



# Automatic detection of the mental foramen for estimating mandibular cortical width in dental panoramic radiographs: the seventh survey of the Tromsø Study (Tromsø7) in 2015–2016

Journal of International Medical Research  
2022, Vol. 50(11) 1–17  
© The Author(s) 2022  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/03000605221135147  
[journals.sagepub.com/home/imr](https://journals.sagepub.com/home/imr)



Isak Paasche Edvardsen<sup>1</sup>, Anna Teterina<sup>2</sup>,  
Thomas Johansen<sup>3</sup> , Jonas Nordhaug Myhre<sup>3</sup>,  
Fred Godtliebsen<sup>4</sup> and  
Napat Limchaichana Bolstad<sup>2</sup> 

## Abstract

**Objective:** To apply deep learning to a data set of dental panoramic radiographs to detect the mental foramen for automatic assessment of the mandibular cortical width.

**Methods:** Data from the seventh survey of the Tromsø Study (Tromsø7) were used. The data set contained 5197 randomly chosen dental panoramic radiographs. Four pretrained object detectors were tested. We randomly chose 80% of the data for training and 20% for testing. Models were trained using GeForce RTX 2080 Ti with 11 GB GPU memory (NVIDIA Corporation, Santa Clara, CA, USA). Python programming language version 3.7 was used for analysis.

**Results:** The EfficientDet-D0 model showed the highest average precision of 0.30. When the threshold to regard a prediction as correct (intersection over union) was set to 0.5, the average precision was 0.79. The RetinaNet model achieved the lowest average precision of 0.23, and the

<sup>1</sup>Department of Physics and Technology, Faculty of Science and Technology, UiT The Arctic University of Norway, PO Box 6050, Langnes, N-9037 Tromsø, Norway

<sup>2</sup>Department of Clinical Dentistry, Faculty of Health Sciences, UiT The Arctic University of Norway, PO Box 6050, Langnes, N-9037 Tromsø, Norway

<sup>3</sup>Department of Energy and Technology, NORCE Norwegian Research Center, Sykehusveien 23, 9294 Tromsø, Norway

<sup>4</sup>Department of Mathematics and Statistics, Faculty of Science and Technology, UiT The Arctic University of Norway, PO Box 6050, Langnes, N-9037 Tromsø, Norway

### Corresponding author:

Napat Limchaichana Bolstad, Associate Professor of Oral and Maxillofacial Radiology, Department of Clinical Dentistry, Faculty of Health Sciences, UiT The Arctic University of Norway, PO Box 6050, Hansens str., 86, 9019, Langnes, N-9037 Tromsø, Norway.  
Email: [napat.l.bolstad@uit.no](mailto:napat.l.bolstad@uit.no)



precision was 0.64 when the intersection over union was set to 0.5. The procedure to estimate mandibular cortical width showed acceptable results. Of 100 random images, the algorithm produced an output 93 times, 20 of which were not visually satisfactory.

**Conclusions:** EfficientDet-D0 effectively detected the mental foramen. Methods for estimating bone quality are important in radiology and require further development.

## Keywords

Dentistry, artificial intelligence, panoramic radiography, machine learning, mental foramen, mandibular cortical width

Date received: 24 June 2022; accepted: 10 October 2022

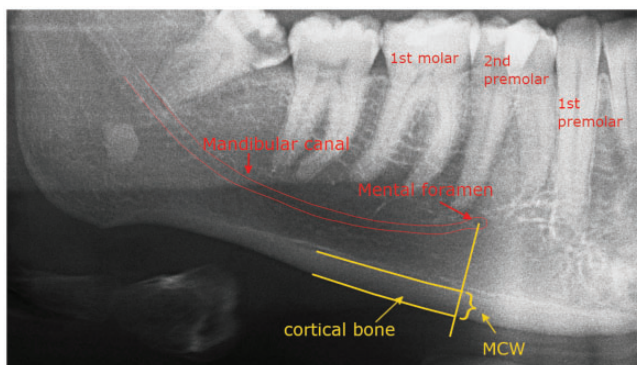
## Introduction

Dental panoramic radiographs (DPRs) are a standard diagnostic tool in dental practice because they provide valuable and comprehensive information about oral health and have a relatively low radiation dose. Approximately 16 million DPRs are annually taken in the general dental service in England and Wales,<sup>1</sup> 10 million in Japan,<sup>2</sup> and 5.55 million in Norway.<sup>3</sup> DPRs provide a comprehensive view of the jaw. In many situations, DPRs assist in providing information on the status of the jaw prior to further examination decisions such as those required in patients with jaw trauma, extensive dental or osseous lesions, tooth eruption, and developmental anomalies.<sup>4</sup>

The mental foramen (MF) is a clinically significant landmark for clinicians in several disciplines, such as dentists, oral and maxillofacial surgeons, emergency physicians, and plastic and reconstructive surgeons.<sup>5</sup> For example, to perform a mental nerve block (a type of anesthesia applied in the region of the MF), accurate determination of the position of the MF is paramount to avoid injury to nerves and blood vessels. The MF is also an essential landmark for measuring the mandibular cortical width (MCW) (Figure 1). A recent systematic

review concluded that the MCW measured on DPRs taken for routine dental diagnoses might also be useful as a screening tool for osteoporosis.<sup>6</sup> However, previous studies showed low reliability of the MCW when manually measured by different dentists.<sup>7,8</sup> Therefore, development of an automatic algorithm with which to measure the MCW was proposed.<sup>8</sup> Finding the correct position of the MF is the most important step in building such an automatic algorithm.

The MF is commonly located in the projection of the root apex of the second premolar or between the first and second premolar apices. Irregular tooth alignments or missing teeth make it challenging to determine the location of the MF.<sup>9</sup> Most patients have a single MF. However, variations such as supernumerary (accessory), curling, looping, or missing MFs are also encountered by clinicians. An accessory MF can occur because the mental nerve splits into several nerve fibers before the development of the MF, resulting in double, triple, or quadruple MFs. However, an accessory MF is more common than an absent MF.<sup>9</sup> An accessory MF is present in approximately 1% to 6% of people in different populations. A literature review showed that the MF was



**Figure 1.** Visualization of a region on a dental panoramic radiograph with essential markings such as the mental foramen, mandibular canal, and cortical bone. The MCW is measured between the border of the bone along the line drawn through the mandibular foramen perpendicular to the tangent of the lower edge of the bone.

MCW, mandibular cortical width.

detectable in approximately 87% to 94% of DPRs but clearly visible in only 49% to 64% of DPRs.<sup>10</sup> Jacobs et al.<sup>11</sup> reported detection of the MF in 94% of 545 DPRs; however, only 49% were considered visible by two independent observers (oral radiologists).

