

## Original Article

# The application of artificial neural networks in the detection of mandibular fractures using panoramic radiography

Maryam Shahnava<sup>1</sup>, Hosein Mohamadrahimi<sup>2</sup>

<sup>1</sup>Department of Oral and Maxillofacial Radiology, School of Dentistry, Aja University of Medical Sciences, <sup>2</sup>Department of Orthodontics, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran

## ABSTRACT

**Background:** Panoramic radiography is a standard diagnostic imaging method for dentists. However, it is challenging to detect mandibular trauma and fractures in panoramic radiographs due to the superimposed facial skeleton structures. The objective of this study was to develop a deep learning algorithm that is capable of detecting mandibular fractures and trauma automatically and compare its performance with general dentists.

**Materials and Methods:** This is a retrospective diagnostic test accuracy study. This study used a two-stage deep learning framework. To train the model, 190 panoramic images were collected from four different sources. The mandible was first segmented using a U-net model. Then, to detect fractures, a model named Faster region-based convolutional neural network was applied. In the end, a comparison was made between the accuracy, specificity, and sensitivity of artificial intelligence and general dentists in trauma diagnosis.

**Results:** The mAP50 and mAP75 for object detection were 98.66% and 57.90%, respectively. The classification accuracy of the model was 91.67%. The sensitivity and specificity of the model were 100% and 83.33%, respectively. On the other hand, human-level diagnostic accuracy, sensitivity, and specificity were  $87.22 \pm 8.91$ ,  $82.22 \pm 16.39$ , and  $92.22 \pm 6.33$ , respectively.

**Conclusion:** Our framework can provide a level of performance better than general dentists when it comes to diagnosing trauma or fractures.

**Key Words:** Deep learning, dental radiography, mandibular fractures, panoramic radiography

Received: 10-Oct-2022  
Revised: 06-Nov-2022  
Accepted: 20-Dec-2022  
Published: 14-Feb-2023

Address for correspondence:  
Dr. Maryam Shahnava,  
School of Dentistry,  
Aja University of  
Medical Sciences, Misaq  
Complex, 13<sup>th</sup> East Street,  
Ajodanieh, Tehran, Iran.  
E-mail: maryamshahnavazi@  
gmail.com

## INTRODUCTION

The field of dentistry has undergone a significant transformation over the past few decades, and new technologies based on artificial intelligence (AI) have played an essential role in this transformation. The use of these intelligent technologies has been used as a powerful tool for the prediction and diagnosis of diseases as well as for the provision of appropriate treatment plans by dentists.<sup>[1,2]</sup> It

is possible for dentists to use AI technology to make more accurate diagnoses and better clinical decisions. AI is the ability of a system to imitate human-like intelligence.<sup>[3]</sup> Machine learning and deep learning are the main subbranches of AI, which mainly predict or make decisions about new data based on training by sample data or “training data.”<sup>[4]</sup>

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** WKHLRPMedknow\_reprints@wolterskluwer.com

**How to cite this article:** Shahnava M, Mohamadrahimi H. The application of artificial neural networks in the detection of mandibular fractures using panoramic radiography. Dent Res J 2023;20:27.

Access this article online	
	Website: <a href="http://www.drj.ir">www.drj.ir</a> <a href="http://www.drjjournal.net">www.drjjournal.net</a> <a href="http://www.ncbi.nlm.nih.gov/pmc/journals/1480">www.ncbi.nlm.nih.gov/pmc/journals/1480</a>

Dentists and maxillofacial surgeons use panoramic radiography as a standard diagnostic imaging method in their routine practices.<sup>[5,6]</sup> Previous studies have shown that physician training plays a crucial role in interpreting medical images. In addition, dental professionals evaluate radiographic images differently due to differences in knowledge, skills, and errors. As a result, their diagnosis may be different.<sup>[7]</sup> The ability of dental professionals to read panoramic radiographs varies, which can lead to erroneous diagnoses or nondiagnoses. A recent review shows dentists have low sensitivity in the radiographic diagnosis of dental caries with a diagnostic odds ratio of 0.24–0.42.<sup>[8]</sup>

