



Two-step deep learning models for detection and identification of the manufacturers and types of dental implants on panoramic radiographs

Yoshiko Arijji¹ · Kaoru Kusano² · Motoki Fukuda¹ · Yo Wakata² · Michihito Nozawa¹ · Shinya Kotaki¹ · Eiichiro Arijji³ · Shunsuke Baba²

Received: 9 April 2024 / Accepted: 7 August 2024

© The Author(s), under exclusive licence to The Society of The Nippon Dental University 2024

Abstract

The purpose of this study is to develop two-step deep learning models that can automatically detect implant regions on panoramic radiographs and identify several types of implants. A total of 1,574 panoramic radiographs containing 3675 implants were included. The implant manufacturers were Kyocera, Dentsply Sirona, Straumann, and Nobel Biocare. Model A was created to detect oral implants and identify the manufacturers using You Only Look Once (YOLO) v7. After preparing the image patches that cropped the implant regions detected by model A, model B was created to identify the implant types per manufacturer using EfficientNet. Model A achieved very high performance, with recall of 1.000, precision of 0.979, and F1 score of 0.989. It also had accuracy, recall, precision, and F1 score of 0.98 or higher for the classification of the manufacturers. Model B had high classification metrics above 0.92, exception for Nobel's class 2 (Parallel). In this study, two-step deep learning models were built to detect implant regions, identify four manufacturers, and identify implant types per manufacturer.

Keywords Deep learning · Oral implant · Classification · Detection · Panoramic radiograph

Introduction

The advent of osseointegration implants marked a major advance in dental prosthetics [1–3]. Currently, treatment options that combine aesthetic and functional rehabilitation are now available [1]. As the reliability and success rate of dental implant treatment improves, the number of patients increases [4–6]. Because of the continued development of dental implants, hundreds of manufacturers have produced thousands of types of implants since 2000 [7, 8]. Implants consist of fixators, abutments, superstructures, and screws. Their internal structure and installation tools vary between manufacturers and systems [2–4]. If biological

complications (peri-implantitis) occur and the continued use of the implant becomes difficult, the implant must be removed [3, 4]. To enable easy removal, it is necessary to understand the types of implants [4, 9]. If mechanical complications (fracture or loosening of prosthesis or fixation components) occur, the types of implants must be identified to determine the appropriate treatment [4, 10–13]. If implant surgery or the prosthetic procedure was performed at another hospital, or if the previous dentist cannot be contacted because the clinic is closed, it may be difficult to obtain the details of implants [9]. It has also been reported that it is difficult to identify the types of implants, which makes continuous maintenance impossible and forces the removal of implants [14]. Clinicians have used radiographic images along with clinical information to confirm implant types [9]. Even if implant characteristics can be identified on radiographic images, clinicians without sufficient experience and knowledge of implant systems may have difficulty identifying them by type [9].

Deep learning has resulted in great advances in the medical field. It has demonstrated excellent performance, particularly in the classification of medical images [15]. It improves performance by learning large amounts of data and

✉ Yoshiko Arijji
ariji-y@cc.osaka-dent.ac.jp

¹ Department of Oral Radiology, Osaka Dental University, 1-5-17, Otemae, Chuo-ku, Osaka 540-0008, Japan

² Department of Oral Implantology, Osaka Dental University, Osaka, Japan

³ Department of Oral and Maxillofacial Radiology, Aichi-Gakuin University, Nagoya, Japan

adjusting neural network weights as required [16]. Recently, deep learning researchers in dentistry have reported the preliminary value of clinical applications and reliable performance comparable to that of dentists [9]. Studies have been conducted in which deep learning has been used to identify implant types [10, 12, 13, 17, 18]. The results have demonstrated that the performance of implant type identification on periapical and panoramic radiographs can be significantly improved by turning a large number of deep parameters [10, 12, 13, 17, 18]. In these studies, researchers evaluated 3–11 types of implants [10, 12, 13, 17]. However, in actual clinical practice, thousands of types of dental implants with similar shapes are sold [16].

Two major challenges exist in creating deep learning models to clarify implant types on panoramic radiographs. The first challenge is to create a model that can identify thousands of types of implants. Since the 2000s, hundreds of manufacturers and thousands of implant systems have been produced, and the number continues to grow [8, 19]. Kong et al. were challenged to create a model to identify 130 different implants [16]. However, the identification of each implant type was insufficient with only 73–74% of top-1 accuracy, and was predictable within the top-5. This may be because the performance of the classification deep learning model decreased as the number of classes increased [20]. To overcome this limitation, a different approach is required.

The second challenge is to build a system to automatically detect and diagnose implant regions. In many studies in which implant types have been identified, researchers have created classification models by manually cropping the image patches of the implant regions on panoramic radiographs. With this method, even if a practical model was created, the input test data must be a manually cropped image of the implant area, rather than the panoramic radiograph itself. This challenge may be solved using neural networks for object detection. You Only Look Once (YOLO) is a neural network for object detection. It divides the entire image into grids, and predicts the bounding boxes and corresponding classes for each grid [21]. It is open source and widely applied because of its high detection sensitivity and precision [22]. If the implant area can be detected, an automatically cropped image of that area can be easily prepared. Furthermore, by identifying the type of implant by manufacturer, it will be possible to construct a model for more accurate type identification.

The purpose of this study is to develop deep learning models that can automatically detect implant regions on panoramic radiographs and identify implant types. First, we created a model (model A) to detect implant regions and classify the manufacturers on panoramic radiographs using the detection neural network YOLO. After preparing the image patches that cropped the implant regions detected by

model A, second, we created model B to identify implant types by manufacturer using classification networks.