Studies on automatic image analysis from DPRs have been conducted in recent years, and such analysis is challenging because of the inherent complexity of DPRs. The challenge lies in identifying and recognizing specific structures and their morphometry. Morphometry involves assessment of the mandibular cortical bone and MCW for diagnosis of osteoporosis. Before considering an automatic system, Arifin et al.<sup>12</sup> created a manual computer-aided system for measuring the MCW based on gradient analysis of edges in 2006. Because the dentists had to manually determine the position of the MF, Arifin et al.<sup>12</sup> claimed that the experience of the examiners might greatly influence their decision, resulting in poor intra- and inter-examiner agreement. Other studies have focused on automatic segmentation of the mandible.<sup>13–15</sup> The approaches involved techniques such as horizontal integral

projections, use of a modified Canny edge detector, morphological operations, thresholding, and use of active contour models. Methods relying on isolation of the cortical bone region are prone to obstacles due to the unclear border of the bone and sometimes its irregular shape. Active contour models, or snakes, require a clear distinction of pixel intensity levels so that the snakes can follow the border of the mandible.<sup>16</sup> Aliaga et al.<sup>17</sup> considered these factors when developing an automatic system for computing mandibular indices in DPRs. The resulting algorithm computed indices inside two regions of interest that tolerated flexibility in sizes and locations, making this process adequately robust. However, they used morphological operations to locate the MF and reported that the proposed approach failed in 5% of 310 cases.<sup>17</sup> Lee et al.<sup>18</sup> used transfer learning for screening osteoporosis in DPRs with a limited data set (680 images). The highest overall accuracy achieved was 84%. Their results showed that transfer learning with pretrained weights and fine-tuning techniques could be helpful and reliable in the automated screening of osteoporosis.

The main objectives of this study were to explore the feasibility of detecting the MF in DPRs with pretrained object detection models and to investigate the possibility of developing an automatic measurement tool of the MCW.

## Materials and Methods

### Concepts

The main idea behind deep learning is the ability to solve tasks without explicitly designing a rule-based system to do so. Instead, deep learning resolves an assignment by learning from data and adapting to the present task. Hence, the data are often referred to as training data and are essential for proper functioning of deep learning models. A given model that is pretrained and has gained knowledge for a specific task can be further trained to resolve a similar task without the extensive need for data and computing time; e.g., a model trained to recognize apples can be trained to recognize pears. Fine-tuning is an approach of transfer learning that allows implementation of various strategies in which the model is initialized with knowledge (parameters) from a pretrained model. For instance, a model can be initialized with all the parameters from a pretrained model and adjust them regarding the present task, or a selection of these parameters can be set aside and not adjusted for the new task.

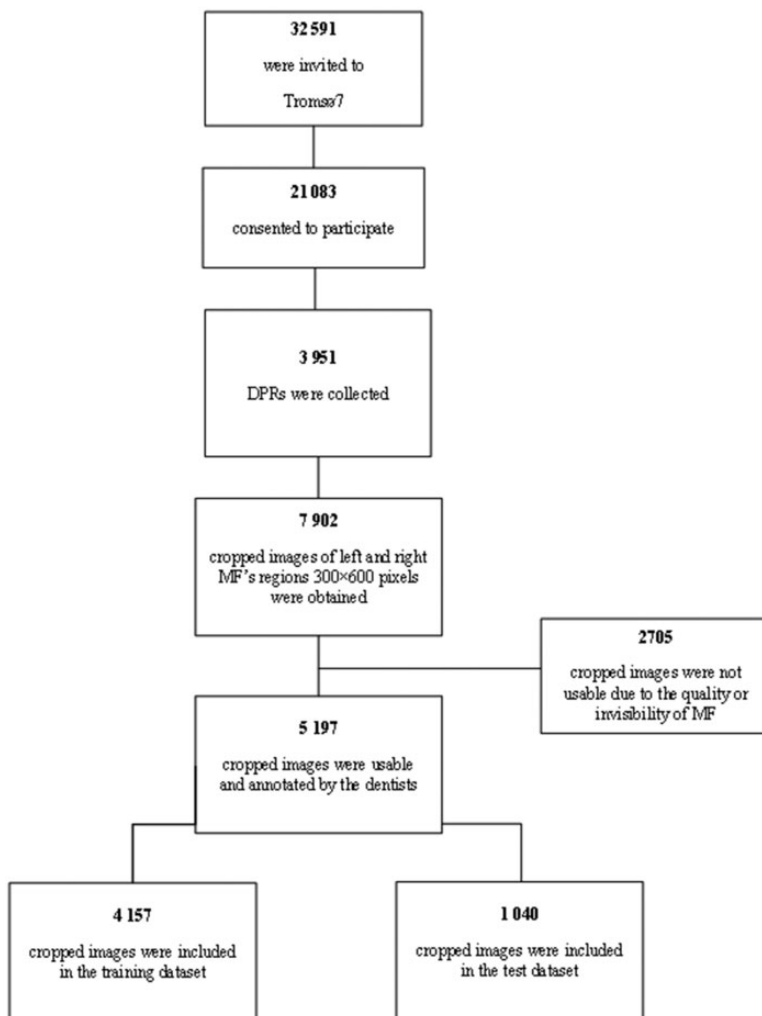
### Selection of DPRs and image annotation

The data set used in the present study consisted of DPRs taken during the seventh survey of the Tromsø Study (Tromsø7) from 2015 to 2016. The Tromsø Study is a population-based study carried out in repeated cross-sectional surveys.<sup>19</sup> Tromsø7 consisted of a questionnaire-based survey and clinical examinations, including DPRs. The survey enrolled 21,083 participants aged

40 to 99 years.<sup>20</sup> In total, 3951 DPRs were collected following the clinical dental examination (Figure 2). The DPRs were  $2821 \times 1376$  pixels, were in TIF format, and had 257 dots per inch. Knowing the dots per inch makes it possible to convert between pixels and physical size. In addition, two regions of interest were automatically cropped out for every image at an exact location. The resulting crops were  $300 \times 600$  (height  $\times$  width) pixels. The fixed cropping region did not always capture the jaw because of the varying patient positioning during the examination; such crops were discarded. Distorted images and images with obstructing artifacts were also rejected. Finally, the image was rejected if the experts did not recognize the position of the MF. Of 7902 crops, 5197 were usable (Figure 2), and the MF was annotated by the experts using VIA annotation software.<sup>21</sup> The data were divided into 4157 training images and 1040 test images (Figure 2). Two dentists experienced in oral radiology handpicked 100 “easy images” in which the MF was distinguishable and 101 “complex images” in which the MF was challenging to locate. These handpicked images were used to further analyze the model.

