Transfer learning in a deep convolutional neural network for implant fixture classification: A pilot study

Hak-Sun Kim (ID)[1], Eun-Gyu Ha (ID)[1], Young Hyun Kim (ID)[1], Kug Jin Jeon (ID)[1], Chena Lee (ID)[1], Sang-Sun Han (ID)[1,2,*]

[1]*Department of Oral and Maxillofacial Radiology, Yonsei University College of Dentistry, Seoul, Korea*
[2]*Center for Clinical Imaging Data Science, Yonsei University College of Medicine, Seoul, Korea*

ABSTRACT

Purpose: This study aimed to evaluate the performance of transfer learning in a deep convolutional neural network for classifying implant fixtures.

Materials and Methods: Periapical radiographs of implant fixtures obtained using the Superline (Dentium Co. Ltd., Seoul, Korea), TS III (Osstem Implant Co. Ltd., Seoul, Korea), and Bone Level Implant (Institut Straumann AG, Basel, Switzerland) systems were selected from patients who underwent dental implant treatment. All 355 implant fixtures comprised the total dataset and were annotated with the name of the system. The total dataset was split into a training dataset and a test dataset at a ratio of 8 to 2, respectively. YOLOv3 (You Only Look Once version 3, available at https://pjreddie.com/darknet/yolo/), a deep convolutional neural network that has been pretrained with a large image dataset of objects, was used to train the model to classify fixtures in periapical images, in a process called transfer learning. This network was trained with the training dataset for 100, 200, and 300 epochs. Using the test dataset, the performance of the network was evaluated in terms of sensitivity, specificity, and accuracy.

Results: When YOLOv3 was trained for 200 epochs, the sensitivity, specificity, accuracy, and confidence score were the highest for all systems, with overall results of 94.4%, 97.9%, 96.7%, and 0.75, respectively. The network showed the best performance in classifying Bone Level Implant fixtures, with 100.0% sensitivity, specificity, and accuracy.

Conclusion: Through transfer learning, high performance could be achieved with YOLOv3, even using a small amount of data. *(Imaging Sci Dent 2022; 52: 219-24)*

KEY WORDS: Artificial Intelligence; Deep Learning; Dental Implants; Dental Radiography

Introduction

Dental implant treatment is a way of restoring the function of a permanent tooth after its loss.[1] As implant treatment has grown more popular,[2] more manufacturers have produced various implant systems.[3-6] This diversity provides dentists with more options among implant systems. However, since each manufacturer produces implant systems with unique implant parts,[3,5,6] if patient records containing the exact name of the implant system are not available, it can be quite challenging to identify the implant system.

Periapical radiographs or panoramic radiographs are often taken to help identify the implant system.[7] However, without extensive experience in implant treatment using various systems, it is almost impossible to differentiate among systems based on imaging results,[8] and sometimes dentists end up trying to apply every kind of implant part that would fit. To resolve this laborious situation, websites[4] and programs[6] have been developed to facilitate identification based on several characteristics of implant fixtures.

More recently, several studies have utilized deep convolutional neural networks (DCNNs) to detect or classify various lesions or objects in dental radiographs.[8-15] DCNNs alone can be trained to show excellent performance, but transfer learning to DCNNs is considered to be more efficient.[8,10,11] The concept of "transfer learning" refers to training neural networks that have already been trained before,

Fig. 1. Examples of annotating implant fixtures using LabelImg software. A. Superline fixtures are annotated with light green boxes. B. TS III fixtures are annotated with purple boxes. C. Bone Level Implant fixtures are annotated with red boxes.

with related, but new data.[16,17] For example, a DCNN can be pretrained with image data to develop a final network for the detection or classification of certain objects in a new image dataset, and pretraining can be applied to language data for a new language dataset in the same way. The Microsoft COCO (Common Objects in Context) image dataset is commonly used to pretrain DCNNs, and YOLOv3 (You Only Look Once version 3) is pretrained with the COCO dataset. Within the broader category of plain radiographs, periapical radiographs have better resolution than panoramic radiographs and are recommended for the assessment of implants.[7] Therefore, this study aimed to evaluate the performance of an implant fixture classification model through transfer learning to YOLOv3 with periapical radiographs.