Materials and methods

This study was approved by the author's University Ethics Committee (Approval No. 111195–0). The requirement for informed consent from the study subjects was waived by the Committee because of the retrospective study design. Ethics approval was also obtained from the co-authors' institution for the use of external testing dataset (Approval No. 586). All procedures were conducted in accordance with the ethical guidelines of the 1964 Helsinki Declaration and its subsequent Ethical Guidelines for Medical and Health Research.

Dataset

Subjects were selected from patients who underwent pre-operative panoramic radiography at the Department of Oral and Maxillofacial Radiology, the author's University Hospital. They received oral implants at the Department of Oral Implantology from January 2016 to August 2023. The final study consisted of 730 patients (267 men and 463 women, age range 18 to 88 years, mean 56.1 ± 13.7 years). A total of 1,574 panoramic radiographs containing 3,675 implants were included. No images were excluded due to image quality. Table 1 shows the number of patients, panoramic radiographs, and oral implants by manufacturer, and number of implants by type. The implant manufacturers were Kyocera, Dentsply Sirona, Straumann, and Nobel Biocare. The implant types were Kyocera's Finesia, Dentsply Sirona's Astra Tech OsseoSpeed EV and TX, Straumann's Standard (S)/Standard Plus (SP) and Bone Level (BL)/Bone-Level Tapered (BLT), and Nobel's Active, Replace, and Parallel.

Panoramic radiographs were taken using Veraviewepocs (Morita, Kyoto, Japan) at 73–76 kV, 10 mA, and 8.0 s. The panoramic radiographs were cropped to include the upper and lower dentition, downloaded from the image database of the hospital at 1000×700 pixels, 96 dpi, and 24-bit depth, and saved in JPEG format.

The study procedure is shown in Fig. 1. First, deep learning model A was created to detect oral implants and identify the manufacturers on panoramic radiographs using YOLO-v7 neural network. After preparing the image patches that cropped the implant regions detected by model A, next, models B was created to identify the implant types per manufacturer using EfficientNet.

Table 1 Summary of subjects

Manufactures	No of patients	No of panoramic radiographs	Implant types	No of implants per type	Total No of implants
Kyocera	86	159	Finesia	263	263
Dentsply Sirona	263	588	Astra EV	1427	1650
			Astra TX	223	
Straumann	242	528	Straumann S/SP	20	1097
			Straumann BL/BLT	1077	
Nobel Biocare	139	299	Nobel Active™	65	665
			Nobel Replace	480	
			Nobel Parallel™	120	
Total	730	1574		3675	3675

Astra EV Astra Tech Implant System Osseospeed™ EV, *TX* Astra Tech Implant System Osseospeed™ TX, *Straumann S* tissue level standard, *SP* tissue level standard plus, *BL* bone level, *BLT* bone level tapered

Fig. 1 Diagram of the deep learning system. Deep learning model A was created for the detection of implant regions and classification of manufacturers on panoramic radiographs using YOLO-v7. After preparing the image patches that cropped the implant regions detected by model A, models B was created to identify the implant types per manufacturer or identify the diameters of Straumann implants using EfficientNet

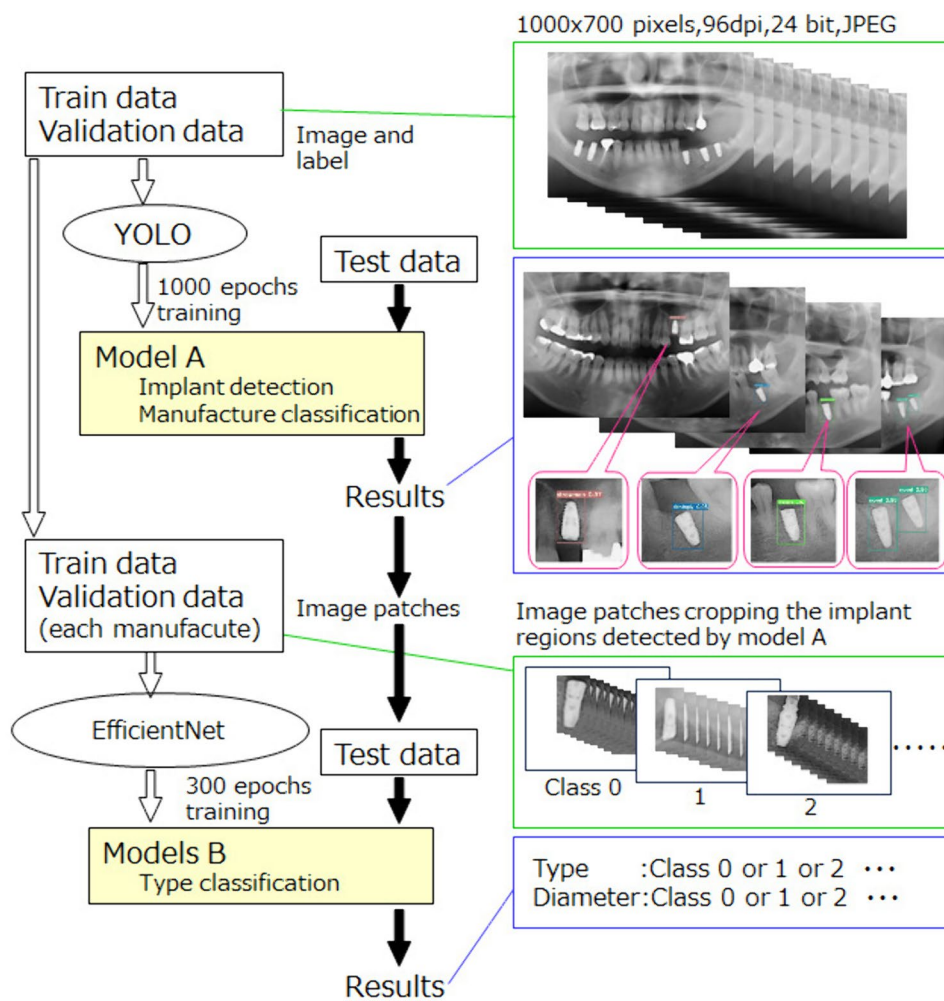


Table 2 Image assignment for model A

Class	Manufactures	No of images	Training data	Validation data	Test data
0	Kyocera	159	129	15	15
1	Dentsply Sirona	588	472	58	58
2	Straumann	528	422	53	53
3	Nobel biocare	299	239	30	30

Model A

Image assignment and annotation

The panoramic radiographs were randomly assigned as training, validation, or test data in the ratio 8:1:1. The number of images per manufacturer was adjusted to the aforementioned ratio. The images of patients assigned as test data were not assigned as training or validation data. The details are shown in Table 2.

