Journal of Prosthodontic Research

Official Journal of Japan Prosthodontic Society

Original Article

A system for designing removable partial dentures using artificial intelligence. Part 1. Classification of partially edentulous arches using a convolutional neural network

Toshihito Takahashi [a*], Kazunori Nozaki [b], Tomoya Gonda [a], Kazunori Ikebe [a]

[a] Department of Prosthodontics, Gerodontology and Oral Rehabilitation, Osaka University Graduate School of Dentistry. Japan
[b] Division of Medical Information, Osaka University Dental Hospital. Japan

Abstract

Purpose: The purpose of this study was to develop a method for classifying dental arches using a convolutional neural network (CNN) as the first step in a system for designing removable partial dentures.

Methods: Using 1184 images of dental arches (maxilla: 748 images; mandible: 436 images), arches were classified into four arch types: edentulous, intact dentition, arches with posterior tooth loss, and arches with bounded edentulous space. A CNN method to classify images was developed using Tensorflow and Keras deep learning libraries. After completion of the learning procedure, the diagnostic accuracy, precision, recall, F-measure and area under the curve (AUC) for each jaw were calculated for diagnostic performance of learning. The classification was also predicted using other images, and percentages of correct predictions (PCPs) were calculated. The PCPs were compared with the Kruskal-Wallis test (p = 0.05).

Results: The diagnostic accuracy was 99.5% for the maxilla and 99.7% for the mandible. The precision, recall, and F-measure for both jaws were 0.25, 1.0 and 0.4, respectively. The AUC was 0.99 for the maxilla and 0.98 for the mandible. The PCPs of the classifications were more than 95% for all types of dental arch. There were no significant differences among the four types of dental arches in the mandible.

Conclusions: The results of this study suggest that dental arches can be classified and predicted using a CNN. Future development of systems for designing removable partial dentures will be made possible using this and other AI technologies.

Keywords: *Removable partial denture, Machine learning, Artificial intelligence, Convolutional neural network*

1. Introduction

The design of removable partial dentures (RPDs) influences not only the outcome of the prosthesis itself, but also the prognosis of the abutment and residual teeth [1, 2]. The design also influences oral functions, such as mastication, pronunciation, and swallowing, and general quality of life factors, such as enjoyment of food [3] and cognitive function [4]. For these reasons, clinicians bear a heavy responsibility to design a high quality RPD for each patient, using all the available information about RPD design. However, clinicians are usually limited by their inadequate knowledge and experience of designing RPDs, and many clinicians will design different RPDs even if the oral situation is the same. Hence, quality of RPDs is affected by the clinicians' knowledge, experience or skills of RPD treatment. Therefore, there is a need for an automated system for designing RPDs based on the oral conditions of the patient, and that is not dependent on the clinician's knowledge and experience.

In recent years, artificial intelligence (AI) technologies have been applied in various fields, obviating the need for human input in some cases. In the medical field, studies and clinical applications using AI have been developed in the diagnosis of breast cancer [5] and leukemia [6], yielding results comparable to or superior to diagnoses made by humans. In the dental field, AI has been applied in the diagnosis of dental caries [7] and oral cancer [8] with accuracy comparable to human beings. These applications have involved the use of machine learning, in which a convolutional neural network (CNN) is developed to make predictions from various target images.

Several trials have been conducted to develop an AI system for designing RPDs. When designing RPDs using AI, the computer itself needs to acquire information about the dental arch and the residual teeth to formulate predictions.

The aim of this study was to develop a system for designing RPDs using AI involving the classification of the dental arch using a CNN as the first step.

* Corresponding author at: Department of Prosthodontics, Gerodontology and Oral Rehabilitation, Osaka University Graduate School of Dentistry, 1-8 Yamadaoka, Suita, Osaka 565-0871, Japan.
E-mail address: toshi-t@dent.osaka-u.ac.jp (T.Takahashi).

2. Materials and Methods

2.1. Data collection

A total of 1184 oral photographic images (748 maxillary and 438 mandibular) were obtained from patients. The images consisted of four types of dental arches: edentulous, arches with posterior tooth loss (distal extension missing), arches with bounded edentulous space (intermediate missing), and intact dentition (without missing) in each jaw. These images were judged the arch types previously by authors using the same oral images. They were JPEG files which were resized to 224 × 224 pixels. The images were randomly divided into two datasets: one for training (1016 images; 656 maxillary and 360 mandibular) and one for testing (168 images; 92 maxillary and 76 mandibular). This study protocol was approved by the Ethical Committee (H30-E26).