Panoramic radiography can be used to detect various conditions, including mandibular lesions and traumas. The interpretation of trauma and mandibular injuries can be challenging even for experienced professionals due to their complexity. This is primarily due to the panoramic radiography procedure in which the source-detector assembly rotates around the patient's head so that all bony structures of the facial skeleton are superimposed.<sup>[9]</sup> For example, it has been reported that clinicians' diagnostic accuracy when using panoramic radiography for detecting condylar fractures is about 70%.<sup>[10]</sup> In spite of these problems, only a limited number of studies have used AI algorithms to detect maxillofacial traumas.<sup>[11,12]</sup>

Hence, we decided to investigate the use of deep learning to create an image analysis algorithm for automatically detecting mandibular trauma and fractures on panoramic radiographs in this study. We also compared the performance of our model with the diagnostic performance of general dentists. It is possible to use this algorithm in clinical practice as an aid in clinical decision-making.

## MATERIALS AND METHODS

### Study design

This is a retrospective diagnostic test accuracy study. A two-stage deep learning framework was used in this study. First, a U-net model was used to segment the region of interest, which was the mandible. Then, a model named Faster region-based convolutional neural network (Faster R-CNN) was applied to determine the presence and the position of fractures in the mandible through panoramic radiographs. The \*Aja University of Medical Sciences' ethics committee approved the study (IR.AJAUMS.REC.1400.204). In accordance

with the Checklist for Artificial Intelligence in Medical Imaging,<sup>[13]</sup> the study was conducted, and the results were reported.

### Data description

A total of 190 panoramic radiographs were collected from the patients. Due to limitations in acquiring relevant data, we gathered them from various resources, which were as follows.

1. Imam Hossein Hospital, Tehran, Iran
2. A private maxillofacial radiology center, Isfahan, Iran
3. Radiopaedia website (<https://radiopaedia.org/>)
4. Open-access biomedical image search engine provided by NIH (<https://openi.nlm.nih.gov/>).

All the images were exported to JPEG. The inclusion criteria were the presence of any sign of at least one fracture on the hard tissue of the mandible. The exclusion criteria were as follows.

1. Low-quality or corrupted images (e.g., blurry or noisy images)
2. Duplicate data
3. Data that cannot be identified as ground truth for any reason.

The pretreatment images of a patient were chosen if both pretreatment and posttreatment images were available.

### Diagnostic criteria and data labeling

For the first model, the aim was to segment the region of interest. For this purpose, a dentist annotated all 190 images by drawing polygons around the mandible using LabelMe software (the MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, Massachusetts, USA). Another dentist double-checked the annotated data and edited the polygons if there was any issue. To develop the trauma detection model, all radiographic images were annotated by two oral and maxillofacial radiologists through a consensus process. The location of the fracture was determined using bounding boxes with LabelMe software. In case of any disagreements, the final decision was made through consensus.

### Data partitions and preprocessing

Finally, 190 images were divided randomly into the training ( $n = 154$ ), validation ( $n = 18$ ), and test sets ( $n = 18$ ). The validation set was used for early stopping. Before feeding both models, all images were resized to  $224 \times 224$ . In addition, histogram equalization was used to adjust the contrast of an image based on its histogram. To enhance the object

detection model, the number of samples was increased by five times before the model was used. Applied augmentation techniques were as follows.

1. Random crop
2. Random color jitter (e.g., applying random changes in brightness, contrast, saturation, and hue)
3. Random affine (e.g., applying random rotation, translating, and scaling)
4. Adding random Gaussian noise
5. Random horizontal flip.

### Model architecture and training details

To develop our deep learning models, we have used the Python programming language and the PyTorch library to implement them. For the region of interest segmentation, we used a randomly initialized U-net model. The output of this model was used for training the object detector. We used the Faster R-CNN model based on ResNet101 pretrained on the COCO detection dataset for object detection.

To avoid overfitting, we decided to use the early stopping strategy. According to this strategy, the best weights of the model based on their performance on the validation set are stored and used in the next run of the model. Finally, to tune the hyperparameters, a randomized search strategy was used. A Tesla T4 Graphics Processor Unit was used to carry out the training procedure.

### Comparing results to the human-level detection

In the final step, the test set of panoramic radiographs and another 18 random radiographs without any sign of fractures were given to five general dentists (H.M.R., F.S., T.S., Z.P., and A.O.). Then, we asked them to classify images if there were any fractures in the samples or not. Then, the diagnosis of the AI model and dentists were compared to each other.