Labels containing the classes and coordinates of the implant manufacturers were created on the training and validation images. Classification was determined as follows: Kyocera to class 0, Dentsply Sirona to class 1, Straumann to class 2, and Nobel Biocare to class 3. Rectangles surrounding the implant regions were set by one radiologist (YA) using labelImg software (<https://github.com/heartexlabs/labelImg/releases>), and the upper left and lower right coordinates were recorded.

Deep learning neural network architecture

The deep learning procedure was performed on a desktop computer equipped with a 24-GB GPU (NVIDIA GeForce RTX 3090, Santa Clara, CA, USA) and running Windows 10 Pro (Microsoft).

The YOLO-v7 network was downloaded from the GitHub repository (<https://github.com/WongKinYiu/yolov7>). YOLO-v7 consists of four general modules: input terminal, backbone, head, and prediction [22, 23]. Additionally, it has five basic components: community bridging services, max pooling, efficient layer aggregation network (ELAN), ELAN-higher, and spatial pyramid pooling-cross stage partial convolution. YOLO works by dividing images into grid cells and predicting bounding boxes for each cell. For each bounding box, it predicts the class probability and confidence score. The confidence score is the probability that the bounding box contains a particular object.

Training details

The YOLO-v7 architecture was used unchanged. A total of 1,000 learning epochs were performed using the training and

validation image and label datasets. The number of learning epochs was determined after confirming that the loss during learning converged sufficiently. The initial learning rate was set to 0.01, and was adopted, because learning with this setting yielded sufficient diagnostic accuracy. As the confidence score affects the balance between precision and recall, it was set to the score that showed a high F1 value after analysis across all bounding boxes, and was selected from 0.1, 0.3, and 0.5.

Evaluation metrics for model A

Test data were applied to the created model A to evaluate the model's performance.

To evaluate the detection performance for implants on panoramic radiographs, the following metrics were acquired:

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}).$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}).$$

$$\text{F1 score} = 2 \times \text{Recall} \times \text{Precision} / (\text{Recall} + \text{Precision}).$$

$$\text{False positive per images} = \text{FP} / \text{number of images},$$

where TP denoted true-positive, which is the number of cases in which the trained model correctly detected areas where implants actually exist; FN denoted false-negative, which is the number of cases in which the model failed to detect areas where implants actually exist; and FP denoted false-positive, which indicates is the number of cases in which the model incorrectly detected areas where implants actually did not exist.

For the evaluation of the identification (classification) of implant manufacturers, the accuracy, recall, precision, and F1 score were acquired. Among the trained models created using the three confidence scores, the model with the highest performance was adopted; that is, the model with the confidence score of 0.1.

To investigate the external validation, 19 panoramic radiographs containing 64 implants (16 Dentsply Sirona, 29 Straumann, and 19 Nobel Biocare) were collected from other institution. These were used as external test dataset to evaluate the performance of Model A to detect implants and identify the manufacturers.

Table 3 Image patch assignment for model B

Manufactures	Class	Implant types	No of image patches	Training data	Validation data	Test data
Dentsply	0	Astra EV	1427	1200	114	113
Sirona	1	Astra TX	223	177	22	24
Straumann	0	Straumann S/SP	20	13	4	3
	1	Straumann BL/BLT	1077	864	106	107
Nobel biocare	0	Nobel Active™	65	50	8	7
	1	Nobel Replace	480	395	42	43
	2	Nobel Parallel™	120	96	12	12

Astra EV Astra Tech Implant System Osseospeed™ EV, *TX* Astra Tech Implant System Osseospeed™ TX, *Straumann S* tissue level standard, *SP* tissue level standard plus, *BL* bone level, *BLT* bone level tapered

Models B

Image assignment and annotation

The implant regions detected by model A were cropped on the panoramic radiographs, and then, image patches were created. Image patches were randomly assigned as the training, validation, and test data in the ratio 8:1:1. Image patches for the same patient were not split between the test and training/validation databases. Details of the image patch assignment are shown in Table 3.

Classification by implant type was as follows: Dentsply Sirona's implants were classified as class 0 for the Astra Tech Implant System Osseospeed™ EV (Astra EV), and class 1 for the Astra TX. Straumann's implants were classified as class 0 for Straumann Tissue Level Standard (S)/Tissue Level Standard Plus (SP), and class 1 for Bone Level (BL)/Bone-Level Tapered (BLT). Nobel Biocare's implants were classified as class 0 for Active™, class 1 for Replace, and class 2 for Parallel™.

Deep learning neural network architecture

The same computer as mentioned above was used. EfficientNet B0 was used as neural networks for classification. EfficientNet is a model created by uniformly scaling the depth/width/resolution dimensions of a neural network, designing a new baseline network using neural architecture exploration, and then scaling it up [24, 25]. It achieves high performance and efficiency with fewer parameters. EfficientNet-b0 is a convolutional neural network trained on over 1 million images from the ImageNet database and can classify images into 1000 object categories.