2.2. Algorithm of the convolutional neural network

In a deep CNN procedure, Keras was used as the library with Tensorflow as the backend. The ImageNet pretrained model of a 152-layer residual network model (ResNet152) [9] with fine tuning was used for preprocessing, and the dataset was trained to classify the type of dental arch. The training dataset was separated into 10 batches for every epoch and 50 epochs were run at a learning rate of 0.01.

2.3. Statistical analysis

The diagnostic accuracy, precision, recall, F-measure, and the area under the receiver operating characteristic (ROC) curve (i.e., the AUC) of each jaw were calculated to assess the test dataset. After completion for the learning procedure, predictions of the dental arch in both jaws were conducted using 20 other images of each arch type, and the percentages of correct prediction (PCPs) were recorded. Comparisons were made of the median of PCPs for each arch type using the Kruskal-Wallis test with a post-hoc comparison using the Bonferroni method. All statistical analyses were performed using SPSS statistics Ver22 (IBM, Armonk, NY).

3. Results

In the maxilla, diagnostic accuracy was 99.5%, precision was 0.25, recall was 1.0, F-measure was 0.4, and AUC was 0.99. In the mandible, diagnostic accuracy was 99.7%, precision was 0.25, recall was 1.0, F-measure was 0.4, and AUC was 0.98 (Table 1).

Figure 1 shows an example of PCP results in the maxilla. In these images, the PCPs of edentulous, distal extension missing, and without missing group were 100%, and that of intermediate missing group was 99.8%. The median of PCP results in the maxilla were 100.0% (99.7-100.0) (edentulous), 99.4% (99.2-99.9) (distal extension missing), 97.5% (95.1-99.3) (intermediate missing), and 97.9% (97.7-99.9) (without missing). There was a significant difference between edentulous and intermediate missing group in the maxilla (p = 0.017) (Fig. 2). In the mandible, the median of PCP results were 100.0% (99.7-100.0) (edentulous), 99.3% (96.3-100.0) (distal extension missing), 98.8% (97.9-99.9) (intermediate missing), and 98.8% (96.4-100.0) (without missing). There were no significant differences among the four types of dental arches in the mandible (p = 0.358) (Fig. 2).

4. Discussion

The development of a system for designing RPDs using AI technology was firstly reported in 1980s as an "expert system" [10, 11]. Such systems were programmed by clinicians with special knowledge, techniques, and enough experience with RPDs so that, in a real sense, the rules for the design of the RPD were decided by clinicians.

Table 1. Diagnostic accuracy, precision, recall, F-measure, and AUC of the maxilla and mandibula.

	diagnostic accurary(%)	precision	recall	F-measure	AUC
maxilla	99.5	0.25	1	0.4	0.99
mandibula	99.7	0.25	1	0.4	0.98

Although this new system is also programmed with information about past RPDs provided by clinicians, the computer itself creates the rules for the design of RPDs via autonomous learning. This system can archive information through AI technology, such as big data and machine learning, and predict the design of RPDs from the results. The flow of this classification system is as below. The computer reads the training dataset which is already classified into true arch types and creates (predicts) the classification rule by learning them. The testing dataset is used to test the predicted rule and the accuracy of learning is improved by learning again and again. After finishing the learning, the computer classifies new images of dental arch with prediction rate based on the rule. Clinicians finally have to judge whether the prediction is correct or not.

Basic information about the intraoral condition, such as the location of missing teeth, and the prosthetic situation of the residual teeth and occlusion, is necessary for the computer to learn and create the design of an RPD. Therefore, in this study, the dental arches were classified using oral photographic images and a CNN (an AI technology) as the first step in developing a system for designing RPDs, and the diagnostic accuracy of learning and the prediction of dental arches was evaluated.

When evaluating the effectiveness of machine learning, diagnostic accuracy was the most important factor, and the ideal value for diagnostic accuracy was 100%. The ideal values of other indicators, such as recall, precision, and, importantly, the F-measure, were 1.0. The AUC is an index of learning performance, and the ideal value was close to 1.0. In this study, the diagnostic accuracy of learning was more than 99% and the AUC was more than 0.98 in both jaws. These accuracy values were higher than those of previous studies using a CNN [12], and a value of more than 90% was considered suitable for application in a clinical setting. Although recall values for both jaws were almost 1.0, the precision and F-measure values were limited to 0.25 and 0.4, respectively. These results indicate that correct answers were fully judged as correct, but several wrong answers were also judged as correct in this system. In other words, this system still contains some uncertainties, despite the high diagnostic accuracy and AUC. To improve the quality of the learning for use in clinical settings, wrong answers need to be correctly judged as wrong and therefore the value of the F-measure must be raised to at least 0.8.

