A Single-Step Prediction of Inferior Alveolar Nerve Injury After Mandibular Third Molar Extraction Using Contrastive Learning and Bayesian Auto-Tuned Deep Learning Model

Myoungho Lee

The Graduate School

Yonsei University

Department of Dentistry

A Single-Step Prediction of Inferior Alveolar Nerve Injury After Mandibular Third Molar Extraction Using Contrastive Learning and Bayesian Auto-Tuned Deep Learning Model

Directed by Professor Kee-Deog Kim

A Dissertation Thesis

Submitted to the Department of Dentistry

and the Graduate School of Yonsei University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Dental Science

Myoungho Lee

June 2024

This certifies that the Doctoral dissertation

of Myoung-Ho Lee is approved.

Thesis Supervisor: Kee-Deog Kim

Thesis Committee Member: Hyung-Jun Kim

Thesis Committee Member: Wonse Park

Thesis Committee Member: Jong-Eun Choi

Thesis Committee Member: Jun-Young Kim

The Graduate School

Yonsei University

June 2024

감사의 글

20살부터 물리학과를 전공하고 군대를 가고, 다시 공부하여 치과의사가 되고, 결혼을 하였고, 통합치의학과 수련을 마치고 두 아이의 아빠가 되었으며 또 개원을 하였습니다. 정신없이 앞만 보고 달린 20년을 마무리하면서 도전했던 치의학박사의 꿈, 개원의에게는 많이 힘들었고 포기하고 싶은 적도 많았습니다. 이룰 수 있도록 도와주었던 많은 분들께 진심으로 감사드립니다. 앞으로도 그분들과 가족들에게 부끄럽지 않은 인생을 살도록 더욱 열심히 정진하겠습니다.

통합치의학과 수련을 받는 동안 늘 저의 존경의 대상이 였던 김기덕 지도 교수님 감사합니다. 그동안 주신 은혜와 배려 졸업 후에도 잊지 않겠습니다. 바쁜 신 중에도 부족한 제자를 끝까지 포기 하지 않고 끌고 와 주신, 가장 고생 많으셨던 저의 선생님 박원서 교수님께도 진심으로 감사드립니다. 또한 구강악안면외과 제자들 지도만으로도 바쁘실 텐데 저의 논문까지 지도해주시고 따뜻한 말로 안아 주신 김형준 교수님 감사드립니다. 그리고 논문을 가장 꼼꼼히 봐주시고, 마지막까지 검토해 주셨던 김준영 교수님 은혜 잊지 않겠습니다. 또 최종은 교수님, 저의 지식과 치과대학의 데이터만으로는 불가능했던 부분을 채워 주셨던 은혜, 그리고 저의 논문까지 지도해 주셨던 은혜 감사드립니다. 더불어 같이 연구 해주셨던 윤규백 교수님께도 감사의 인사를 드립니다. 마지막으로 2012년부터 치과의사로서의 기본을 가르쳐 주시고 지금까지 정진할 수 있게 해주신 은사님 정복영 교수님, 방난심 교수님께도 감사드립니다.

저의 연구를 도와주셨던 김재연선생님과 최이슬선생님, 4년 동안 선생님들의 소통과 도움이 없었으면, 절대로 이룰 수 없었던 일이였습니다. 감사하고 또 감사드립니다. 앞으로 추가 연구를 위한 노력에 최대한 도움이 되도록 하겠습니다. 또 4년동안 박사과정을 잘 마칠 수 있도록 도와준 연세고마운치과 박대근 팀장님 이하 실장님 및 직원분들께도 감사드립니다. 마지막으로 연구와 논문이 막히고 포기하고 싶을 때 항상 믿음을 준 김한상, 정선호, 김민성, 고동우, 이상현, 유병관 저의 죽마고우들 감사합니다. 항상 힘든 일이 있을 때 위로가 되주고, 치의학 자문을 준 우리 연세고마운 치과 네트워크 원장님들도 모두 감사합니다.