The dentists divided the workload, not annotating the same image to save time. However, to establish the intersection over union (IoU) between them, 706 images were annotated by both experts once. The IoU metric determines the amount of overlap between two boxes compared with their size (Figure 3). True positives are defined based on the IoU being greater than or equal to a threshold (i.e.,  $IoU(y^{(i)}, y^{(j)}) > T$ , where  $T$  is a defined threshold). The IoU between two bounding boxes  $A$  and  $B$  is defined in Equation 1.

$$IoU(A, B) = \frac{area(A \cap B)}{area(A \cup B)} \quad (1)$$



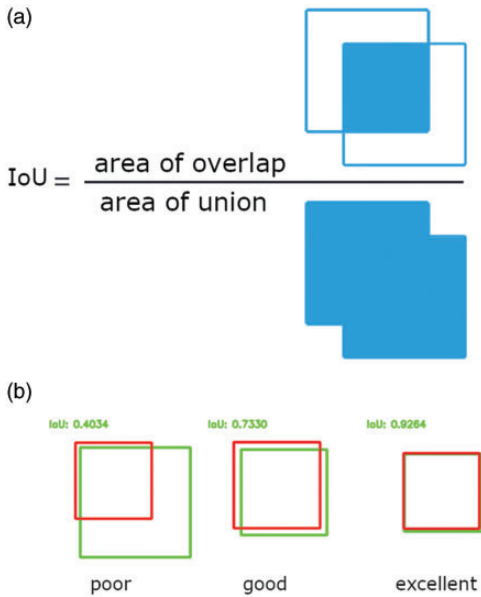
**Figure 2.** Flow chart of participants included in this study. DPR, dental panoramic radiograph; MR, mental foramen.

### *Data availability, ethical permissions, and funding*

The current study was based on data owned by the Tromsø Study, Department of Community Medicine, UiT The Arctic University of Norway. The data are available to interested researchers as approved by the Regional Committee for Medical and Health Research Ethics, the Norwegian Data Inspectorate, and the

Tromsø Study. Guidelines on data access and the application process are available at <https://uit.no/research/tromsostudy>.

The Tromsø Study was conducted in accordance with the World Medical Association Declaration of Helsinki.<sup>22</sup> The Regional Committee on Research and Ethics (REK North) and the Norwegian Data Protection Authority (Datatilsynet) approved the Tromsø Study. All participants provided written informed consent.



**Figure 3.** Performance evaluation. (a) Calculation of IoU and (b) poor IoU (0.40), good IoU (0.73), and excellent IoU (0.92). The poor IoU would not be considered a true positive if the threshold was 0.5. IoU, intersection over union.

In addition, we received separate approval from REK North (reference number 68128) and the Norwegian Centre for Research Data (NSD) to use the data from the Tromsø Study database.

The Arctic University of Norway (UiT), Northern Norway Regional Health Authority (Helse Nord RHF), the University Hospital of North Norway (UNN), and different research funds financed the Tromsø Study. The Department of Clinical Dentistry, Faculty of Health Science (UiT) fully financed the current study. We declare no conflict of interest in this study.

## Experiment

We performed a feasibility study showing that it is possible to fine-tune an object detector to be adequate in detecting the MF in X-ray images, which is useful for automatic measurement of the MCW.

Such a process needs to be able to measure at an appropriate location. Therefore, the first barrier is to detect the MF. The testing and fine-tuning were performed on a GeForce RTX 2080 Ti with 11 GB GPU memory (NVIDIA Corporation, Santa Clara, CA, USA).

The following models, pretrained on the COCO data set,<sup>23</sup> were “fine-tuned” to our data set using the TensorFlow framework:<sup>24</sup>

1. Faster R-CNN with ResNet50<sup>25</sup> as the backbone
2. CenterNet with HourGlass104<sup>26</sup> as the backbone
3. EfficientDet-D0 with EfficientNet-B0<sup>27</sup> as the backbone
4. RetinaNet<sup>28</sup> with ResNet50 as the backbone

Pretrained models (i.e., models that have already been given a data set of input and output pairs and taught to reproduce the correct output for each input) can be useful for solving other tasks involving data that are structured similarly to the original data set. Using pretrained models and training them on a different but similar data set is called fine-tuning. We placed the term “fine-tuning” in quotation marks above because the COCO data set is far from similar to ours, and “trained” hereafter implies “fine-tuned.”

## Experiment setup

For experiments on object detectors, the  $I_{oU}$  threshold  $\phi_{IoU}$  and confidence score threshold  $\phi_c$  used during non-maximum suppression (NMS) were set to 0.5 and virtually 0, respectively, for all models except CenterNet, which does not use NMS. Setting  $\phi_c$  to 0 means all proposals are accepted at the beginning of NMS. We assume that this is beneficial in challenging scenarios in which the predicted scores can be poor. Each model was trained with two

configurations (Setup 1 and Setup 2), and the results are presented in Table 1 and Table 2 (one for each configuration).

The batch size was set to 6 for all experiments (unless something else was specified), and we trained for 30 epochs. Because the training data comprised 4157 examples, processing 6 simultaneously (1 batch) resulted in approximately 693 gradient updates (training steps) to cycle through the training data once (1 epoch). Therefore, training for 30 epochs with a batch size of 6 required approximately 21,000 steps. Empirically, using the moving average of the trained parameters has been shown to be better than using trained parameters directly. However, we did not employ a moving average in any experiment because of technical limitations.

### Agreement between different models and dental experts

To evaluate each model, the test images were used to compute the accuracy

(i.e., the proportion of images for which the model outputs a correct bounding box). In addition, the handpicked images were used to evaluate the models under the circumstances in which the MF was and was not easy to distinguish. Both experts manually inspected these results because several images were not labeled. The experts reported whether they agreed with the predicted results. The experts performed the inspection of the results once, and the weighted kappa value was calculated.

### Procedure to estimate MCW

The procedure to estimate the MCW is briefly described in Algorithm 1. The procedure included the trained object detector. Further, the stop criterion in Algorithm 1 was a user-defined threshold representing the percentage of the line segment  $L$  overlapping with black pixels in the binary image  $I_b$  (Figure 4). The threshold was set to 0.7 in this study. After Algorithm 1

**Table 1.** Test results from the object detector with **Experimental Setup 1** of the object detectors presented in the Experiment subsection using the Tromsø7 data set described in the Selection of DPRs and image annotation subsection.

	mAP	mAP at IoU of 0.50	mAP at IoU of 0.75	AR at 100
Faster R-CNN	0.24	0.68	0.069	0.33
CenterNet	0.22	0.68	0.064	0.34
EfficientDet-D0	0.23	0.7	0.007	0.21
RetinaNet	0.21	0.62	0.010	0.46

mAP, mean average precision; IoU, intersection over union; AR, aspect ratio.