### Performance measurements and statistical analysis

For the segmentation model, our main performance measurements were intersection over union (IoU) and dice coefficient. For the object detection model, our main performance measurements were mean average precision calculated at the IoU threshold of 0.5 (mAP50) and 0.75 (mAP75). In addition, the accuracy, specificity, and sensitivity of AI and dentists were compared. If the AI model found any fracture in the image, we considered it a positive predicted sample. Otherwise, we considered it a negative predicted sample.

## RESULTS

The amount of IoU and dice coefficient for the segmentation model were 94.53% and 91.77% for the test set images, respectively. Three samples of model outcomes are presented in Figure 1. For the object detection model, mAP50 and mAP75 were 98.66% and 57.90% for the test set images, respectively. Figure 2 illustrates two sample outcomes of the whole framework. It seems the model currently deals with the overdiagnosis problem.

The accuracy of the classification model was determined to be 91.67%. Moreover, it was shown that the model had a sensitivity of 100 and a specificity of 83.33%. The confusion matrix of models' prediction and the ground truth is presented in Figure 3. When compared to human performance, the model overperformed human-level diagnosis regarding accuracy (91.67 vs.  $87.22 \pm 8.91$ ) and sensitivity ( $82.22 \pm 16.39$ ) on average. Only two out of five raters were able to diagnose trauma more accurately [Table 1].

## DISCUSSION

Misdiagnosis is one of the most common causes of malpractice in health care. Clinicians may misinterpret radiographic fractures for a variety of reasons, including fatigue, a lack of specialized expertise, and inconsistency in readings.<sup>[14,15]</sup> It has been reported that using an AI algorithm makes it possible to perform radiographic interpretation done by dentists.<sup>[16]</sup> Our aim was to develop a deep learning framework to detect and localize trauma and fractures in the mandible.

According to our results, we achieved a mAP50 of 98.66% and a mAP75 of 57.90% using our framework. It can be interpreted that the framework

**Table 1: Comparison of artificial intelligence model and dentists in diagnosing traumas**

Rater	Clinical experience (years)	Accuracy (%)	Sensitivity (%)	Specificity (%)
Dentist 1	5	86.11	88.89	83.33
Dentist 2	2	75	55.56	94.44
Dentist 3	2	83.33	77.78	88.89
Dentist 4	4	94.44	94.44	94.44
Dentist 5	8	97.22	94.44	100
Average		87.22±8.91	82.22±16.39	92.22±6.33
AI model		91.67	100	83.33

AI: Artificial intelligence

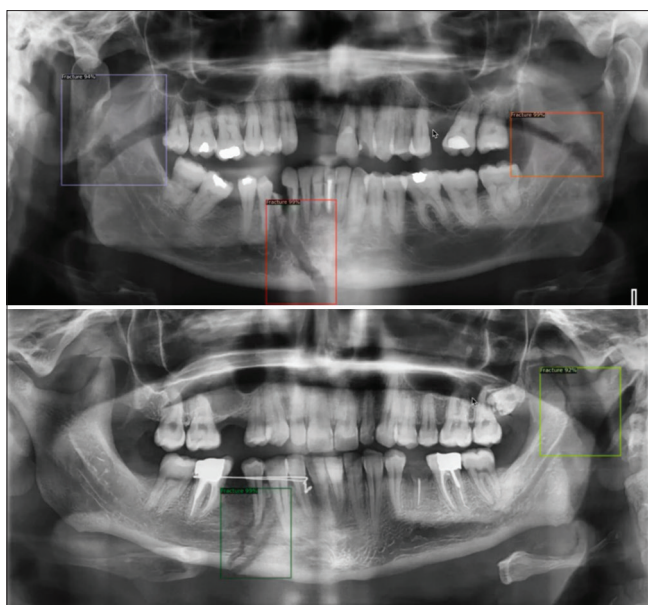


Figure 1: Samples of the segmentation model outputs.