Materials and Methods

Data preparation

Periapical radiographs of adult patients who had undergone dental implant treatment at Yonsei University Dental Hospital were collected between April 2020 to July 2021 in Digital Imaging and Communication in Medicine (DICOM) format. The images were anonymized to prevent identification, and the requirement for patient consent was waived by the Institutional Review Board (IRB) of Yonsei University Dental Hospital (IRB No. 2-2021-0104) for the retrospective collection of images. The DICOM files were con-

**Table 1.** The number of fixtures for each implant system in the dataset

Implant system	Superline	TS III	Bone Level Implant	Total
Train dataset	95	92	97	284
Test dataset	26	23	22	71
Total dataset	121	115	119	355

verted to bitmap files via ImageJ software for training and testing.

The implant systems in the radiographs were noted, and 3 systems - Superline (Dentium Co. Ltd., Seoul, Korea), TS III (Osstem Implant Co. Ltd., Seoul, Korea) and Bone Level Implant (Institut Straumann AG, Basel, Switzerland) - were selected.

Every implant fixture in all the bitmap images was annotated in LabelImg software (ver. 1.8.4, available at https://github.com/tzutalin/labelImg). Rectangular boxes were drawn around only the fixtures and labeled with their class name (i.e., the name of the implant system) (Fig. 1): Superline, TS III, or Bone Level Implant. The information of these annotation boxes was saved as eXtensible Markup Language (XML) files, separate from the bitmap files.

In total, 263 periapical radiographs with 355 implant fixtures constituted the dataset. The images of fixtures were allocated to the training dataset (284 fixtures in 212 images,
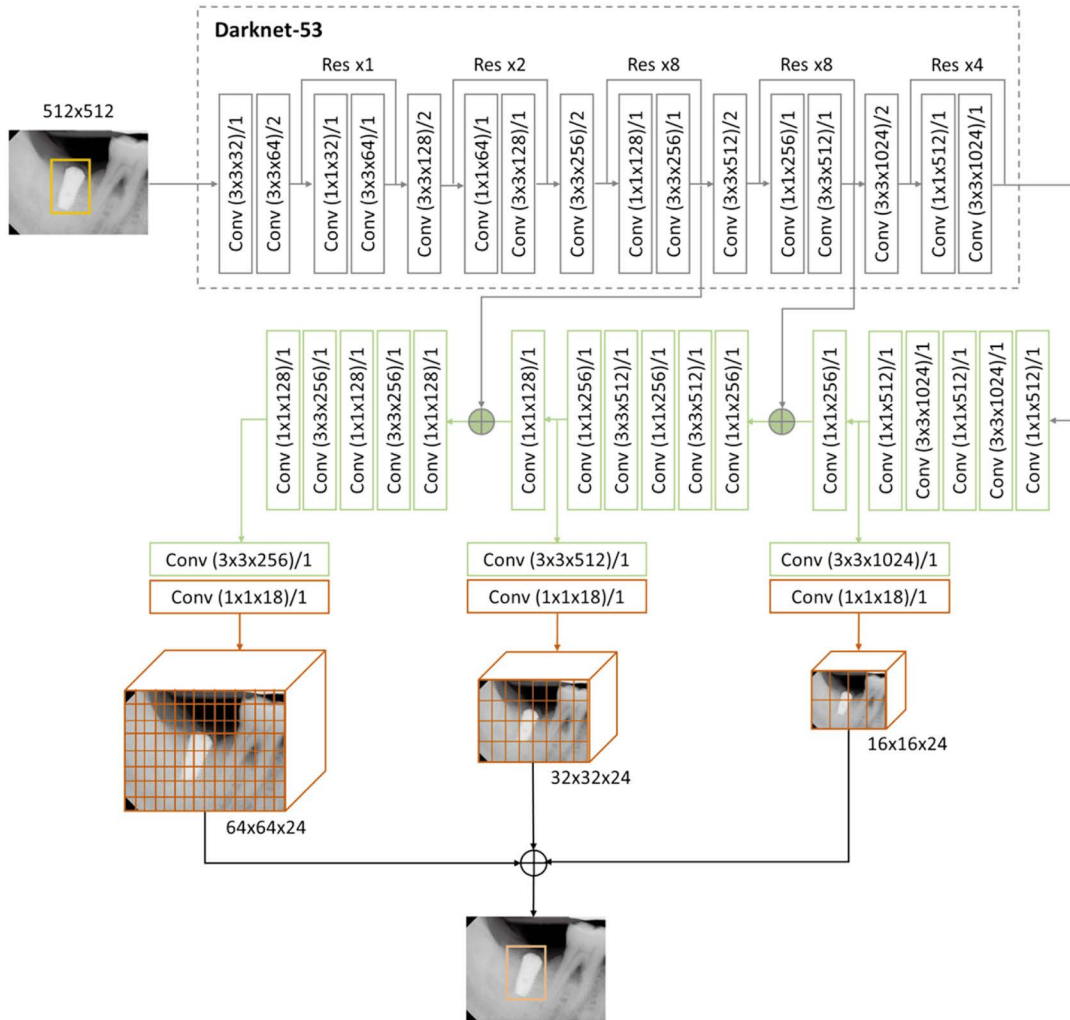
Fig. 2. The structure of YOLOv3 used in this study. In training, the input data to Darknet-53, the backbone of this network, are resized images to 512 (width) × 512 (height) pixels and their annotation files. Then, 3 feature maps in different resolutions are produced with classification during the training via convolution layers. The final weights acquired from this procedure are used in testing.