**Table 2.** Test results from the object detector with **Experimental Setup 2** of the object detectors presented in the Experiment subsection using the Tromsø7 data set described in the Selection of DPRs and image annotation subsection.

	mAP	mAP at IoU of 0.50	mAP at IoU of 0.75	AR at 100
Faster R-CNN	0.25	0.72	0.08	0.39
CenterNet	0.28	0.75	0.13	0.39
EfficientDet-D0	0.30	0.79	0.14	0.43
RetinaNet	0.23	0.64	0.01	0.47

mAP, mean average precision; IoU, intersection over union; AR, aspect ratio.

terminated, the width of the bone was defined as the distance between two parallel lines: the initial line and the resulting line. The distance was calculated with Equation 2, where  $c_1$  and  $c_2$  are the y-intercepts of the lines and  $m$  is the slope.

**Algorithm 1:** Method for bone width measurement, which was improved with an object detector. Please see the supplemental material for more information.

---

#### Identification of lowest edge of bone

1. Find MF's location  $P$  with an object detector
2. Convert image to grayscale and apply median filtering with kernel size 11
3. Apply a variance filter with kernel size 5, and follow with Canny edge detector
4. Use morphology to remove objects smaller than 150 pixels with a neighborhood of 500 pixels
5. Use probabilistic Hough transform<sup>29</sup> to retrieve possible line segments representing the lower bone edge, and save line segment  $L$  closest to  $P$

#### Identification of upper edge of bone (part 1)

1. Convert image to grayscale and apply variance filter with kernel size 8
2. Follow with exposure equalization to obtain  $I_v$
3. Apply a uniform filter with kernel size 11 to  $I_v$  to obtain  $I_m$
4. Calculate the binary image

$$I_b = \begin{cases} 1, & \text{if } I_m - I_v \leq \sigma^2 \\ 0, & \text{otherwise} \end{cases}$$

Where  $\sigma^2$  is the variance of  $I_v$ .

#### Identification of upper edge of bone (part 2)

- Initialize:

Place line segment  $L$  on  $I_b$

**while** stop criterion not fulfilled **do**

| Move  $L$  toward  $P$

**end**

---

$$d = \frac{|c_1 - c_2|}{\sqrt{1 + m^2}} \quad (2)$$

## Results

The configurations of the hyperparameters of the different algorithms are listed below.

### 1. Faster R-CNN

**Setup 1:** The stochastic gradient descent optimizer<sup>30</sup> was used with momentum 0.9 and  $L_2$  regularization ( $decay = 4 \times 10^{-4}$ ). The learning rate grew linearly from  $1 \times 10^{-2}$  to  $4 \times 10^{-2}$  for 2000 steps, then transitioned down using a cosine decay rule.<sup>31</sup> Rectified linear unit activation was

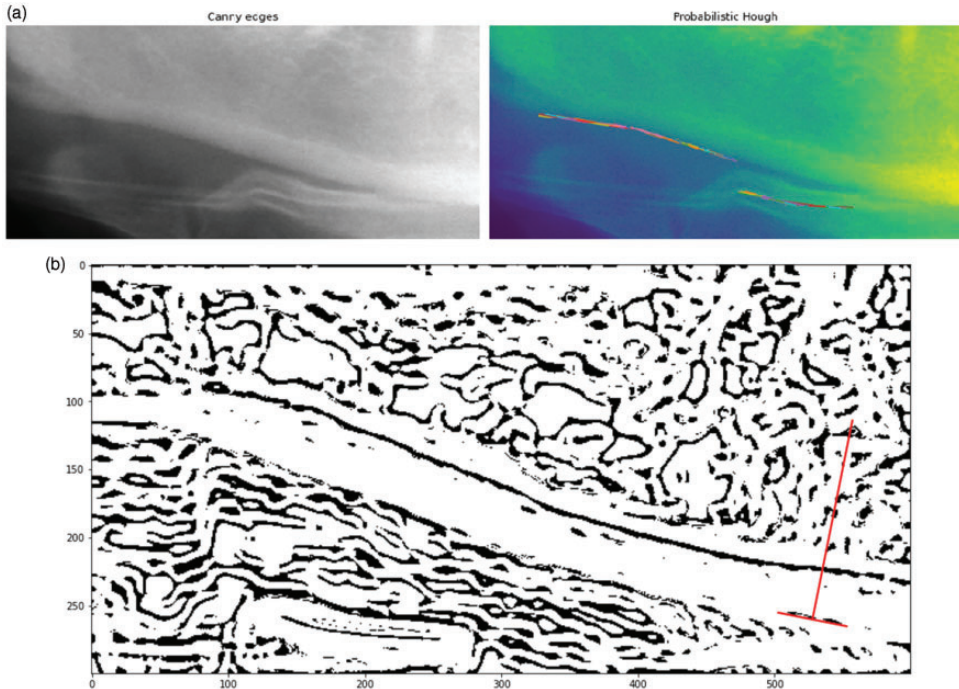
employed between convolutional layers. The anchor generator used aspect ratios (1/2, 1, 2) at scales (1/4, 1/2, 1, 2). The training images had a 50% probability of being flipped horizontally.

**Setup 2:** From the first setup, we changed to the following (the rest was unchanged): Adam optimizer ( $\epsilon = 1 \times 10^{-7}$ ) was used with a learning rate of  $2 \times 10^{-4}$ , which dropped to  $2 \times 10^{-4}$  at epoch 6, then to  $8 \times 10^{-5}$  at epoch 10 and  $4 \times 10^{-5}$  at epoch 15.

### 2. RetinaNet

**Setup 1:** The stochastic gradient descent optimizer<sup>30</sup> was used with momentum 0.9 and  $L_2$  regularization ( $decay = 4 \times 10^{-4}$ ).





**Figure 4.** Two cases in which the measuring algorithm needed improvements. (a) Canny edges will be retrieved from the left image and fed to the probabilistic Hough transform to find the best edge candidate. However, an artifact breaks the jawline, and the segment closest to the mental foramen here will be incorrect. (b) A case of a “pit” where the line segment has been initialized on the binary image\_1, satisfying the stopping criteria (overlapping black pixels).

The learning rate grew linearly from  $1 \times 10^{-2}$  to  $4 \times 10^{-2}$  for 2000 steps, then transitioned down using a cosine decay rule.<sup>31</sup> Synchronized batch normalization was added after every convolution with batch norm decay of 0.99 with  $\epsilon = 1 \times 10^{-3}$ . Rectified linear unit activation was employed but was capped at 6. Standard smooth L1 was the localization loss, and focal loss with  $\alpha = 0.25$  and  $\gamma = 2$  was the classification loss. The anchor generator used aspect ratios (1/2, 1, 2). The training images had a 50% probability of being flipped horizontally. The feature pyramid used minimum level 3 and maximum 7.