끝으로 초등학교 입학부터 마지막 박사 공부를 끝낼 때까지 항상 묵묵히 지켜봐 주시고 저보다 먼저 많이 걱정하고 기뻐해 주셨던, 이종선 아버님, 최희자 어머님 감사합니다. 그리고 저의 옆에 있는 것만으로도 힘이 되는 가장 소중한 가족 와이프 문효진과 아들 이형후, 딸 이슬아에게도 고맙다는 말을 하고 싶습니다. 항상 내일처럼 기뻐해주고 응원해주신 문진주 장인어른, 정금례 장모님께도 진심으로 감사드립니다.

2024 년 6 월

이 명 호

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

A Single-Step Prediction of Inferior Alveolar Nerve Injury After Mandibular Third Molar Extraction Using Contrastive Learning and Bayesian Auto-Tuned Deep Learning Model

Myoungho Lee

Department of Dentistry

The Graduate School, Yonsei University

(Directed by Professor Kee-Deog Kim, D.D.S., M.S.D., Ph.D.)

In recent years, artificial intelligence (AI) and deep learning have revolutionized various scientific and technological fields, showing tremendous potential in solving complex problems. Among the numerous applications of deep learning, image classification has emerged as a powerful technique that can identify and classify objects within images with impressive accuracy.

The most frequently performed oral and maxillofacial surgery is mandibular

third molar extraction, and inferior alveolar nerve damage is one of the most serious postoperative complications that can occur temporarily or permanently. Therefore, many studies are being conducted to predict the occurrence of nerve damage after tooth extraction through the analysis of pre-extraction radiographs. However, there are limitations to evaluating the possibility of sensory abnormalities using images. If the clinician has little experience or lacks knowledge, risk factors may be difficult to detect and interobserver error may occur. Additionally, no matter how experienced the clinician or oral radiologist is, errors are likely to occur if the radiograph is unclear or multiple findings overlap.

To overcome these limitations, research has recently been conducted using AI to evaluate the relationship between the mandibular third molars and the inferior alveolar nerve. However, studies using deep learning also have unique limitations.

This is a step to detect or identify the inferior alveolar nerve and mandibular third molar, but no research has been conducted to analyze the correlation with clinical data regarding actual nerve damage after tooth extraction or to predict nerve damage using cone beam computed tomography (CBCT). This issue has arisen because most existing deep learning studies are limited to the use of panoramic images owing to the lack of CBCT data.

Additionally, manually selecting the most appropriate hyperparameters for a deep learning model requires an extensive and time-consuming trial-and-error process when working with limited amounts of data, making it difficult to produce an accurate model.

Therefore, in this paper, qualitative improvement of images using contrast limited adaptive histogram equalization (CLAHE) and quantitative improvement using data augmentation reduced the possibility of damage to the inferior alveolar nerve during tooth extraction using CBCT, and a model was designed using a contrastive learning method and a Siamese network. We propose creating an accurate model through high-parameter auto-tuning using transfer learning and Bayesian optimization.

In contrast to previous studies which have focused on the segmentation of the mandibular third molar and the inferior alveolar nerve and evaluated the possibility of contact, post-tooth extraction was performed in a single step without a segmentation process in our study by learning the results of pre-extraction radiographs and actual clinical sensory abnormalities. A model was created to predict the possibility of sensory abnormalities using preoperative radiographs.

In order to overcome the limitation that the number of patients who developed sensory abnormalities after tooth extraction was significantly lower than the number of patients who did not, the accuracy of the prediction results was secured by using contrastive learning and Bayesian optimization after image processing and amplification of the data. .

We will extend our current research to find a more accurate model to evaluate the possibility of sensory abnormalities before mandibular wisdom tooth extraction by using various cuts of multiple radiographs we have collected. By determining the evaluation areas of the most accurate radiographs and incorporating AI technology in the editing process of these radiographs, we can create a model that is more accurate and practical for clinical application.

Keywords: Third molar extraction; inferior alveolar nerve; numbness; deep learning; CLAHE; data augmentation; contrastive learning; *Siamese network;* transfer learning; Bayesian optimization.

A Single-Step Prediction of Inferior Alveolar Nerve Injury After Mandibular Third Molar Extraction Using Contrastive Learning and Bayesian Auto-Tuned Deep Learning Model

Myoungho Lee

Department of Dentistry

The Graduate School, Yonsei University

(Directed by Professor Kee-Deog Kim, D.D.S., M.S.D., Ph.D.)