Training details

EfficientNet was trained for 300 epochs using type-specific training and validation datasets for Dentsply, Straumann, and Nobel implants. Before training, the input images were

resized to 256 × 256 pixels and automatically augmented by resizing, horizontally and vertically flipping, rotating, and adjusting the brightness/contrast of images. The initial learning rate was set to 0.0001. Training was run three times on each dataset, which each yielded three trained model Bs.

Evaluation metrics for model B

To evaluate classification performance, each test dataset by type or diameter was applied to each of the three trained model Bs. The evaluation metrics were accuracy, recall, precision, and F1 score. The most accurate model among the three was applied. The metrics for that model are shown in the Results section.

Regarding the classification of the two types of Dentsply and Straumann implants, ROC curves were drawn and the areas under the curve (AUCs) were obtained.

Results

Performance of model A

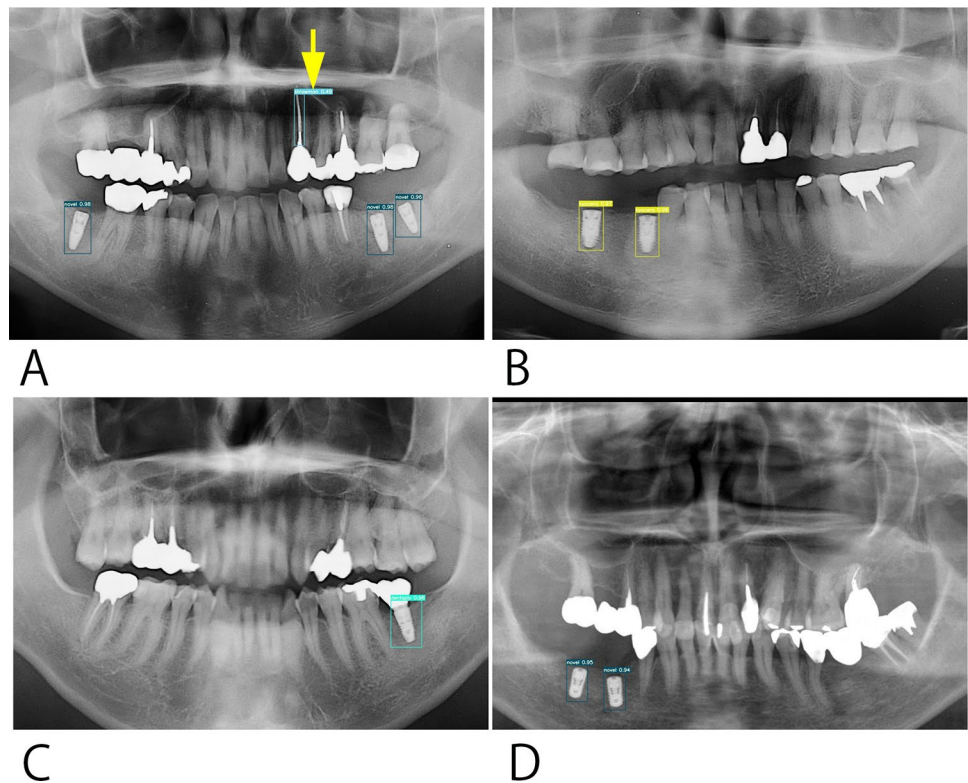
Performance for the detection of implants

Table 4 shows the performance of model A trained using YOLO to detect implants on panoramic radiographs. It achieved very high performance with a recall of 1.000, precision of 0.979, and F1 score of 0.989. The false positive per

Table 4 Performance of model A using YOLO for the detection of implants

Evaluation metrics	Values
Recall	1.000
Precision	0.979
F1 score	0.989
False positive per images	0.038

Fig. 2 Successful and unsuccessful examples for the detection of implant regions and classification of manufacturers by model A. **A** Post/root filling in the root canal of the left maxillary canine (arrow) was incorrectly detected as an implant (false positive). Class 3 Nobel implants in bilateral mandibular molars were correctly detected and correctly classified as class 3. **B** Class 0 Kyocera implants in the right mandibular molars were correctly detected and correctly classified as class 0. **C** Class 1 Dentsply implant in the left mandibular second molar was correctly detected and correctly classified as class 1. **D** Class 2 Straumann implants in the right mandibular molars were correctly detected but incorrectly classified as class 3 (Nobel)



		Model A's classification			
		Class 0 (Kyocera)	Class 1 (Dentsply)	Class 2 (Strauman)	Class 3 (Nobel)
Ground truth	Class 0 (Kyocera)	22	1	0	0
	Class 1 (Dentsply)	0	108	0	0
	Class 2 (Strauman)	0	0	98	2
	Class 3 (Nobel)	0	0	0	45

Fig. 3 Confusion matrix of model A using YOLO for the classification of implant manufacturers

images was very low at 0.038. There was no difference in detection performance depending on implant site. In external validation, recall, precision, and F1 score for implant detection were all 1.000.

Figure 2A shows an example of a false positive. Figure 2B–D shows examples of successful implant detection.

Table 5 Performance of model A using YOLO for the classification of implant manufacturers

Evaluation metrics	Total	Class 0 (Kyocera)	Class 1 (Dentsply)	Class 2 (Straumann)	Class 3 (Nobel)
Accuracy	0.989				
Recall	0.984	0.957	1.000	0.980	1.000
Precision	0.987	1.000	0.991	1.000	0.957
F1 score	0.985	0.978	0.995	0.990	0.978

Performance of the classification of implant manufacturers

The confusion matrix for the classification of implant manufacturers by model A trained using YOLO is shown in Fig. 3. All implants of class 1 Dentsply and class 3 Nobel were correctly classified into their respective classes. Only one of the class 0 Kyocera implants was incorrectly classified as class 1 (Dentsply), and two of the class 2 Straumann implants were incorrectly classified as class 3 (Nobel).