**Setup 2:** From the first setup, we changed to the following (the rest was unchanged): Adam optimizer,<sup>32</sup> where the

learning rate grew linearly from  $2 \times 10^{-4}$  to  $2 \times 10^{-3}$  for 2100 steps, then transitioned down using a cosine decay rule.

### 3. CenterNet

**Setup 1:** The Adam optimizer was used ( $\epsilon = 1 \times 10^{-7}$ ) for training with a constant learning rate of  $9.9 \times 10^{-4}$ . For the penalty-reduced pixel-wise logistic regression with focal loss,  $\alpha$  and  $\beta$  were set to 2 and 4, respectively. The loss was scaled by  $\lambda_{size} = 0.1$  and  $\lambda_{off} = 1.0$ . The training images had a 50% probability of being flipped horizontally, cropped, contrast-adjusted, or brightness-adjusted.

**Setup 2:** From the first setup, we changed to the following (the rest was

unchanged): The Adam optimizer was used ( $\epsilon = 1 \times 10^{-7}$ ) for training with a learning rate of  $5 \times 10^{-4}$  for 30 epochs, dropping  $10\times$  at epochs 18 and 24.

#### 4. EfficientDet-D0

**Setup 1:** The Adam optimizer ( $\epsilon = 1 \times 10^{-7}$ ) was used with a learning rate of  $2 \times 10^{-2}$  for 30 epochs, dropping  $10\times$  at epochs 18 and 24. Synchronized batch normalization was added after every convolution with batch norm decay of 0.99 and  $\epsilon = 1 \times 10^{-3}$ . Swish-1<sup>33</sup> (commonly called SiLu) activation was employed. Standard smooth L1 was the localization loss, and focal loss with  $\alpha = 0.25$  and  $\gamma = 1$  was the classification loss. The anchor generator used aspect ratios (1/2, 1, 2, 4). The training images had a 50% probability of being flipped horizontally. The feature pyramid used minimum level 3 and maximum 7.

**Setup 2:** From the first setup, we changed to the following (the rest was unchanged): Adam optimizer ( $\epsilon = 1 \times 10^{-7}$ ) was used with a learning rate of  $2 \times 10^{-4}$ , which dropped to  $1 \times 10^{-4}$  at epoch 6, then to  $8 \times 10^{-5}$  at epoch 10 and  $4 \times 10^{-5}$  at epoch 15. Random cropping was added as well, and the batch size was increased to 8.

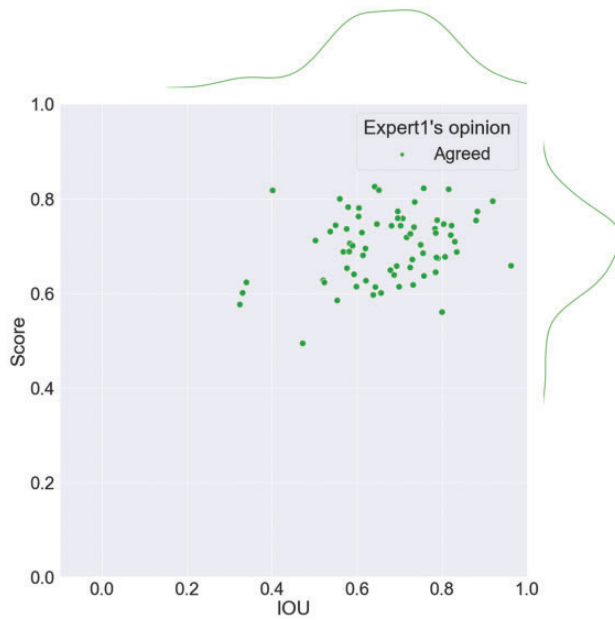
The different performances of the models with respect to detection of the MF are shown in Tables 1 and 2. EfficientDet-D0 clearly performed better in terms of average precision. This was true for both cases in which an IoU of 0.50 and 0.75 was the threshold for a prediction labeled as true positive. The three other models demonstrated relatively fair results. Notably, EfficientDet-D0 only uses a fraction of the number of parameters compared with the other models. However, CenterNet produced very similar results, and RetinaNet had a higher average recall regarding 100 detections. In addition, we noticed that the second configuration of

every model produced better mean average precision than the first.

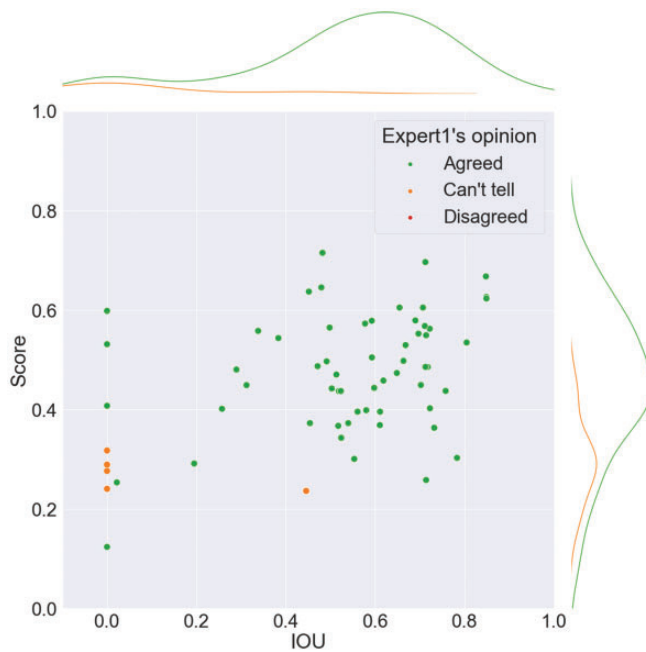
The initial hypothesis of this study was that existing models trained on the COCO data set could be fine-tuned to detect the MF. EfficientDet-D0 demonstrated sufficient precision and correct predictions at a threshold of 50% IoU compared with the other well-known models tested in this study; thus, the first hypothesis was concluded to be true. This conclusion was drawn by comparing the average precisions in Tables 1 and 2.

Figures 5, 6, 7, and 8 show the agreement between the dental experts when assessing the results of automatic detection of the MF on the handpicked images. This further investigation showed that the experts agreed on every prediction using the easy images and disagreed on some of the more complex images (Tables 3 and 4). It is apparent that annotating complex images is exceptionally challenging, and in the worst cases, annotation relies only on the best guess. When using three categories (“agree,” “unsure,” and “disagree”), the kappa value was 0.18, indicating slight agreement.<sup>34</sup> However, the kappa value can be misleading when the distribution between categories is unequal,<sup>35</sup> as in our case where only 10 of 101 predictions fell into the category “disagree.”