I. INTRODUCTION

The impaction of the mandibular third molars, commonly referred to as wisdom teeth, is a frequent occurrence, and their extraction is a routine dental surgery performed by both oral and maxillofacial surgeons and general dentists.

According to Jung Y and Cho B(2013), 60% of individuals over 25 years of age in the Republic of Korea choose to have their mandibular third molars extracted. However, this commonly performed surgery is associated with a variety of complications such as swelling, pain, bleeding, infection, dry socket, and perforation of the maxillary sinus. A study by the Korea Dental Medical Policy Research Institute (2023) revealed that 57.8% of dentists have encountered complications related to the extraction of impacted mandibular third molars, indicating a high frequency of complications. Among the numerous complications, inferior alveolar nerve (IAN) damage is considered one of the most serious side effects. Unlike other complications that can be improved with proper treatment and medication, IAN damage poses a high risk of irreversible and permanent damage, causing lifelong physical discomfort and trauma to the patient. Furthermore, IAN damage accounts for a significant proportion (23.3%) of complications (Korea Dental Medical Policy Research Institute, 2023). Consequently, dentists bear an increasing responsibility to predict and manage these adverse outcomes.

In light of these issues, various studies have been conducted to predict the risk of IAN damage in order to prevent nerve damage after dental extractions and to ensure safe surgical procedures. In particular, numerous approaches involving the analysis of pre-surgical radiographic images have been employed. Rood J et al. (1990) reviewed the literature and proposed seven criteria[1] that can be used to determine the possibility of IAN damage based on panoramic radiographs. However, the analysis of panoramic radiographs has limitations, such as the inability to accurately assess the anatomical shape of the mandibular third molar roots and their relationship with the mandibular canal owing to limitations of the radiographic equipment (Kim J et al., 2006). Consequently, in an effort to overcome these limitations, the evaluation of cone beam computed tomography (CBCT) images prior to the extraction of the mandibular third molars became standardized, and research was conducted on the prediction of nerve damage using CBCT. According to a paper by Suomalainen A et al. (2010), CBCT can more accurately determine the number of tooth roots and their relationship with the mandibular nerve compared with other radiographic images, contributing to the prediction of risks prior to extraction. Nevertheless, despite advances in CBCT techniques, the use of radiographic imaging for predictions still has certain limitations. For instance, radiographs may be unclear or contain duplications or errors, and the prediction of nerve damage by the analyzing dentist can vary because of different levels of expertise and knowledge, making accurate prediction challenging.

[1] 1. Darkening of the root, 2. Deflected roots, 3. Narrowing of the root, 4. Dark and bifid root, 5. Interruption of the white lines, 6. Diversion of the inferior alveolar canal, and 7. Narrowing of the inferior alveolar canal.

These limitations have been dramatically improved with the integration of artificial intelligence (AI) technology in the field of dental medical imaging. AI and deep learning have recently contributed to significant advancements and have been applied to solve complex problems across various scientific and technological fields (Goodfellow I et al. 2016). Among the applications of deep learning, image classification has emerged as a powerful technique capable of identifying and classifying objects within images with impressive accuracy (Krizhevsky A et al. 2017). Notably, the advancement of Convolutional Neural Network (CNN) techniques within the realm of deep learning has been particularly striking across various AI applications (LeCun Y et al. 2015). The advantage of CNNs in capturing metrical features from raw images has been actively employed not only in the medical field but also in dental treatment.

The application of deep learning-based CNN models in the field of dentistry has exhibited high accuracy and efficiency, demonstrating the potential for use in various clinical scenarios. For example, Lee J-H et al. (2018) presented an AI model that recognizes and classifies dental caries based on panoramic radiographic images, and Krois J et al. (2019) introduced a deep learning model for periodontal assessment using radiographs. Furthermore, Yu H et al. (2020) developed an AI model for new skeletal diagnosis using lateral cephalometric radiographs, and Lee K et al. (2020) developed a deep learning model for diagnosing periarticular inflammation of the jaw joint using CBCT images.

