Journal of Dentistry xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Dentistry

journal homepage: www.elsevier.com/locate/jdent

Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm

Jae-Hong Lee[a,*], Do-Hyung Kim[a], Seong-Nyum Jeong[a], Seong-Ho Choi[b]

[a] Department of Periodontology, Daejeon Dental Hospital, Institute of Wonkwang Dental Research, Wonkwang University College of Dentistry, Daejeon, Republic of Korea
[b] Department of Periodontology, Research Institute for Periodontal Regeneration, Yonsei University College of Dentistry, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Artificial intelligence
Dental caries
Machine learning
Supervised machine learning

ABSTRACT

Objectives: Deep convolutional neural networks (CNNs) are a rapidly emerging new area of medical research, and have yielded impressive results in diagnosis and prediction in the fields of radiology and pathology. The aim of the current study was to evaluate the efficacy of deep CNN algorithms for detection and diagnosis of dental caries on periapical radiographs.
Materials and methods: A total of 3000 periapical radiographic images were divided into a training and validation dataset (n = 2400 [80%]) and a test dataset (n = 600 [20%]). A pre-trained GoogLeNet Inception v3 CNN network was used for preprocessing and transfer learning. The diagnostic accuracy, sensitivity, specificity, positive predictive value, negative predictive value, receiver operating characteristic (ROC) curve, and area under the curve (AUC) were calculated for detection and diagnostic performance of the deep CNN algorithm.
Results: The diagnostic accuracies of premolar, molar, and both premolar and molar models were 89.0% (80.4–93.3), 88.0% (79.2–93.1), and 82.0% (75.5–87.1), respectively. The deep CNN algorithm achieved an AUC of 0.917 (95% CI 0.860–0.975) on premolar, an AUC of 0.890 (95% CI 0.819–0.961) on molar, and an AUC of 0.845 (95% CI 0.790–0.901) on both premolar and molar models. The premolar model provided the best AUC, which was significantly greater than those for other models ($P < 0.001$).
Conclusions: This study highlighted the potential utility of deep CNN architecture for the detection and diagnosis of dental caries. A deep CNN algorithm provided considerably good performance in detecting dental caries in periapical radiographs.
Clinical significance: Deep CNN algorithms are expected to be among the most effective and efficient methods for diagnosing dental caries.

1. Introduction

Dental caries are common chronic infectious oral diseases affecting most teenagers and adults worldwide [1]. According to the National Health and Nutrition Examination Survey, the prevalence of dental caries in the United States is 41% among children 2–11 years of age (in their primary teeth), 42% in children and adolescents 6–19 years of age, and approximately 90% among adults ≥20 years of age (in their permanent teeth) [2–4]. Most studies have reported that socially and economically disadvantaged individuals, including low-income minorities, those with lower levels of education, and disability groups, are at higher risk for dental carries. In addition, several epidemiologicaland clinical studies have reported that tooth loss caused by oral disease, including dental caries, is related to detrimental dietary changes, and potentially modifiable risk factors or risk indicators for cardiovascular disease and cognitive impairment [5,6].

Various retention and restoration methods, which have been proposed and improved for the treatment of dental caries, have been successfully developed over the past few decades [7,8]. However, there has not yet been a significant improvement in the diagnostic methodology for detecting dental caries due to various anatomical morphologies of teeth and the shapes of restorations. In particular, when deep fissures, tight interproximal contacts, and secondary lesions are present, it is difficult to detect early-stage disease and, eventually, many lesions are detected in the advanced stages of dental caries. Therefore, although dental radiography (including panoramic, periapical, and bitewing views), and explorer (or dental probe), which are widely used and regarded to be highly reliable diagnostic tools for the detection of

* Corresponding author at: Department of Periodontology, Daejeon Dental Hospital, Wonkwang University College of Dentistry, 77, Dunsan-ro, Seo-gu, Daejeon 35233, Republic of Korea.
E-mail address: ljaehong@gmail.com (J.-H. Lee).

https://doi.org/10.1016/j.jdent.2018.07.015
Received 4 June 2018; Received in revised form 21 July 2018; Accepted 25 July 2018

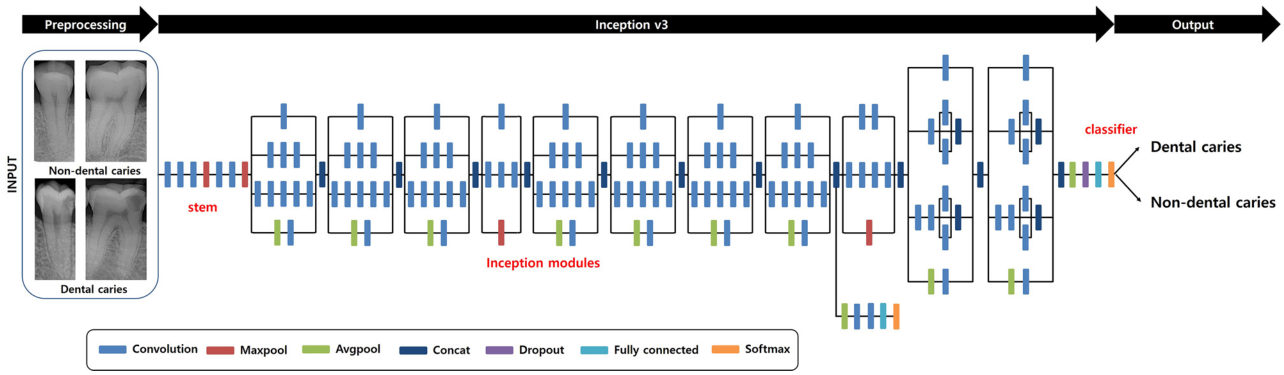

Fig. 1. Overall architecture of the pre-trained convolutional neural network model. The architecture has a total of 9 inception modules, including an auxiliary classifier, and two fully connected layers. The output layer performs binary classification between dental caries and non-dental caries using a softmax function.

dental caries, much of the screening and final diagnosis tends to rely on empirical evidence.