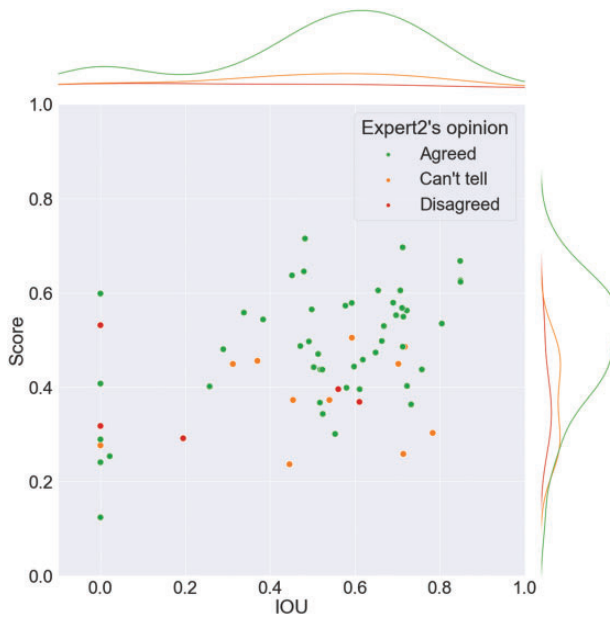
The second hypothesis followed the first, assuming the first was true: Can an object detector help accomplish automatic measurement of the MCW in DPRs? Using the results obtained from testing the first hypothesis, it was possible to make an algorithm that automates the measuring process. Of 100 random images (not necessarily in the training or test data set), the algorithm produced an output 93 times, 20 of which were not visually satisfactory. Therefore, the resulting algorithm needs improvement, and it is not yet generalized to handle image regions with high complexity even though the MF was found.



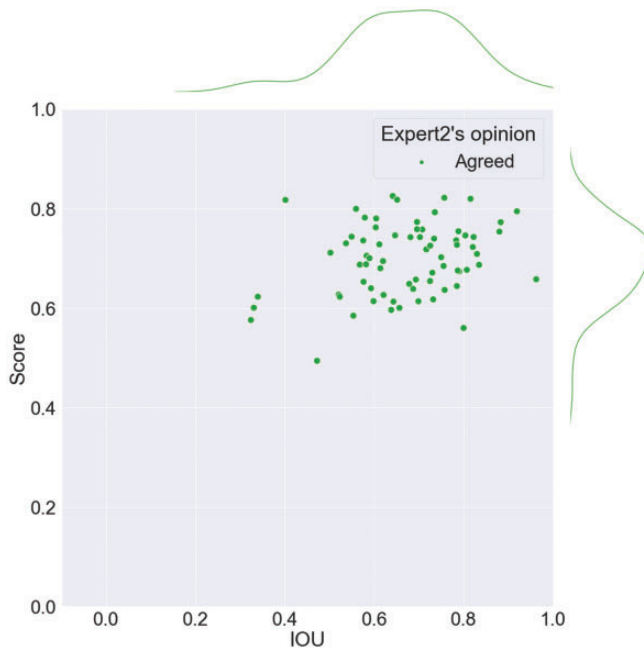
**Figure 5.** Predicted score versus IoU. Expert I has manually inspected the results and indicated whether they agree with the predicted results. IoU, intersection over union.



**Figure 6.** Predicted score versus IoU. Expert I has manually inspected the results and indicated whether they agree with the predicted results. IoU, intersection over union.



**Figure 7.** Predicted score versus IoU. Expert 2 has manually inspected the results and indicated whether they agree with the predicted results. IoU, intersection over union.



**Figure 8.** Predicted score versus IoU. Expert 2 has manually inspected the results and indicated whether they agree with the predicted results. IoU, intersection over union.

**Table 3.** Evaluation of 101 complex images by two dentists.

	Expert 2 (agree)	Expert 2 (unsure)	Expert 2 (disagree)
Expert 1 (agree)	67	12	7
Expert 1 (unsure)	7	5	2
Expert 1 (disagree)	0	1	0

**Table 4.** Combined evaluation of 101 complex images.

	Expert 1 (agree)	Expert 1 (disagree)
Expert 2 (agree)	91	7
Expert 2 (disagree)	0	3

Therefore, the algorithm was semi-capable of measuring the bone from visual reports, and the second hypothesis cannot be considered true.

## Discussion

Our investigation of the predictive ability of EfficientDet-D0 using easy images showed that both experts agreed with every prediction, even when several predictions had a relatively low IoU ( $<0.5$ ). However, even when the IoU was poor, overlap was still present between the ground truth and the predicted bounding box. Consequently, the prediction can result in a good suggestion of the position of the MF that largely agrees with dental experts.

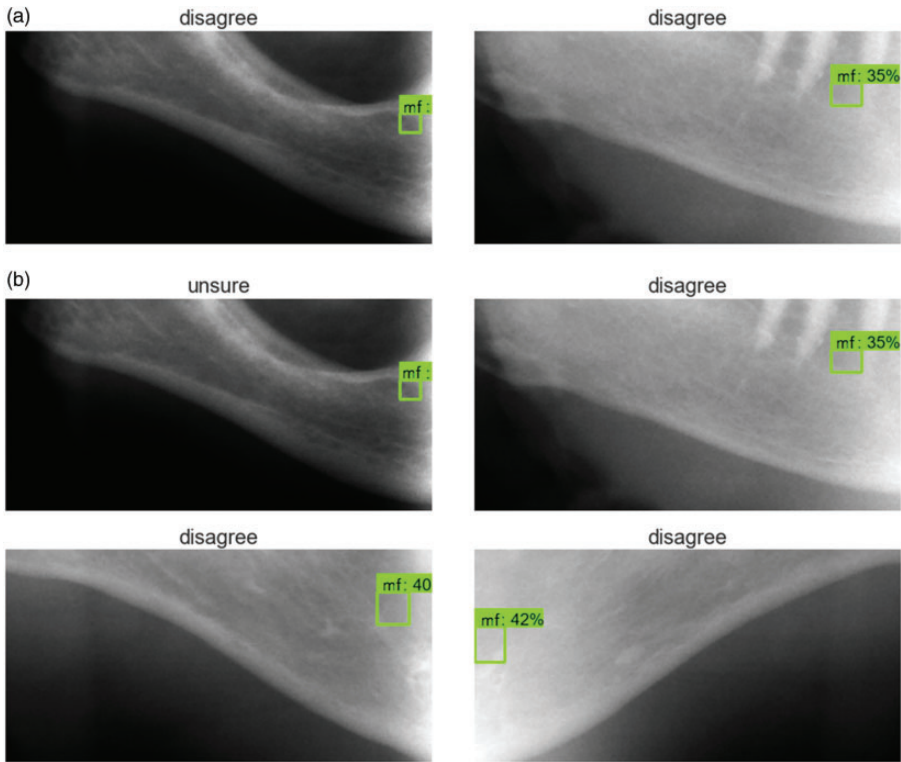
Analysis of the prediction ability of EfficientDet-D0 using complex images produced some interesting results. First, it should be stated that ground truths are not absolute. The experts agreed or were unsure about predictions for which the IoU was 0. All of these predicted bounding boxes lay on the mandibular canal next to the ground truth. Therefore, these predictions possibly contained the MF. The expert verdict explained that other

predicted regions seemed to contain part of the tooth's root apex, which could be a dark region in some cases and is challenging to distinguish from the MF.