The learning procedure of this system consisted of 50 epochs, a batch size of 10 and a learning rate of 0.01. The number of epochs is an important factor in the accuracy of deep learning and, in general, the higher the number of epochs, the more accurate the learning will be. The batch size and learning rate are also important factors in the diagnostic accuracy. The greater the batch size, the longer the learning time will be. In the pre-learning procedure to decide learning conditions, several trials were conducted with different epoch numbers and batch sizes. As the result, more than 50 epochs did not increase the accuracy, and a batch size of more than 10 only prolonged the learning time. A learning rate of 0.01 was usually used in past studies using CNN [12]. Other rates were also tried in pre-learning procedure and resulted in lower accuracy. Therefore, these three values were selected in this study.

In this classification system, each image was presumably classified into arch type according to color distribution. In the edentulous group, an image is consisted of only red color of residual ridge, palate and floor of mouth. On the other hand, in other three groups it is consisted of combination of white color of residual teeth, metallic colors of
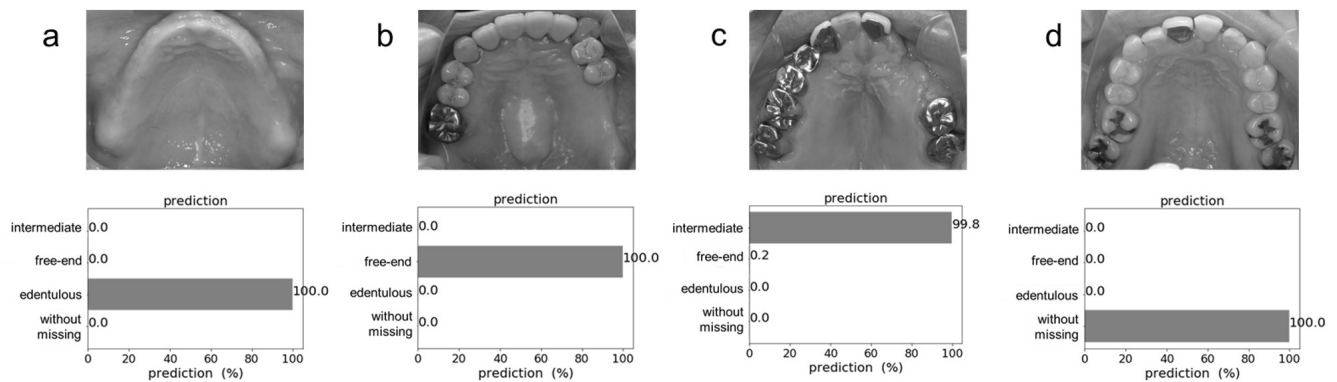
Fig. 1. Photographic images and graphs showing an example of the percentage of correct predictions (PCPs) in the maxilla. a: edentulous, b: arches with posterior tooth loss (distal extension missing), c: arches with bounded edentulous space (intermediate missing), and d: intact dentition (without missing).
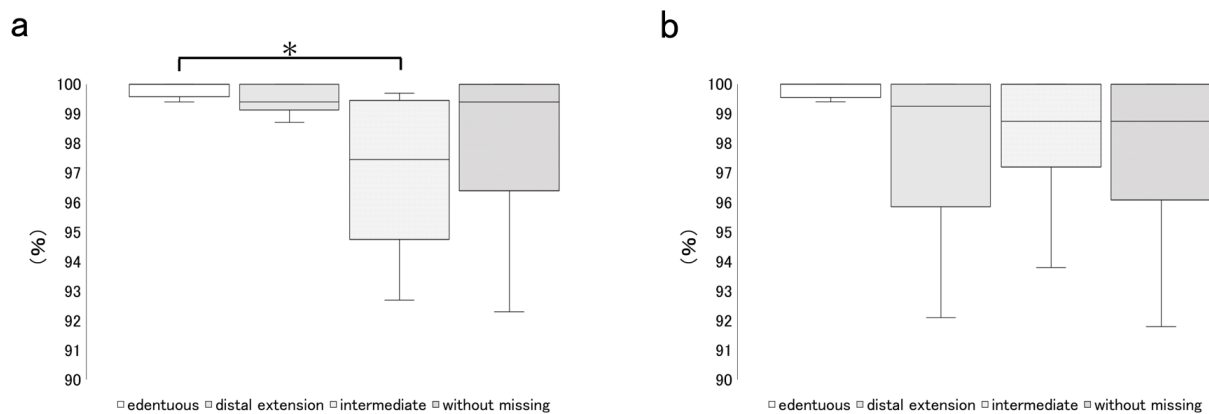