80%) and test dataset (71 fixtures in 51 images, 20%). The detailed composition of the datasets is shown in Table 1.

### Transfer learning of a deep convolutional neural network

A DCNN (YOLOv3) was used for training and testing. The Keras library 2.2.4 (available at https://github.com/keras-team/keras/releases/tag/2.2.4) was used to implement the network, and TensorFlow 1.14 (available at https://github.com/tensorflow/tensorflow) was used as the backend on a desktop with TITAN RTX Graphics Processing Unit (NVIDIA Corp., Santa Clara, CA, USA).

First, from the XML files of the training dataset, the upper left (X_1, Y_1) and the lower right (X_2, Y_2) corners' coordinates of the annotations and their class names were extracted and saved in a comma-separated values (CSV) file. The

training dataset was then divided into a 9 : 1 ratio, with 10% of the data allocated to the validation process regardless of the class names. The training dataset images were resized to 512 (width) × 512 (height) pixels and used as input to the Darknet-53 backbone of YOLOv3 with the CSV file.

Darknet-53 has 53 convolutional layers with batch normalization applied. Its activation function was the leaky rectified linear unit. The output from this model is 3 feature maps, where implant fixtures are automatically classified at different resolutions (Fig. 2).

In training, weights produced from pretraining YOLOv3 with COCO dataset were used. This network was trained for 100, 200, and 300 epochs. For the first half of the epochs, only the last 3 layers' weights were trained with a batch size of 2. The other half of the epochs used the whole network's weights for training with a batch size of 4. This process of
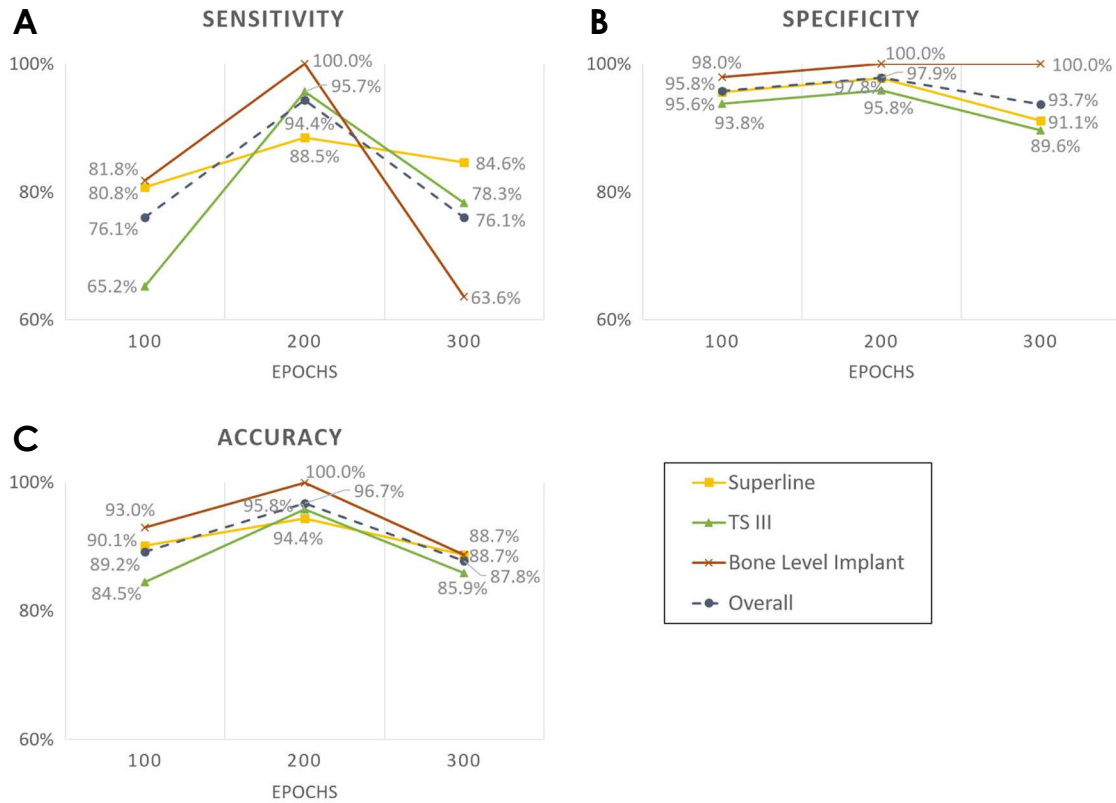
Fig. 3. The performance of implant fixture classification according to the number of epochs of training for YOLOv3. A. Sensitivity. B. Specificity. C. Accuracy.
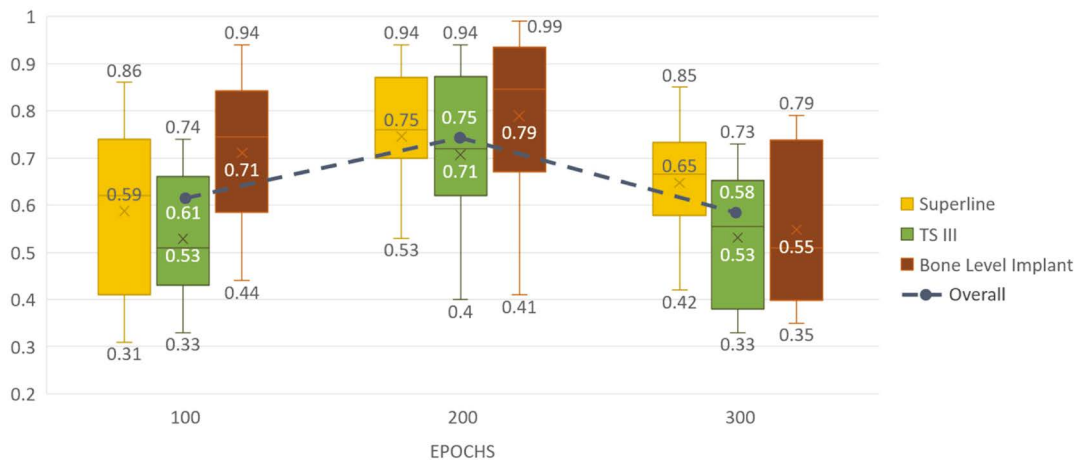

Fig. 4. Confidence score range (top and bottom) and mean values (middle) of correct classifications during testing according to the number of epochs of training for YOLOv3.

training an already pretrained network with a new dataset is called transfer learning.

The final weights were produced after training and utilized in testing. After testing, the sensitivity, specificity, accuracy, and confidence score of correct classifications were assessed according to the number of epochs.

Results

The sensitivity, specificity, and accuracy of implant fixture classification according to the number of epochs are shown in Figure 3. The highest overall accuracy was 96.7% at 200 epochs, and the accuracy values for Superline, TS
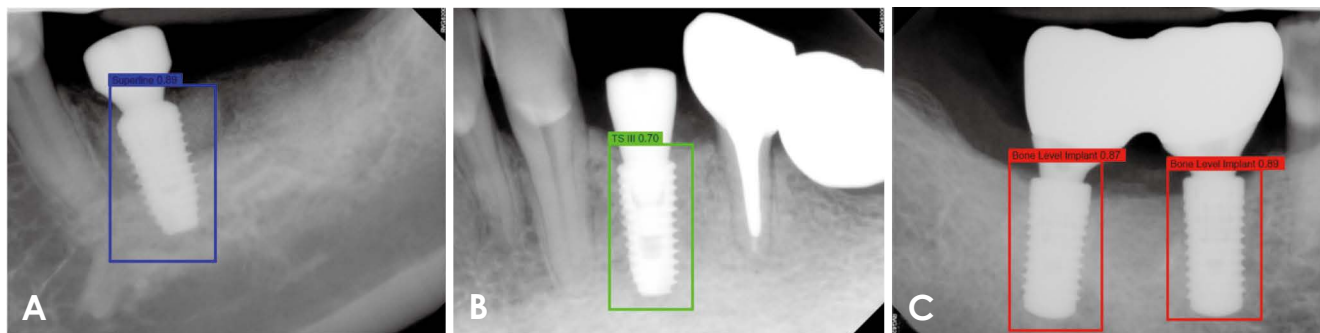
Fig. 5. Test result examples of correctly classified implant fixtures with the confidence score for each decision. A. Superline. B. TS III. C. Bone Level Implant.