The classification performance of model A trained using YOLO is shown in Table 5. Model A achieved an accuracy, recall, precision, and F1 score of 0.98 or higher. The recall values for Dentsply and Nobel implants were 1.000. Other metrics exceeded 0.95. The performance of model A on the external dataset was 0.969 for precision, 0.978 for recall, 0.965 for precision, and 0.970 for F1 score in distinguishing implant manufacturers.

Dentsply's implant type		Model B's classification	
		Class 0 (EV)	Class 1 (TX)
Ground truth	Class 0 (EV)	112	1
	Class 1 (TX)	2	23

Straumann's implant type		Model B's classification	
		Class 0 (S/SP)	Class 1 (BL/BLT)
Ground truth	Class 0 (S/SP)	3	0
	Class 1 (BL/BLT)	0	107

Nobel's implant type		Model B's classification		
		Class 0 (Active)	Class 1 (Replace)	Class 2 (Parallel)
Ground truth	Class 0 (Active)	7	0	0
	Class 1 (Replace)	0	43	0
	Class 2 (Parallel)	0	3	9

Fig. 4 Confusion matrix of model B using EfficientNet for classification of implant types and diameters

Examples of successful and unsuccessful classification are shown in Fig. 2. Figure 2A–C shows examples of the successful classification of class 3 (Nobel), class 0 (Kyocera), and class 1 (Dentsply) implants, respectively.

Figure 2D shows an example where a class 2 Straumann implant was incorrectly classified as class 3 (Nobel).

Performance of model B

The confusion matrix and classification performance of model B using EfficientNet are shown in Fig. 4 and Table 6. Regarding the classification of implant types of Dentsply, model B achieved high metrics above 0.92. Figure 5A shows a successful example where a class 0 Astra EV implant was correctly classified as class 0. Figure 5B shows an unsuccessful example where a class 1 Astra TX implant was incorrectly classified as class 0 (EV).

Regarding the Nobel Biocare's implant types, models B showed high values above 0.91 for all metrics, except the class 2 (Parallel) values. Figure 5E shows a successful example where a class 0 Nobel Active implant was correctly classified as class 1. Figure 5F shows a successful example where a class 1 Nobel Replace implant was correctly classified as class 1. Figure 5G shows an unsuccessful example where a class 2 Parallel implant was incorrectly classified as class 1 (Replace).

Table 6 Performance of model B using EfficientNet for the classification of implant types and diameters

	Dentsply Sirona's implant types			
	Total	Class 0 (EV)	Class 1 (TX)	
Accuracy	0.978			
Recall	0.956	0.991	0.920	
Precision	0.978	0.983	0.958	
F1 score	0.963	0.987	0.939	
AUC	0.996			
	Straumann's implant types			
	Total	Class 0 (EV)	Class 1 (TX)	
Accuracy	1.000			
Recall	1.000	1.000	1.000	
Precision	1.000	1.000	1.000	
F1 score	1.000	1.000	1.000	
AUC	1.000			
	Nobel Biocare's implant types			
	Total	Class 0 (3.3 mm)	Class 1 (4.1 mm)	Class 2 (4.8 mm)
Accuracy	0.952			
Recall	0.917	1.000	1.000	0.750
Precision	0.978	1.000	0.935	1.000
F1 score	0.941	1.000	0.967	0.857

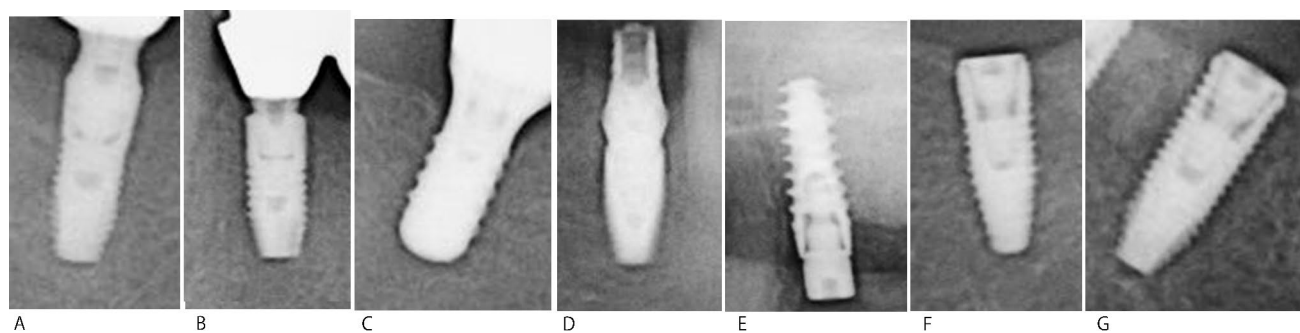


Fig. 5 Examples of the classification of implant types by models B using EfficientNet. **A** Class 0 Dentsply Astra EV implant was correctly classified as class 0. **B** Class 1 Dentsply Astra TX implant was incorrectly classified as class 0 (EV). **C** Class 0 Straumann S/SP implant was correctly classified as class 0. **D** Class 1 Strau-

mann BL/BLT implant was correctly classified as class 1. **E** Class 0 Nobel Active implant correctly classified as class 0. **F** Class 1 Nobel Replace implant was correctly classified as class 1. **G** Class 2 Nobel Parallel implant was incorrectly classified as class 1 (Replace) classified as class 1 (Replace) by model C

Discussion

Implant shapes are complex, and thousands of implant types exist. The experienced dentists achieved reliable accuracy similar to deep learning methods in identifying several types of implants [12]; however, even experienced dentists have difficulty in accurately identifying thousands of implants. The deep learning could solve this problem [3].