Research using AI models to analyze radiographic images has also been conducted in the field of wisdom tooth extraction. Jaskari J et al. (2020) and Zhu T et al. (2021) compared AI recognition of wisdom teeth and the IAN with actual dentist recognition, achieving satisfactory results. Moreover, Yoo J et al. (2021) and Kim B et al. (2021) attempted to assess the possibility of sensory abnormalities after extraction by classifying the difficulty of impacted mandibular third molar based on the relationship between the AI-recognized wisdom tooth and the IAN. In addition, Jeon K et al. (2023) presented an AI model for evaluating the actual approach between the impacted mandibular third molar and the nerve based on panoramic radiographic images.

This study contributes to the effective pre-emptive prediction and evaluation of the mandibular third molar by demonstrating deep learning methods to enhance the accuracy and efficiency of the diagnosis and analysis of radiographic images. Despite recent advancements, the scarcity of CBCT data remains a common limitation among many deep learning studies. Effective predictions using CBCT require ample training data; however, the number of cases involving post-extraction sensory abnormalities is significantly lower than the total number of extraction cases, leading to a shortage of CBCT training data. Networks trained with a limited amount of CBCT data are known to overfit on arbitrary data, reducing the accuracy of predictions (Casalegno F et al., 2019). Moreover, using a limited dataset necessitates a time-consuming and extensive trial-and-error process to manually select the most suitable hyperparameters for the deep learning model, thus further complicating the task.

To address this issue, the current study aims to collect a large quantity of training data from various radiographic images of patients with sensory abnormalities and to determine the most accurate type of radiographic image for the pre-extraction assessment of sensory abnormalities. Furthermore, it intends to develop a model capable of accurate pre-emptive predictions using a limited dataset through the application of recent deep learning advancements such as transfer learning or contrastive learning. By developing a CBCT deep learning model applicable to a small dataset with an auto-tuning feature for hyperparameters through Bayesian optimization, this study seeks to overcome the limitations of existing CBCT models and offer an improved method for accurately predicting post-extraction IAN damage in the mandibular third molar.

II. MATERIALS and METHODS

Data Acquisition

A. Dataset

Retrospective dental radiographic images from patients who visited the advanced general dentistry or oral and maxillofacial surgery departments at Yonsei University Dental Hospital between January 2018 and December 2020 were investigated. The research protocol was approved by the Institutional Review Board (IRB) of Yonsei University Dental Hospital (IRB No.2-2021-0110). All patient data was anonymized, and the requirement for written consent was waived in view of the retrospective study design. The study was conducted in accordance with the principles of the Declaration of Helsinki.

For the collection of dental radiographic images, patients who had undergone pre-extraction radiographs (panoramic view or CBCT) of the mandibular third molar were selected, and 1,000 individuals were randomly chosen. The inclusion criteria applied in the process of collecting dental radiographic images were as follows: (1) patients who had pre-extraction dental radiographs of the mandibular third molars and (2) patients whose IAN numbness symptoms were evaluated through follow-up observation after extraction. Conversely, the exclusion criteria were as follows: (1) patients with unclear IAN damage and (2) patients with incomplete or unclear radiographic images owing to imaging errors. Based on these inclusion and exclusion criteria, dental radiographic images were collected in JPG format. Panoramic radiographs including the IAN and third molar were manually cropped to sizes of 200x200, 400x400, 600x600, 800x800, and 1000x1000 pixels (px)(Figure 1). In a similar manner, periapical views including the IAN and the roots of the mandibular third molar were cropped to a size of 1980x1440 px. For reconstructed axial CBCT images, an experienced dentist with over ten years of practice manually separated the section of the IAN and the root of the mandibular third molar based on a line connecting the center of the teeth, cutting vertically at 1 mm intervals and horizontally at 4 mm intervals to sizes of 70x70, 140x140, and 210x210 px(Figure 2). In CBCT cross-sections along curves, the cross-sectional radiographs of the IAN and the root surface of the third molar were first observed together based on a line connecting the centers of the teeth, and manually cut at 1 mm intervals towards the mesial direction to a size of 250x400 px(Figure 3). For the panoramic CBCT view, an experienced dentist cropped the reconstructed panoramic view in which the inferior alveolar nerve and the third molar overlapped the most to sizes of 100x100, 150x150, and 200x200 px(Figure 4).