III, and Bone Level Implant classification were 94.4%, 95.8%, and 100.0%, respectively. Overall sensitivity and specificity were also the highest at 200 epochs, (94.4% and 97.9%, respectively); the sensitivity for Superline, TS III, and Bone Level Implant classification was 88.5%, 95.7%, and 100.0% and the specificity was 97.8%, 95.8%, and 100.0%, respectively.

The mean values and the range of confidence scores for correct classifications according to the number of epochs are shown in Figure 4. They were also the highest at 200 epochs for all classes, with an overall mean score of 0.75. Specifically, the mean confidence score for correct classification was 0.75 for the Superline implant system, 0.71 for TSIII, and 0.79 for the Bone Level Implant. Figure 5 shows examples of correctly classified implant fixtures for each system with confidence scores. In these examples, not only did the network identify the implant systems, but the confidence scores of the decision were also above the preset threshold.

Discussion

Through transfer learning of YOLOv3, the network was trained for 100, 200, or 300 epochs. The performance metrics of the network were the highest when it was trained for 200 epochs. Bone Level Implant fixtures were the best-classified system with 100.0% sensitivity, specificity, and accuracy. Simultaneously, the confidence scores were the highest at 200 epochs. These results demonstrate that after training for a certain number of epochs, performance does not necessarily show an incremental change.

The confidence score indicates the certainty of classification during testing, and its threshold can be customized within the interval between 0 and 1. If a network can classify an implant fixture - that is, if the confidence score of the classification is above the threshold - it draws a box around the recognized implant fixture in the periapical image. In contrast, if the network cannot classify the implant fixture, or if the confidence score of the classification is below the threshold, it returns the periapical image without any box. Thus, for a given network, a higher threshold will yield fewer classification outcomes, while a lower threshold will lead to more classification outcomes being drawn.

There has been a dramatic increase in the number of DCNN studies in recent years.[12] Various networks have been applied for the detection or classification of lesions or anatomical structures.[8-11,13-15] Several studies have also evaluated the performance of DCNNs in implant fixture classification in a similar way. They mostly used networks other than YOLOv3,[9] such as basic CNN,[10] Visual Geometry Group,[10] and GoogLeNet Inception v3.[8,11] Takahashi et al.[9] assessed the performance of implant fixture classification of 4 implant systems based on panoramic radiographs using the same network as in this study. Their dataset consisted of more than 200 implants per system (as many as 1919 implants), and they trained the network for 1000 epochs. However, the highest sensitivity was only 0.82, which is lower than the lowest sensitivity observed in this study with 200 epochs of training. Other implant classification studies also trained networks for much more than 200 epochs, such as 700[10] or 1000 epochs.[8,9,11] Training for more epochs might improve performance to a certain extent, but it also consumes more time and leads to a higher probability of overfitting.[17]

Only periapical radiographs constituted the dataset of this study. Panoramic radiographs, which were used in other implant classification studies,[8-11] are acceptable as a modality for implant evaluation.[7] Nevertheless, panoramic radiographs have substantial drawbacks, such as blurring due to patient motion, superimposition of the normal anatomy, and

geometric distortions.[18] Periapical radiography offers a limited field of view, but with a paralleling technique; thus, it has superb resolution and the least distortion, which makes it the first-choice modality after the installation of implant fixtures.[7,19]

Most other studies had datasets of thousands of images, much more than this study.[8-11] The performance results of this study, however, were not inferior to them, as some studies showed lower accuracy.[9-11] This implies that the saying "the more, the merrier" does not apply to the number of epochs and the size of the dataset in developing DCNNs for implant classification. Even though some research has shown almost perfect accuracy in implant classification,[8] the increased number of implant systems used in real-world settings might require a larger dataset to maintain high accuracy. It is worthwhile to search for more efficient ways to maintain dataset quality.