Recently, one aspect of artificial intelligence and deep learning—-convolutional neural networks (CNNs)—has demonstrated excellent performance in computer vision including object, facial and activity recognition, tracking, and three-dimensional mapping and localization [9]. Medical segmentation and diagnosis is one of the most important fields in which image processing and pattern recognition procedures have been adopted. In particular, detection and classification of diabetic retinopathy, skin cancer, and pulmonary tuberculosis using deep learning-based CNN models have already demonstrated very high accuracy and efficiency, with promising clinical applications [10,11]. In contrast, however, there have been few studies based on deep CNN architectures in the dental field, and research investigating detection and diagnosis of dental caries is also more limited [12]. Accordingly, the aim of the present study was to evaluate the efficacy of deep CNN algorithms for the detection and diagnosis of dental caries in periapical radiographs.

2. Materials and methods

2.1. Datasets

This study was conducted at the Department of Periodontology, Daejeon Dental Hospital, Wonkwang University and approved by the Institutional Review Board of Daejeon Dental Hospital, Wonkwang University (approval no. W1804/003-001). Anonymized periapical radiographic image datasets, acquired between January 2016 and December 2017, were obtained from the authors' dental hospital's PACS system (Infinitt PACS, invented by Infinitt Co., Seoul, Korea), and classified and labeled based on electronic medical records (EMR). All images were clearly revalidated, and dental caries, including enamel and dentinal carious lesions (excluding deciduous teeth), were distinguished from non-dental caries by four calibrated board-certified dentists. The dataset excluded all periapical radiographic images in which the diagnosis of the four examiners did not match, and included the periapical radiographic images for which all four examiners agreed to the diagnosis of dental caries.

The dataset consisted of a total of 3000 periapical radiographic images of 778 (25.9%) maxillary premolars, 769 (25.6%) maximally molars, 722 (24.1%) mandibular premolars, and 731 (24.4%) mandibular molars. There were 781 (23.9%) premolars and 772 (25.7%) molars that were diagnosed as dental caries, and 719 (26.1%) premolars and 728 (24.3%) molars diagnosed as non-dental caries. Periapical radiograph images diagnosed as dental caries and non-dental caries were cropped to show only one tooth per image and optimal position. Images with moderate-to-severe noise, haziness, distortion, and shadows, and those with a full crown or large partial onlay restoration were excluded from the analysis. Periapical radiographic images included in the dataset were then calibrated to standardize contrast between gray/white matter and lesions.

2.2. Preprocessing and image augmentation

A total of 3000 periapical radiographic images were finally selected and resized to 299 × 299 pixels and converted into JPEG file format. All maxillary teeth images were reverted to the mandibular teeth form through a vertical flip. A randomization sequence was created using SPSS (IBM Corporation, Armonk, NY, USA), and used to divide the dataset into a training and validation dataset (n = 2400 [80%]), and a test dataset (n = 600 [20%]). The training and validation dataset consisted of 1200 dental caries and 1200 non-dental caries, and the test dataset consisted of 300 dental caries and 300 non-dental caries in the same ratio. The training dataset was randomly augmented 10 times using rotation (range of 10°), width and height shifting (range, 0.1), zooming (range, 0.8–1.2), shearing (range, 0.5), and horizontal flip [13].

2.3. Architecture of the deep convolutional neural network algorithm

A pre-trained GoogLeNet Inception v3 CNN network was used for preprocessing, and the datasets were trained using transfer learning. The Inception v3 architecture, which demonstrated excellent performance in the 2014 ImageNet Large Scale Visual Recognition Challenge, has preliminarily learned approximately 1.28 million images consisting of 1000 object categories. It consists of 22 deep layers, and it is possible to obtain different scale features by applying convolutional filters of different sizes in the same layer. A total of 9 inception modules were used, including an auxiliary classifier, two fully connected layers, and softmax functions [14]. The training dataset was separated randomly into 32 batches for every epoch, and 1000 epochs were run at a learning rate of 0.01. To provide better detection of dental caries, fine-tuning was used to optimize the weights and improve the output power by adjusting the hyperparameters [15,16] (Fig. 1).

2.4. Statistical analysis

The training and validation dataset was used to estimate and create optimal deep CNN algorithm weight factors. All deep CNNs in this study were implemented using the Keras library on top of TensorFlow in Python. The diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), receiver operating characteristic (ROC) curve, and area under the ROC curve (AUC) of the test dataset were assessed. P values < 0.05 were considered to be statistically significant, and 95% confidence intervals (CIs) were calculated.