The determination of the presence or absence of IAN damage was carried out by dental specialists who classified the cases into two categories based on the analysis of electronic medical records (EMR) and dental radiographs: class 0 for cases without IAN damage and class 1 for those with IAN damage. Following the application of the inclusion and exclusion criteria, 992 panoramic radiographs (252 indicating IAN damage and 740 normal), 839 axial CBCT radiographs (212 indicating IAN damage and 627 normal), and 834 cross-sectional CBCT radiographs (212 indicating IAN damage and 622 normal) were collected.

For the development of the deep learning model, 902 panoramic radiographs size 400x400 (162 indicating IAN damage and 740 normal), 750 axial CBCT radiographs Reference plane (139 indicating IAN damage and 611 normal), and 750 cross-sectional CBCT radiographs Reference plane (139 indicating IAN damage and 611 normal) were utilized.

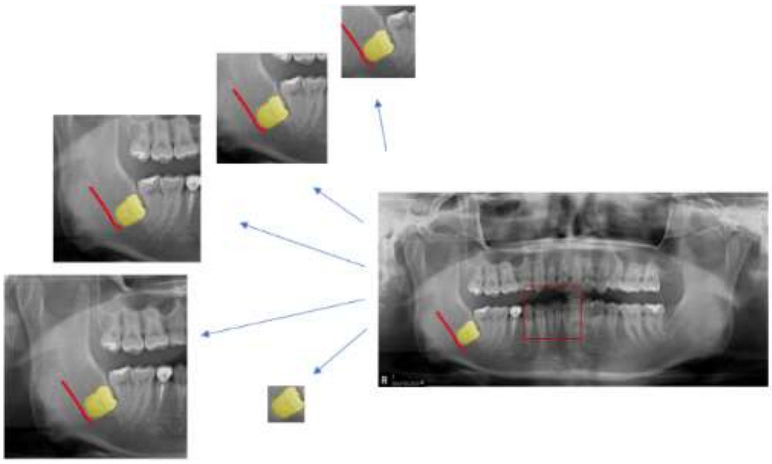


Figure 1. **Panoramic radiographs.**
The images are cropped around the areas where the nerve canal and the tooth are closest to each other and are shown in five different sizes.

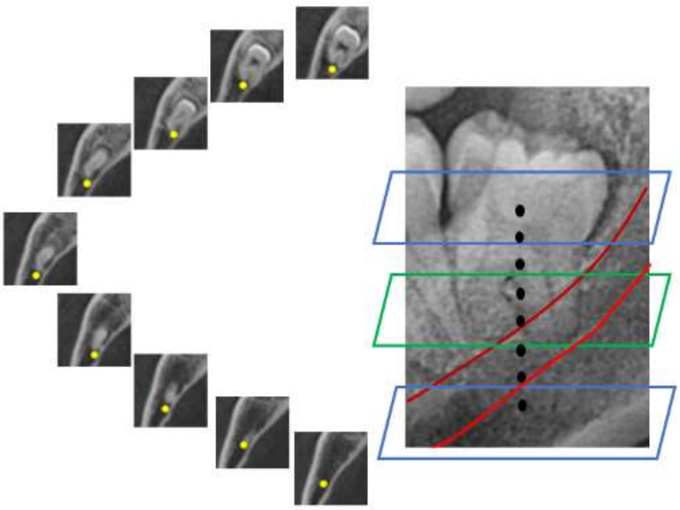


Figure 2. **CBCT - Reconstructed axial CBCT images.**
In these images, an experienced dentist with over ten years of practice has based the cuts in the axial direction on a curve connecting the centerline of the tooth, focusing on the area where the nerve canal and the tooth are closest, cropping at 1 mm vertical intervals.
CBCT, cone beam computed tomography.