Another case (Figure 7) disagreed with two predictions with a relatively high IoU ( $>0.5$ ), which may seem contradictory. This shows that the cropped images were challenging to label with a ground truth bounding box; labeling could only be accomplished by the best guess. Additionally, the entire image was available to aid the evaluation of a prediction in cases where the cropped images lacked information on other important landmarks, such as the premolars, which might explain this scenario. If no other landmarks are present when evaluating a prediction of the location of the MF, explainable artificial intelligence (AI) is needed to provide insight into the reason behind the predictions.<sup>36</sup> This would also allow for an uncertainty measure behind the model, which would benefit clinicians.

As stated above, not all the complex images that were handpicked for inference had ground truth bounding boxes. This occurred because the experts could not locate the MF when creating ground truth bounding boxes. These highly complex images were given to the model, and the experts evaluated the results (see Table 3). In one case, one expert disagreed with the prediction whereas the other expert was unsure. In another case, one expert was unsure but leaned toward disagreeing whereas the other expert disagreed with the prediction. These cases are depicted in Figures 9(a) and (b). For all cases shown in Figures 9(a) and (b), the experts concluded that the model annotated a part of the tooth's root apex, or the experts could not see the MF and therefore disagreed.

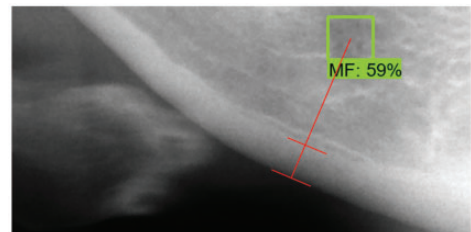
In this study, EfficientDet-D0 was used for inference, while EfficientDet-D7 is available with almost twice the mean average precision on the COCO data set. Future



**Figure 9.** Incorrect prediction from EfficientDet-D0 as judged by the (a) first and (b) second experts. MF, mental foramen.

research should utilize explainable AI to improve the trustworthiness of the AI system.<sup>36</sup>

When estimating the MCW, the proposed Algorithm 1 operates *fully* automatically given an image region. In Figure 10, we see that the algorithm effectively locates the MF and estimates the bone thickness automatically. However, our study did not compare automated MCW measurements with the actual osteoporosis status based on hip bone mineral density. Unlike our study, the OSTEODENT study used active shape models for automated MCW measurements and compared them with the actual diagnoses. The authors found that an MCW of <3mm could identify postmenopausal women with osteoporosis and stated that their findings were clinically



**Figure 10.** Results from Algorithm 1. The optimistic results are observed in this radiograph. The algorithm has stopped in a sweet spot immediately under porous textures. MF, mental foramen.

important.<sup>37</sup> Thus, in our further study, we plan to determine whether the algorithm measuring MCW can differentiate patients with osteoporosis diagnosed by bone mineral density measurements at the hip.

Moreover, to further improve the MCW measurement algorithm, steps can be taken to check whether the cropped image contains the lower edge of the bone beneath the MF. Alternatively, a dynamic image-cropping procedure based on other landmarks could be implemented to ensure the presence of the edge. Otherwise, the algorithm measures other structures close to the MF, not the bone. Another issue to consider is that the initial lines can become stuck in a “pit” in the binary image  $I_b$  (see Algorithm 1) if the lower border of the bone is unclear. In the most challenging scenario, the binary image  $I_b$  can contain artifacts overlapping either the line’s pathway when traveling toward the MF or other image areas. These artifacts cause an unclear upper bone border, terminating the algorithm at an incorrect location, or the line segment suggested in the first place will suffer (see Figure 4(a)). Therefore, we should also consider possibilities other than the MCW for screening osteoporosis, especially transfer learning, which could be used to learn attributes of DPRs labeled as affected, given a sufficiently large data set.

The use of AI in medicine and dentistry aims at smooth integration into the workflow and saving of time. However, one limitation of AI is that its accuracy depends on the quality of data from which the algorithm has learned. If a human decision is used as a “ground truth,” common human bias can be introduced into the algorithm. In this study, expert assessments were considered a “ground truth.” The proper “ground truth” for the location of the MF should be either a cadaver mandible or a cone-beam computed tomography scan. However, the former would not be approved by an ethics committee, and the latter was unavailable for our study.

Moreover, medical images with multiple overlapping artifacts can lead to unreliable algorithm outputs, which is also a

limitation.<sup>38</sup> Very few studies to date have focused on the automated location of the MF on DPRs. Discussing our findings in the context of previous research is challenging because of the different AI methods used<sup>39</sup> and the lack of guidelines for comparing different studies using AI in medicine, notably dentistry.<sup>38</sup>

## Conclusion

The MF is an important landmark for dental practitioners. Detecting its location on a DPR is the most important step in measuring the MCW, which can be a useful index for osteoporosis screening. In this study, EfficientDet-D0 showed sufficient precision and correct predictions of MF locations. Moreover, it was possible to merge EfficientDet-D0 with the previously made MCW measurement algorithm. This indicates the feasibility of fully automatic measurement of the MCW for osteoporosis detection.

## Authors’ contributions

**Isak Paasche Edvardsen:** Conceptualization, Methodology, Validation, Formal analysis, Writing – Original Draft.

**Anna Teterina:** Methodology, Writing – Review, Editing, Visualization, Supervision.

**Thomas Johansen, Jonas Nordhaug Myhre, Fred Godtliebsen:** Conceptualization, Methodology, Validation, Writing – Review, Editing, Supervision.

**Napat Limchaichana Bolstad:** Conceptualization, Validation, Writing – Review, Editing, Supervision, Project Administration, Funding Acquisition.

## Declaration of conflicting interests

The authors declare that there is no conflict of interest.

## Funding

The authors disclosed receipt (pending publication) of the following financial support for the

research, authorship, and/or publication of this article: UiT The Arctic University of Norway, Northern Norway Regional Health Authority (Helse Nord RHF), the University Hospital of North Norway (UNN), and different research funds financed the Tromsø Study. The Department of Clinical Dentistry, Faculty of Health Science, UiT The Arctic University of Norway fully financed the current study.

### ORCID iDs

Thomas Johansen  <https://orcid.org/0000-0003-3572-4706>

Napat Limchaichana Bolstad  <https://orcid.org/0000-0002-4276-6720>

### Supplemental material

Supplemental material for this article is available online.

