswari R, et al. Dental-YOLO: Alveolar Bone and Mandibular Canal Detection on Cone Beam Computed Tomography Images for Dental Implant Planning. IEEE Access 2022;10:101483–94.
18. Kaya E, Gunec HG, Aydin KC, Urkmez ES, Duranay R, Ates HF. A deep learning approach to permanent tooth germ detection on pediatric panoramic radiographs. Imaging Sci Dent 2022 Sep 1;52(3):275–81.
19. Altan B, Gunec HG, Cinar S, Kutal S, Gulum S, Aydin KC. Detecting Prosthetic Restorations Using Artificial Intelligence on Panoramic Radiographs. Sci Program 2022 Nov 8;2022:1–6.
20. Zhang K, Wu J, Chen H, Lyu P. An effective teeth recognition method using label tree with cascade network structure. Computerized Medical Imaging and Graphics 2018 Sep 1;68:61–70.
21. 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):3840. doi:10.1038/s41598-019-40414-y
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. Applied Sciences (Switzerland) 2020 Aug 1;10(16):1
23. Gielis S, Moris P, Bittremieux W, de Neuter N, Ogunjimi B, Laukens K, et al. Detection of Enriched T Cell Epitope Specificity in Full T Cell Receptor Sequence Repertoires. Front Immunol 2019.