The ideal implant fixture classification by deep learning should classify all kinds of implant fixtures available in the market, although hundreds of systems exist.[4] This is the major limitation of this study and other studies with the same purpose because the largest number of implant systems evaluated in a single study was 11.[10] Another limitation is that it did not include a testing process with external data. Even though this study showed that high performance in implant fixture classification by YOLOv3 can be accomplished through transfer learning even with a small amount of data, comprehensive studies are expected in the future. Larger studies with more implant systems and heterogeneous datasets from various centers would help refine automatic implant fixture classification to the point that it is ready for use in clinical settings.

Conflicts of Interest: None

References

1. Carlsson L, Röstlund T, Albrektsson B, Albrektsson T, Brånemark PI. Osseointegration of titanium implants. Acta Orthop Scand 1986; 57: 285-9.
2. Elani HW, Starr JR, Da Silva JD, Gallucci GO. Trends in dental implant use in the U.S., 1999-2016, and projections to 2026. J Dent Res 2018; 97: 1424-30.
3. Jokstad A, Braegger U, Brunski JB, Carr AB, Naert I, Wennerberg A. Quality of dental implants. Int Dent J 2003; 53(6 Suppl 2): 409-43.
4. Terrabuio BR, Carvalho CG, Peralta-Mamani M, Santos PS, Rubira-Bullen IR, Rubira CM. Cone-beam computed tomography artifacts in the presence of dental implants and associated factors: an integrative review. Imaging Sci Dent 2021; 51: 93-106.
5. Karl M, Irastorza-Landa A. In vitro characterization of original and nonoriginal implant abutments. Int J Oral Maxillofac Implants 2018; 33: 1229-39.
6. Michelinakis G, Sharrok A, Barclay CW. Identification of dental implants through the use of Implant Recognition Software (IRS). Int Dent J 2006; 56: 203-8.
7. Tyndall DA, Price JB, Tetradis S, Ganz SD, Hildebolt C, Scarfe WC, et al. Position statement of the American Academy of Oral and Maxillofacial Radiology on selection criteria for the use of radiology in dental implantology with emphasis on cone beam computed tomography. Oral Surg Oral Med Oral Pathol Oral Radiol 2012; 113: 817-26.
8. 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. Medicine (Baltimore) 2020; 99: e20787.
9. Takahashi T, Nozaki K, Gonda T, Mameno T, Wada M, Ikebe K. Identification of dental implants using deep learning - pilot study. Int J Implant Dent 2020; 6: 53.
10. Sukegawa S, Yoshii K, Hara T, Yamashita K, Nakano K, Yamamoto N, et al. Deep neural networks for dental implant system classification. Biomolecules 2020; 10: 984.
11. Hadj Saïd M, Le Roux MK, Catherine JH, Lan R. Development of an artificial intelligence model to identify a dental implant from a radiograph. Int J Oral Maxillofac Implants 2020; 35: 1077-82.
12. Putra RH, Doi C, Yoda N, Astuti ER, Sasaki K. Current applications and development of artificial intelligence for digital dental radiography. Dentomaxillofac Radiol 2022; 51: 20210197.
13. Lee JH, Han SS, Kim YH, Lee C, Kim I. Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs. Oral Surg Oral Med Oral Pathol Oral Radiol 2020; 129: 635-42.
14. Kise Y, Shimizu M, Ikeda H, Fujii T, Kuwada C, Nishiyama M, et al. Usefulness of a deep learning system for diagnosing Sjögren's syndrome using ultrasonography images. Dentomaxillofac Radiol 2020; 49: 20190348.
15. Ariji Y, Fukuda M, Kise Y, Nozawa M, Yanashita Y, Fujita H, et al. Contrast-enhanced computed tomography image assessment of cervical lymph node metastasis in patients with oral cancer by using a deep learning system of artificial intelligence. Oral Surg Oral Med Oral Pathol Oral Radiol 2019; 127: 458-63.
16. Morid MA, Borjali A, Del Fiol G. A scoping review of transfer learning research on medical image analysis using ImageNet. Comput Biol Med 2021; 128: 104115.
17. Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, et al. Deep learning in medical imaging: general overview. Korean J Radiol 2017; 18: 570-84.
18. Ramesh A. Panoramic imaging. In: Mallya SM, Lam EW. White and Pharoah's oral radiology. 8th ed. St. Louis: Elsevier; 2019. p. 132-50.
19. Chang E. Dental Implants. In: Mallya SM, Lam EW. White and Pharoah's oral radiology. 8th ed. St. Louis: Elsevier; 2019. p. 248-70.