Fig. 2. Graphs of the median of percentages of correct predictions (PCPs). * denotes a significant difference (p < 0.05). a: maxilla, b: mandible. Edentulous: edentulous, distal extension: arches with posterior tooth loss, intermediate: arches with bounded edentulous space, and without missing: intact dentition.

restorations and red color, and it is classified into each dental arch according to this color distribution. The results of prediction were higher in edentulous and without missing because color distribution was simple and uniform, but those of other two arches were lower because color distribution was various depend on the distribution of missing teeth and residual teeth. Especially in the intermediate missing group, there were more patterns of missing teeth and residual teeth than in the other three arch types. This variation made it difficult to make accurate prediction and the difference of prediction accuracy in intermediate missing group between two jaws. The number of images of maxilla was much more than that of mandibula and therefore, the patterns of missing teeth and residual teeth in the intermediate missing group for maxilla were more various than those of mandibula. This made the prediction accuracy of maxilla lower, and difference between two jaws was occurred.

In this study, images used in both learning were selected without special exclusion criteria in order to correct images as many as possible. Therefore, there were some images with problems such as small lacking of dental arch, out of focus and these images may reduce the learning performance. In the case of prediction procedure, when using imperfect images aforementioned above, computer will recognize them as they are and predict the dental arch. As the result, computer will predict them as wrong type of dental arch. Considering these, complete images in these points should be necessary to predict correctly.

The PCPs of this study were more than 95% in all dental arches.

These scores are almost comparable with the visual ability of human beings. Although the PCPs were much higher than those of our preliminary study because of the use of new architecture with greater accuracy, there were still significant differences between the intermediate missing teeth and the other arch types in the maxilla. To improve the accuracy, more images are required, especially of intermediate missing in the maxilla. Our finding that the PCPs were greater than 95% in all types of dental arches suggests that this system can be applied and a system for designing RPDs can be developed when combined with other systems using artificial intelligence. Further steps to be developed in the future include the development of image recognition using photographic images of components of RPDs and panoramic X-rays or images of residual teeth.

5. Conclusion

Within the limitations of this study, the following conclusions were drawn.

1. The diagnostic accuracy was 99.5% for the maxilla and 99.7% for the mandible.

2. The percentage of correct predictions was more than 95% in all dental arches. In the maxilla, there was a significant difference between edentulous and arches with bounded edentulous space. No significant differences were observed among four arch types in the mandible.

Conflicts of interest

The authors report no conflicts of interest related to this study.

References

[1] Tada S, Allen PF, Ikebe K, Zheng H, Shintani A, Maeda Y. The impact of the crown-root ratio on survival of abutment teeth for dentures. J Dent Res 2015; 94: 220S-5S.

[2] Mizuno Y, Gonda T, Takahashi T, Tomita A, Maeda Y. Root Fracture of Abutment Teeth for Partial Removable Dental Prostheses. Int J Prosthodont 2016; 29: 461-6.

[3] Inomata C, Ikebe K, Okubo H, Takeshita H, Mihara Y, Hatta K, et al. Dietary intake is associated with occlusal force rather than number of teeth in 80-y-old Japanese. JDR Clin Trans Res 2017; 2: 187-97.

[4] Hatta K, Ikebe K, Gondo Y, Kamide K, Masui Y, Inagaki H, et al. Influence of lack of posterior occlusal support on cognitive decline among 80-year-old Japanese people in a 3-year prospective study. Geriatr Gerontol Int 2018; 18: 1439-46.

[5] Zhu Z, Albadawy E, Saha A, Zhang J, Harowicz MR, Mazurowski MA. Deep learning for identifying radiogenomic associations in breast cancer. Comput Biol Med 2019; 109: 85-90.

[6] Rehman A, Abbas N, Saba T, Rahman SIU, Mehmood Z, Kolivand H. Classification of acute lymphoblastic leukemia using deep learning. Microsc Res Tech 2018; 81: 1310-7.

[7] Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent 2018; 77: 106-11.

[8] Jeyaraj PR, Samuel Nadar ER. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. J Cancer Res Clin Oncol 2019; 145: 829-37.

[9] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arXiv 2015: 1512. 03385.

[10] Maeda Y, Tsutsumi S, Okada M, Minoura M, Nobuki T, Okuno Y. An expert system for designing removable partial dentures (Part 1) Selection of Abutments. J Jpn Prosthodont Soc 1987; 29: 833-8.

[11] Maeda Y, Tsutumi S, Minoura M, Okada M, Nokubi T, Okuno Y. An expert system for designing removable partial dentures. Part 2: the role of the data base constructed with clinical cases. J Jpn Prosthodont Soc 1985; 31: 1531-35.

[12] Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical Image Analysis using Convolutional Neural Networks: A Review. J Med Syst 2018; 42: 226.