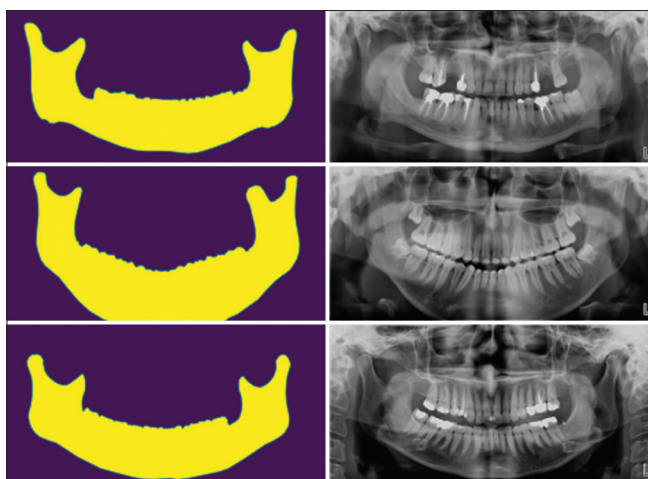


Figure 2: Samples of the final model outputs.

	Actual positive	Actual negative
Predicted positive	18	3
Predicted negative	0	15

Figure 3: Confusion matrix of the model for the diagnosis of the trauma.

has desirable performance in detecting fractures. Nevertheless, there may be some improvements that need to be made to the bounding box area. An increase in the number of samples in the dataset may be able to address this drawback of the model

with regard to detecting fracture extent. Moreover, the sensitivity of the model was 100%, which means the framework can detect any suspicious regions and hardly miss any fractured mandible.

Compared to general practitioners, the model was also outperformed in the case of sensitivity. In practice, most regions without access to oral and maxillofacial radiologists routinely rely on general practitioners to screen patients for mandibular fractures. Thus, general practitioners were included in the comparison of clinician performance with the model in this study. The outcome of the model suggests that it can be used as an assistant by practitioners for the purpose of screening patients who are potentially traumatized.

Similar to our work, Son *et al.*<sup>[11]</sup> tried to detect mandibular fractures using different variations of YOLO object detection algorithms on panoramic radiographs and compared the effect of various preprocessing techniques. They reported their model performance by classification sensitivity at best 79.4%, which was much lower than our framework. It is important to note that they used only 54 panoramic radiographs for training their model. On the other hand, Warin *et al.*<sup>[12]</sup> trained the Faster R-CNN and YOLO object detection models for a similar purpose using 855 images for the training procedures. They reported 87.94% and 86.12% of mAP50 Faster R-CNN and YOLO, respectively. The performance of this model was still lower than our models. It may be due to our region of interest segmentation algorithm, which eliminated nonrelevant areas as part of the process.

To improve the performance of our model, we extracted our region of interest, which was the mandible hard tissue, using a segmentation algorithm. It was intended to assist the object detection model in focusing only on the relevant parts of the image. This region of interest extraction strategy has already been used in AI in medical imaging and dentistry papers. As an example, similar to our study, Yüksel *et al.*<sup>[17]</sup> used a segmentation algorithm to separate each quadrant from the panoramic radiographs. Then, they fed each quadrant to an object detection model for the purpose of tooth enumeration.

Besides the performance of the model, one of the critical advantages of our study, as opposed to similar studies, was that we were able to obtain images from multiple sources of varying types of machines, radiation exposure conditions, sensors,

and image quality. This is because using data from different sources may help the deep learning model generalize better to the data samples from the sources outside our training set.<sup>[18]</sup> In other words, if a model is trained on datasets from a specific source, it will not be generalizable to a different population or a different source of images. As a result, it can only be used within the specific context in which it was developed.<sup>[19]</sup> Therefore, for the purpose of training and evaluating a model, it is recommended to use multiple independent datasets with different properties and populations.

A significant limitation of this study was the fact that we were unable to access the large volume of data that was required. As a first step in tackling this issue, we have collected data from public sources (e.g., PubMed) and pooled it with our data. This approach was already used in biomedical imaging to extend the size of the dataset.<sup>[20,21]</sup> Moreover, we added histogram equalization as a preprocessing step to enhance the image properties from various sources. Image contrast can be improved using histogram equalization in image preprocessing. To achieve this, it spreads out the most frequent intensities of the image, i.e., increases the intensity range of the image.<sup>[22]</sup> Consequently, the model would be able to detect fractures more easily.

## CONCLUSION

As a practical and adaptable tool, our framework also has the potential to provide a level of accuracy that could compete with general dentists when it comes to trauma or fracture diagnosis. The main limitation of the study was the small dataset. It is suggested that future studies to use more extensive datasets. Prospective and clinical studies are also recommended to evaluate the framework outcome in real-world scenarios.

### Financial support and sponsorship

Nil.