### References

- Rushton VE and Horner K. The use of panoramic radiology in dental practice. *J Dent* 1996; 24: 185–201. DOI: 10.1016/0300-5712(95)00055-0.
- Taguchi A, Tsuda M, Ohtsuka M, et al. Use of dental panoramic radiographs in identifying younger postmenopausal women with osteoporosis. *Osteoporos Int* 2006; 17: 387–394. DOI: 10.1007/s00198-005-2029-7.
- Saxeboel G and Olerud HM. *Radiation use in Norway. Useful use and good radiation protection for society, humans and the environment*. 2014. Statens strålevern.
- White SC and Pharoah MJ. *Oral radiology: principles and interpretation*. 7th ed. St. Louis: Elsevier, 2014.
- Laher AE, Wells M, Motara F, et al. Finding the mental foramen. *Surg Radiol Anat* 2016; 38: 469–476. DOI: 10.1007/s00276-015-1565-x.
- Calciolari E, Donos N, Park JC, et al. Panoramic measures for oral bone mass in detecting osteoporosis: a systematic review and meta-analysis. *J Dent Res* 2015; 94: 17S–27S. DOI: 10.1177/0022034514554949.
- Ledgerton D, Horner K, Devlin H, et al. Radiomorphometric indices of the mandible in a British female population. *Dentomaxillofac Radiol* 1999; 28: 173–181. DOI: 10.1038/sj.dmfr.4600435.
- Devlin C, Horner K and Devlin H. Variability in measurement of radiomorphometric indices by general dental practitioners. *Dentomaxillofac Radiol* 2001; 30: 120–125.
- Hasan T. Mental foramen morphology: a must know in clinical dentistry. *J Pak Dent Assoc* 2012; 21: 167–172.
- Greenstein G and Tarnow D. The mental foramen and nerve: clinical and anatomical factors related to dental implant placement: a literature review. *J Periodontol* 2006; 77: 1933–1943. DOI: 10.1902/jop.2006.060197.
- Jacobs R, Mraiwa N, Van Steenberghe D, et al. Appearance of the mandibular incisive canal on panoramic radiographs. *Surg Radiol Anat* 2004; 26: 329–333. DOI: 10.1007/s00276-004-0242-2.
- Arifin AZ, Asano A, Taguchi A, et al. Computer-aided system for measuring the mandibular cortical width on dental panoramic radiographs in identifying postmenopausal women with low bone mineral density. *Osteoporos Int* 2006; 17: 753–759.
- Abdi AH, Kasaei S and Mehdizadeh M. Automatic segmentation of mandible in panoramic X-ray. *J Med Imag* 2015; 2: 044003. DOI: 10.1117/1.JMI.2.4.044003.
- Kavitha MS, Asano A, Taguchi A, et al. The combination of a histogram-based clustering algorithm and support vector machine for the diagnosis of osteoporosis. *Imaging Sci Dent* 2013; 43: 153–161. DOI: 10.5624/isd.2013.43.3.153.
- Naik A, Vinayak Tikhe S, Bhide SD, et al. Designing a feature vector for statistical texture analysis of mandibular bone. *Indian J Sci Technol* 2016; 9: 1–4. DOI: 10.17485/ijst/2016/v9i33/96305.
- Cootes TF, Taylor CJ, Cooper DH, et al. Active shape models-their training and application. *Comput Vis Image Underst* 1995; 61: 38–59.
- Aliaga I, Vera V, Vera M, et al. Automatic computation of mandibular indices in dental panoramic radiographs for early osteoporosis detection. *Artif Intell Med* 2020; 103: 101816. DOI: 10.1016/j.artmed.2020.101816.



18. Lee K-S, Jung S-K, Ryu J-J, et al. Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. *J Clin Med* 2020; 9: 392. DOI: 10.3390/jcm9020392.
19. Jacobsen BK, Eggen AE, Mathiesen EB, et al. Cohort profile: the Tromsø Study. *Int J Epidemiol* 2012; 41: 961–967. DOI: 10.1093/ije/dyr049.
20. The seventh survey of the Tromsø study (Tromsø7), [https://uit.no/research/tromsoudersokelsen/project?p\\_document\\_id=705235&pid=706786](https://uit.no/research/tromsoudersokelsen/project?p_document_id=705235&pid=706786) (2022, accessed 27 April).
21. Dutta A and Zisserman A. The VIA Annotation Software for images, audio and video. In: *MM '19: Proceedings of the 27th ACM International Conference on Multimedia*, Nice, France, 21–25 October 2019, pp.2276–2279. DOI: 10.1145/3343031.3350535.
22. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA: World Medical Association* 2013; 310: 2191–2194.
23. Lin TY, Maire M, Belongie S, et al. *Microsoft COCO: Common Objects in Context*. Cham: Cham: Springer International Publishing, 2014, pp.740–755.
24. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems, <https://www.tensorflow.org/> (2015, accessed 30 May 2022).
25. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016, pp.770–778. DOI: 10.1109/CVPR.2016.90.
26. Newell A, Yang K and Deng J. *Stacked hourglass networks for human pose estimation*. Cham: Cham: Springer International Publishing, 2016, pp.483–499.
27. Tan M and Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, CA, USA, 9–15 June 2019, pp.6105–6114.
28. Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision* 2017, pp.2980–2988.
29. Kiryati N, Eldar Y and Bruckstein AM. A probabilistic Hough transform. *Pattern Recognit* 1991; 24: 303–316. DOI: 10.1016/0031-3203(91)90073-E.
30. Qian N. On the momentum term in gradient descent learning algorithms. *Neural Netw* 1999; 12: 145–151. DOI: 10.1016/S0893-6080(98)00116-6.
31. Loshchilov I and Hutter F. SGDR: Stochastic Gradient Descent with Warm Restarts.
32. Kingma DP and Ba J. Adam: A Method for Stochastic Optimization.
33. Ramachandran P, Zoph B and Le QV. Searching for Activation Functions.
34. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174. DOI: 10.2307/2529310.
35. Szklo M and Nieto FJ. *Epidemiology: Beyond the Basics*. 4 ed. Sudbury: Sudbury: Jones & Bartlett Learning, LLC, 2018.
36. Adadi A and Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access* 2018; 6: 52138–52160.
37. Devlin H, Allen P, Graham J, et al. The role of the dental surgeon in detecting osteoporosis: the OSTEODENT study. *Br Dent J* 2008; 204: E16–E561. DOI: 10.1038/sj.bdj.2008.317.
38. Pethani F. Promises and perils of artificial intelligence in dentistry. *Aust Dent J* 2021; 66: 124–135. DOI: 10.1111/adj.12812.
39. Kats L, Vered M, Blumer S, et al. Neural network detection and segmentation of mental foramen in panoramic imaging. *J Clin Pediatr Dent* 2020; 44: 168–173.