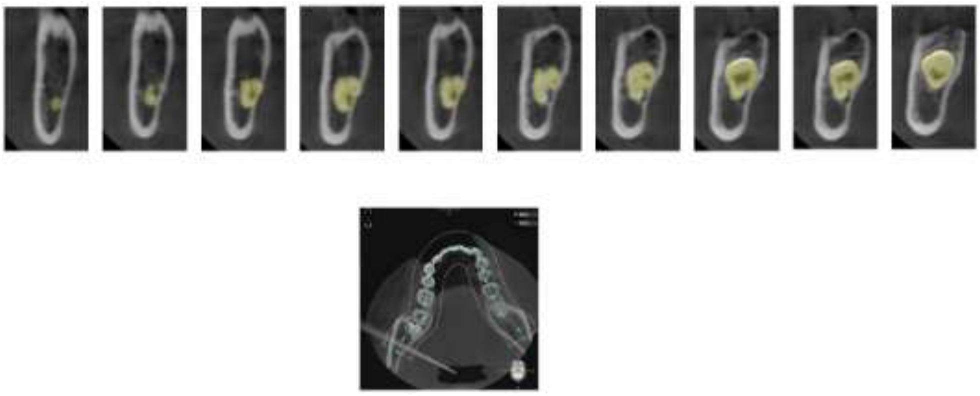


Figure 3. CBCT - Cross-sections along curves.
In these images**,** an experienced dentist used a curve connecting the centerline of the tooth as a guideline to crop in the coronal direction, starting from the surface where the nerve canal and the tooth first appear together and cropping nine images at 1 mm intervals towards the mesial surface.
CBCT, cone beam computed tomography.

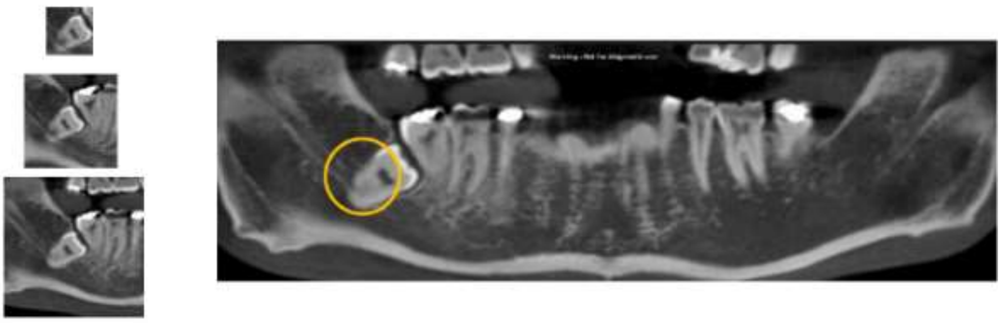


Figure 4. CBCT - Synthetic panoramic image along curves, slightly anterior or posterior.
In these images, an experienced dentist with over ten years of practice has cropped in three different sizes from the determined curve, focusing on the point where the inferior alveolar nerve and the lower third molar are closest, similar to a panoramic view.
CBCT, cone beam computed tomography.

B. Label Reorganization for Contrastive Learning

We employed contrastive learning to counteract the risk of model inaccuracy secondary to overfitting, which is caused by the low number of patients with trigeminal nerve damage relative to those without damage. This approach does not progress through traditional recognition and classification processes, but instead accelerates classification. In the process of implementing contrastive learning, patients are divided into those with and those without sensory abnormalities, and then grouped into classes based on these distinctions.

Pairs of images from patients with sensory abnormalities are combined, and pairs from patients without abnormalities are similarly grouped. A label of 1 is assigned to pairs within the same class (i.e., either all with or all without sensory abnormalities), and a label of 0 is assigned to pairs from different classes. For example, if there are 10 patients with sensory abnormalities and 10 without, rather than having only 20 total data points, the labeling results in 45 pairs within the same class and 100 pairs indicating different classes, generating a total of 145 labeled data points. This method significantly increases the amount of training data, thus mitigating issues related to limited sample sizes and enhancing the ability of the model to generalize.

C. Data Configuration

Table 1. Data configuration

Panorama	Train	Validation	Test	Total
Normal	696	22	22	740
Abnormal	116	23	23	162
CBCT-axial	Train	Validation	Test	Total
Normal	575	18	18	611
Abnormal	101	19	19	139
CBCT-cross	Train	Validation	Test	Total
Normal	575	18	18	611
Abnormal	101	19	19	139

CBCT, cone beam computed tomography.

In view of the limited number of samples in our dataset, compounded by an imbalance between patients with sensory abnormalities and those without, a larger training dataset was required. Consequently, we allocated a higher proportion of the available data to training compared with a typical model development process. The data were randomly split into training, validation, and test datasets at a distribution ratio of 9:0.5:0.5, ensuring that patients were not duplicated across sets.