The ultimate goal of deep learning studies to identify implant types is to create optimal models for the classification of implants used around the world on images, so that they can be used in actual clinical settings [26]. When using classification models, the inference process requires inputting manually cropped images of implant regions into the trained model, instead of inputting the panoramic radiographs themselves. This point is disadvantageous for practical use. This study was designed to create deep learning models in two steps. The first step was to create a model to automatically detect implants and identify their manufacturers on panoramic radiographs. After preparing image patches created by cropping the implant regions detected in the first model, the second step was to create models that categorized implant fixture types for each manufacturer.

Previous studies on implant type identification have included periapical radiographs and dental cone beam CT; however, in most studies, panoramic radiographs were used [16]. Panoramic radiographic data have the advantage of being standardized in size, resolution, and projection angle [3]. Panoramic radiography provides images of the cervical vertebrae and maxillary sinuses that superimposed on the jawbone, which results in distortions depending on the patient's posture, and this distortion can reduce the image quality of the implant fixture [17].

However, overall, panoramic radiographs are considered to be excellent for the specification of the implant location, classification of implant fixtures, classification of implant diameters, and evaluation of bone around implants [4]. For these reasons, we decided to use panoramic radiographs in this study.

YOLO-v7 is a network used for object detection [23, 24]. It divides images into grids, and predicts the object's existence probability and bounding box for each cell. This means that detection and classification are possible simultaneously. Each cell has multiple anchor boxes used to detect objects of different shapes, and many objects can be detected simultaneously. Prior to training with the object detection network, it is necessary to create labels that contain the class names and the coordinates of the implant regions. In this study, we solved the complexity of annotation process using labellmg software.

The object detection networks achieved high detection performance; whereas they have the disadvantage that classification accuracy decreases as the number of classes increases [16]. Due to the complexity of implant structures, it can be inherently difficult to accurately classify implant types using object detection networks [16]. Takahashi et al. attempted to detect dental implants from three manufacturers on 1,282 panoramic radiographs using YOLO-v3, and reported that the mean average precision for detection was 0.71–0.72 [17]. In contrast, this study achieved high performance with recall of 1.000 and precision of 0.979 for the detection of implants from four manufacturers on 1,574 panoramic radiographs using YOLO-v7. This result can be attributed to the technological advances in YOLO and the increase in the number of panoramic radiographs. Furthermore, this study achieved high classification performance with an accuracy of 0.989 for identifying the four manufacturers. We attempted to reduce the number of classes by

classifying by manufacturer instead of by implant type. As a result, the object detection network may have achieved high classification performance. The developers of YOLO v7 compared detection performance and processing speeds with various networks including past YOLO series. The processing of YOLO v7 was so fast that it enabled real-time detection on video, reducing network costs while maintaining high diagnostic accuracy [23].

When using classification networks, it is necessary to create image patches. This is because the model has better classification performance when learning using image patches that crop the regions of interest than when learning using large images themselves. During the inference process, the clinicians should prepare image patches by manually cropped the implant regions from images when inputting test data into the model. This is a time-consuming task. The ROIs in the images include not only the implant fixtures, but also many other complex conditions (surrounding alveolar bone, covers, screws, and prosthesis) [26]. Therefore, some incorrect predictions may occur [27]. In this study, image patches could be created almost automatically by cropping the implant regions detected by Model A.

Classification performance varies greatly depending on the neural network used, amount of training data, and number of classes [3, 4, 26]. EfficientNet, a state-of-the-art classification network, was used in this study. As deep learning layers become deeper, the learning time increases and deterioration occurs, accuracy saturates, and performance decreases [4]. EfficientNet optimizes the network for images by setting composite coefficients and adjusting each of the three elements of depth, width and input image resolution by a specific multiple [25]. It is significantly lighter and more accurate than previous networks. It also achieves high accuracy for transfer learning.

As the number of training images increases, the accuracy of the model increases [3, 26]. This could be improved by collecting a more uniform and sufficient number of images for each type of implant [3]. Data augmentation techniques are often used to increase the number of images for the training model [27]. EfficientNet performs data augmentation by resizing, horizontally and vertically flipping, rotating, and

adjusting the brightness/contrast of images prior to learning. Even when the number of classes increases, data augmentation may be possible to secure the number of images per class and obtain good results [3].

As the number of implant types increases, diagnostic accuracy decreases. In many studies, researchers have investigated the identification of 3–13 types of implants [4, 10–13]. Researchers analyzed classification performance in the previous studies under various conditions; hence, it is difficult to make a simple comparison. However, most of these studies reported high accuracy over 0.9 was reported. Table 7 shows the summary of the previous reports. Prior to this study, we examined the same data using ResNet18, and showed that the results were comparable to Efficient's diagnostic accuracy (data were not shown). For example, Tiryaki et al. [11] used VGG on 11,904 images of five different types and reported a high accuracy of 0.983. Sukegawa et al. [4] used ResNet on 10,191 images of 13 types and reported a high accuracy of over 0.949.

On the other hand, Park et al. [9] identified 25 types of implants using ResNet, but reported a low accuracy of 0.825. Kong et al. [3] classified 130 types of implants using ensemble model of ResNet and EfficientNet. The ensemble model combines multiple networks to improve performance [28]. Using an ensemble model, Kong et al. [3] showed a top-1 accuracy of 0.75 and top-5 accuracy of 0.95, which achieved higher accuracy than a single model [3]. Creating ensemble models requires engineers who specialize in programming, and this will need to be considered in the future. Even implants of the same type may not be compatible depending on the diameter or length [16]; hence, it may be necessary to create models for every implant, which is difficult. Rather than achieving high top-5 accuracy, it may be more convenient to create two-step models that identifies manufacturers, and then classifies types by manufacturer.