### Conflicts of interest

The authors of this manuscript declare that they have no conflicts of interest, real or perceived, financial or nonfinancial in this article.

## REFERENCES

- Mohammad-Rahimi H, Motamedian SR, Rohban MH, Krois J, Uribe SE, Mahmoudinia E, *et al.* Deep learning for caries detection: A systematic review. *J Dent* 2022;122:104115.
- Mohammad-Rahimi H, Motamedian SR, Pirayesh Z, Haiat A, Zahedrozegar S, Mahmoudinia E, *et al.* Deep learning in periodontology and oral implantology: A scoping review. *J Periodontol Res* 2022;57:942-51.
- Katne T, Kanaparthi A, Gotoor S, Muppirala S, Devaraju R, Gantala R. Artificial intelligence: demystifying dentistry – The future and beyond. *Int J Contemp Med Surg Radiol* 2019;4:D6-9.
- Zhang Z, Sejdic E. Radiological images and machine learning: Trends, perspectives, and prospects. *Comput Biol Med* 2019;108:354-70.
- Perschbacher S. Interpretation of panoramic radiographs. *Aust Dent J* 2012;57 Suppl 1:40-5.
- Molander B. Panoramic radiography in dental diagnostics. *Swed Dent J Suppl* 1996;119:1-26.
- Sabarudin A, Tiau YJ. Image quality assessment in panoramic dental radiography: A comparative study between conventional and digital systems. *Quant Imaging Med Surg* 2013;3:43-8.
- Schwendicke F, Tzschoppe M, Paris S. Radiographic caries detection: A systematic review and meta-analysis. *J Dent* 2015;43:924-33.
- Sklavos A, Beteramia D, Delpachitra SN, Kumar R. The panoramic dental radiograph for emergency physicians. *Emerg Med J* 2019;36:565-71.
- Timms L, Deery C. Do panoramic radiographs offer improved diagnostic accuracy over clinical examination and other radiographic techniques in children? *Evid Based Dent* 2021;22:110-1.
- Son DM, Yoon YA, Kwon HJ, An CH, Lee SH. Automatic detection of mandibular fractures in panoramic radiographs using deep learning. *Diagnostics (Basel)* 2021;11:933.
- Warin K, Limprasert W, Suebnukarn S, Inglam S, Jantana P, Vicharueang S. Assessment of deep convolutional neural network models for mandibular fracture detection in panoramic radiographs. *Int J Oral Maxillofac Surg* 2022;51:1488-94.
- Mongan J, Moy L, Kahn CE Jr. Checklist for artificial intelligence in medical imaging (CLAIM): A guide for authors and reviewers. *Radiol Artif Intell* 2020;2:e200029.
- Hallas P, Ellingsen T. Errors in fracture diagnoses in the emergency department – Characteristics of patients and diurnal variation. *BMC Emerg Med* 2006;6:4.
- Guly HR. Diagnostic errors in an accident and emergency department. *Emerg Med J* 2001;18:263-9.
- Suryani D, Shoumi MN, Wakhidah R. Object detection on dental x-ray images using deep learning method. *IOP Conf Ser Mater Sci Eng* 2021;1073:012058.
- Yüksel AE, Gültekin S, Simsar E, Özdemir ŞD, Gündoğar M, Tokgöz SB, *et al.* Dental enumeration and multiple treatment detection on panoramic X-rays using deep learning. *Sci Rep* 2021;11:12342.
- Krois J, Garcia Cantu A, Chaurasia A, Patil R, Chaudhari PK, Gaudin R, *et al.* Generalizability of deep learning models for dental image analysis. *Sci Rep* 2021;11:6102.
- Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. *Nat Med* 2019;25:44-56.
- Santosh K, Wendling L, Antani S, Thoma GR. Overlaid arrow

- detection for labeling regions of interest in biomedical images. IEEE Intell Syst 2016;31:66-75.
21. Mohammad-Rahimi H, Motamadian SR, Nadimi M, Hassanzadeh-Samani S, Minabi MA, Mahmoudinia E, *et al*. Deep learning for the classification of cervical maturation degree and pubertal growth spurts: A pilot study. Korean J Orthod 2022;52:112-22.
  22. Hum YC, Lai KW, Mohamad Salim MI. Multiobjectives bihistogram equalization for image contrast enhancement. Complexity 2014;20:22-36.