Training data is utilized to educate the machine learning model, enabling it to learn data patterns and perform predictions or classifications based on these learnings. It is essential that training data include as many cases as possible to assist the model in generalizing beyond the training examples. Validation data serves to assess the performance of the model and to tune its hyperparameters, acting as an intermediate checkpoint to verify how well the model performs on unseen data.

Validation data thus plays a critical role during the training process by helping to adjust and optimize the model, ensuring that it does not overfit to the training data. Test data is used to evaluate the final performance of the model after training has concluded, testing how well the model operates on entirely new and unseen data.

Test data is not used during the model development process but is crucial in assessing how the model will perform on real-world data, providing a measure of the effectiveness of the model and its readiness for practical application.

Image Preprocessing

A. Image contrast enhancement

To enhance the contrast of images, we utilized a technique known as contrast limited adaptive histogram equalization (CLAHE) (Zuiderveld K , 1994). CLAHE is a method that improves image contrast and is particularly beneficial for medical imaging or low-light images. Traditional histogram equalization distributes the histogram of the entire image uniformly, thereby improving contrast; however, this can lead to the loss of detail in brighter or darker areas.

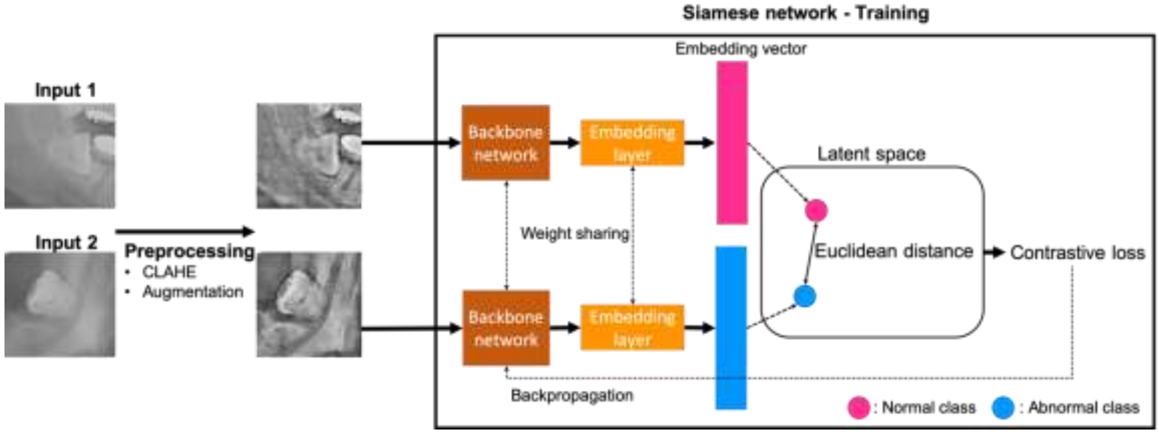
CLAHE addresses this issue by dividing the image into small blocks or tiles and applying histogram equalization independently to each block, thereby enhancing local contrast.

Additionally, CLAHE utilizes "limited contrast" to prevent the excessive application of histogram equalization that might increase noise in certain areas. We implemented CLAHE in our training dataset using the Albumentations library in Python (Buslaev A et al., 2020). To determine the optimal hyperparameters for CLAHE, we conducted a grid search analysis by adjusting the clip limit and tile grid size. The hyperparameters that were ultimately selected based on the minimization of loss in the validation dataset consisted of a clip limit of 2 to 4 and a tile grid size of (8, 8) for both panoramic and CBCT images.