Additionally, models used in clinical practice must be trained using datasets of implant fixtures with various shapes and angles [26]. The report stated that the model trained using datasets of implants with various shapes and angles achieved an accuracy of over 80% [18]. A study using activation mapping technology demonstrated that the AI focuses

Table 7 Reports of accuracy of oral implants classification

Author	Ref No	Numbers of implants	Number of types	Networks	Accuracy
Tiryaki B	11	11,904	5	VGG	0.983
Sukegawa S	10	8859	11	VGG	0.935
Sukegawa S	4	10,191	13	ResNet	0.949
Park W	26	13,291	25	ResNet	0.825
Kong HJ	3	28,112	130	ResNet and EfficientNet	0.75(top1) 0.95(top5)

Ref No Reference No

on the implant fixture itself for clarification, and that angular and shape factors do not significantly impact type classification accuracy [26].

This study has several limitations regarding improving the clinical applicability of deep learning for implant identification. First, the panoramic images used in this study were collected from limited institutions. Model A was validated using external dataset from other institution and demonstrated sufficient performance for implant detection and manufacturer identification. When many types of images are collected from multiple institutions, there may be a lack of standardization of images, for example, differences in the types of imaging device, image quality, and imaging setting conditions, which may reduce the performance of the classification model [3, 16]. Accuracy can be improved using a network that is suitable for implant identification or by fine-tuning the model using extensive datasets [16]. Further verification will be required.

Next, collecting a supervised dataset requires a large number of resources, including funds, time, skilled personnel, hardware, and software. Further research is required to determine whether the number of datasets can be reduced by adopting a network specialized for implant classification [3, 26].

Third, the creation of compatible models has not kept pace with the development of implants. In this study, models were created based on several types of implant images used in our hospital. However, many more implants are now available around the world, and new types of implants are continuously being developed. It will be necessary to input data on new products and continue learning. Therefore, instead of building a learning model once, it may be necessary to develop a federated deep learning system that continuously reflects data uploaded from various facilities.

The fourth issue will require more detailed predictions clinically. This includes unresolved issues, such as predicting implant diameters, distinguishing between similar implant types, and prediction the use of angle-changing screws. The model developed in our pilot study was able to predict the three diameters of Straumann implants; 3.3, 4.1, and 4.8 mm, with an accuracy of 0.95. As panoramic radiographs have many issues regarding the magnification rate and visibility depending on the scanned areas, this result was not included in this study. Further research is needed.

Conclusion

In this study, two-step deep learning models were built to identify implant types on panoramic radiographs. The object detection model detects the implant regions with high performance and identified the four manufacturers of implants with high accuracy. The classification models identified

implant types per manufacturer with high performance. In the future, further modifications will be made to the models by increasing the number of implant types, increasing the number of images for each type, and collecting images from multiple institutions.

Acknowledgements This work was supported by JSPS KAKENHI [Grant No. 20K10194]. The authors thank Edanz (<https://jp.edanz.com/ac>) for editing a draft of this manuscript.

Author contributions Yoshiko Arij: conceptualization, funding acquisition, methodology—software, investigation, data curation, formal analysis, and writing—original draft. Kaori Kusano: conceptualization, investigation, data curation, and writing—review and editing. Motoki Fukuda: methodology—software, data curation, formal analysis, and writing—review and editing. Yo Wakata, Michihito Nozawa, and Shinya Kotaki: data collection. Shunsuke Baba: supervision.

Data availability The datasets used in this study will be available to researchers with legitimate purposes. Please contact the corresponding author to obtain authorization.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval All procedures performed on human participants were in accordance with the ethical standards of our institutional and/or national research committee, and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The ethics committees of Osaka Dental University (Approval No. 111195–0) granted permission for this study. Ethics approval was also obtained from the co-authors' institution for the use of external testing dataset (Approval No. 586).

Informed consent For this type of study (retrospective study), formal consent is not required. Patients were given the opportunity to decline participation via an opt-out format. Declaration regarding generative AI and AI-assisted technologies in the creating process: The generative AI (ChatGPT) was not used in the manuscript writing process.

References

1. Steigenga JT, Al-Shammari KF, Nociti FH, Misch CE, Wang HL. Dental implant design and its relation-ship to long-term implant success. *Implant Dent.* 2003;12:306–17. <https://doi.org/10.1097/01.id.0000091140.76130.a1>.
2. Esposito M, Grusovin MG, Maghaires H, Worthington HV. Interventions for replacing missing teeth: different times for loading dental implants. *Cochrane Database Syst Rev.* 2013. <https://doi.org/10.1002/14651858.CD003878.pub5>.
3. Kong HJ, Eom SH, Yoo JY, Lee JH. Identification of 130 dental implant types using ensemble deep learning. *Int J Oral Maxillofac Implants.* 2023;38:150–6. <https://doi.org/10.11607/jomi.9818>.
4. Sukegawa S, Yoshii K, Hara T, Tanaka F, Yamashita K, Kagaya T, Nakano K, Takabatake K, Kawai H, Nagatsuka H, Furuki Y. Is attention branch network effective in classifying dental implants from panoramic radiograph images by deep learning? *PLoS ONE.* 2022;17:e0269016. <https://doi.org/10.1371/journal.pone.0269016>.

5. Buser D, Weber HP, Brägger U, Balsiger C. Tissue integration of one-stage implants: three-year results of a prospective longitudinal study with hollow cylinder and hollow screw implants. *Quintessence Int.* 1994;25:679–86.
6. Adell R, Eriksson B, Lekholm U, Brånemark PI, Jemt T. Long-term follow-up study of osseointegrated implants in the treatment of to-tally edentulous jaws. *Int J Oral Maxillofac Implants.* 1990;5:347–59.
7. Al-Johany SS, Al Amri MD, Alsaed S, Alalola B. Dental implant len and diameter: a proposed classification scheme. *J Prosthodont.* 2017;26:252–60. <https://doi.org/10.1111/jopr.12517>.
8. Jokstad A, Ganeles J. Systematic review of clinical and patient-reported outcomes following oral rehabilitation on dental implants with a tapered compared to a non-tapered implant design. *Clin Oral Implants Res.* 2018;29:41–54. <https://doi.org/10.1111/clr.13128>.
9. Park W, Schwendicke F, Krois J, Huh JK, Lee JH. Identification of dental implant systems using a large-scale multicenter data set. *J Dent Res.* 2023;102:727–33. <https://doi.org/10.1177/00220345231160750>.
10. Sukegawa S, Yoshii K, Hara T, Yamashita K, Nakano K, Yamamoto N, Nagatsuka H, Furuki Y. Deep neural networks for dental implant system classification. *Biomolecules.* 2020;10:984. <https://doi.org/10.3390/biom10070984>.
11. Tiryaki B, Ozdogan A, Guller MT, Miloglu O, Oral EA, Ozbek IY. Dental implant brand and angle identification using deep neural networks. *J Prosthet Dent.* 2023;S0022–3913:00492–4. <https://doi.org/10.1016/j.prosdent.2023.07.022>.
12. Lee JH, Jeong SN. Efficacy of deep convolutional neural network algorithm for the identification and classification of dental implant systems, using panoramic and periapical radiographs: a pilot study. *Medicine.* 2020;99:e20787. <https://doi.org/10.1097/MD.00000000000020787>.
13. Kim JE, Nam NE, Shim JS, Jung YH, Cho BH, Hwang JJ. Transfer learning via deep neural networks for implant fixture system classification using periapical radiographs. *J Clin Med.* 2020;9:1117. <https://doi.org/10.3390/jcm9041117>.
14. Sukegawa S, Saika M, Tamamura R, Nakano K, Takabatake K, Kawai H, Nagatsuka H, Furuki Y. Clinical retrospective study of dental implant removal: do patients who require implant removal desire implant prosthesis again? *Med Oral Patol Oral Cir Bucal.* 2020;25:e784–90. <https://doi.org/10.4317/medoral.23789>.
15. Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, Liu Y, Topol E, Dean J, Socher R. Deep learning-enabled medical computer vision. *NPJ Digit Med.* 2021;4:5. <https://doi.org/10.1038/s41746-020-00376-2>.
16. Kong HJ, Yoo JH, Lee JH, Eom SH, Kim JH. Performance evaluation of deep learning models for the classification and identification of dental implants. *J Prosthet Dent.* 2023;S0022–3913:00467–75. <https://doi.org/10.1016/j.prosdent.2023.07.009>.
17. Takahashi T, Nozaki K, Gonda T, Mamenno T, Wada M, Ikebe K. Identification of dental implants using deep learning—pilot study. *Int J Implant Dent.* 2020;6:53. <https://doi.org/10.1186/s40729-020-00250-6>.
18. Lee JH, Kim YT, Lee JB, Jeong SN. A performance comparison between automated deep learning and dental professionals in classification of dental implant systems from dental imaging: a multi-center study. *Diagnostics.* 2020;10:910. <https://doi.org/10.3390/diagnostics10110910>.
19. Oshida Y, Tuna EB, Aktören O, Gençay K. Dental implant systems. *Int J Mol Sci.* 2010;11:1580–678. <https://doi.org/10.3390/ijms11041580>.
20. Basha SS, Dubey SR, Pulabaigari V, Mukherjee S. Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing.* 2020;378:112–9. <https://doi.org/10.1016/j.neucom.2019.10.008>.
21. Jiang P, Ergu D, Liu F, Cai Y, Ma B. A review of Yolo algorithm developments. *Procedia Comput Sci.* 2022;199:1066–73. <https://doi.org/10.1016/j.procs.2022.01.135>.
22. Diwan T, Anirudh G, Tembhurne JV. Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimed Tools Appl.* 2023;82:9243–75. <https://doi.org/10.1007/s11042-022-13644-y>.
23. Soeb MJA, Jubayer MF, Tarin TA, Al Mamun MR, Ruhad FM, Parven A, Mubarak NM, Karri SL, Meftaul IM. Tea leaf disease detection and identification based on YOLOv7 (YOLO-T). *Sci Rep.* 2023;13:6078. <https://doi.org/10.1038/s41598-023-33270-4>.
24. Afshar MF, Shirmohammadi Z, Ghahramani SAAG, Noorparvar A, Hemmatyar AMA. An efficient approach to monocular depth estimation for autonomous vehicle perception systems sustainability. *MDPI.* 2023;15(11):8897. <https://www.mdpi.com/2071-1050/15/11/8897>.
25. Tan M, Le QV (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv:1905.11946*. <https://doi.org/10.48550/arXiv.1905.11946>. Accessed 31 May 2023
26. Park W, Huh JK, Lee JH. Automated deep learning for classification of dental implant radiographs using a large multi-center dataset. *Sci Rep.* 2023;13:4862. <https://doi.org/10.1038/s41598-023-32118-1>.
27. Sakai T, Li H, Shimada T, Kita S, Iida M, Lee C, Nakano T, Yamaguchi S, Imazato S. Development of artificial intelligence model for supporting implant drilling protocol decision making. *J Prosthodont Res.* 2023;67:360–5. https://doi.org/10.2186/jpr.JPR_D_22_00053.
28. Opitz D, Maclin R. Popular ensemble methods: an empirical study. *J Artif Intell Res.* 1999;11:169–98. <https://doi.org/10.1145/3631136>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.