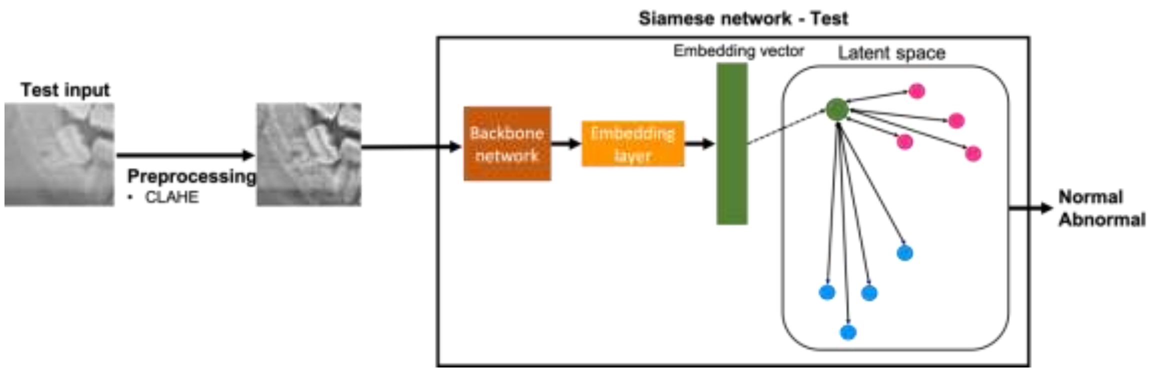
B. Image data augmentation

To address the severe imbalance between the number of patients with sensory abnormalities and those without, as well as the overall scarcity of data, we applied quantitative enhancements to our dataset. Image augmentation is crucial in the medical domain for increasing the precision of deep learning models, as it allows for the generation of diversified images from various positions and orientations, thereby creating a more generalized and robust model.

In our dataset, the number of images from patients without sensory abnormalities was approximately six times greater than that from patients with abnormalities. This discrepancy can lead to overfitting in the images of healthy patients, potentially decreasing the generalization ability and accuracy of the model as the number of images increases. In order to resolve this issue, we applied image augmentation techniques to the images of patients with neural damage in our training dataset. This integration of augmented images into the training dataset helped mitigate the problem of data imbalance. The employed image augmentation techniques included rotation, shifting, and horizontal flipping. Rotations were confined to a maximum angle of 30 degrees (rotate_limit=30), shifting occurred both horizontally and vertically (shift_limit=0.1), and flipping was only horizontal (HorizontalFlip=True).



(a) Model training process.
Two images which can belong to either the same class or different classes are utilized. These images undergo the CLAHE algorithm to enhance their contrast before being fed into the deep learning model as inputs. Subsequently, the two input images pass through the backbone network and the embedding layer, resulting in the generation of 128-dimensional embedding vectors. Through the application of contrastive learning, the deep learning model is trained to ensure that embedding vectors corresponding to images from the same class are positioned closely to each other in the latent space, whereas embedding vectors associated with images from different classes are preferentially located farther apart.
CLAHE, contrast limited adaptive histogram equalization.



(b) Model test process.
The trained deep learning model leverages the acquired knowledge of embeddings to make predictions regarding the proximity of a new test image to either the normal class or the neural damage class within the latent space.

Figure 5. Deep learning model pipeline.

Model Architecture

A. Contrastive learning

To address the issue of limited data in this study, we employed contrastive learning in our approach. Contrastive learning is a pivotal technique used in machine learning(Figure 5), particularly within the domain of self-supervised learning. This method is utilized to learn useful representations from unlabeled data primarily by comparing similarities and differences within the data, thereby aiding the model in recognizing important features. The core concept of contrastive learning involves training the model by comparing similarities and differences between data points, such that the model is trained
to bring pairs from the same category closer together and push pairs from different categories apart. This enables the model to better understand and effectively represent the intrinsic characteristics and patterns of the data.

In contrastive learning, data points are first transformed into embeddings, which are vectors mapped from higher to lower dimensions. The embedded data is then used to measure distances between points, and a contrastive loss function is employed to train the model. As a result, the model maximizes the similarities within images belonging to the normal patient class while minimizing the differences between images of different classes, such as normal patients and those with neural damage.

B. Siamese Network

To implement contrastive learning, we utilized a Siamese network (Koch G et al., 2015), which is a specialized type of neural network used in the field of deep learning. A Siamese network comprises two identical subnetworks, each sharing the same structure and weights, which facilitates the learning and comparison of similarities or relationships between two input data points. This network design focuses on processing two input images through each subnetwork to extract features then uses these features to compute the degree of similarity between the images. The ultimate output of this network is the distance between the two images, which indicates their similarity; a shorter distance suggests greater similarity.

The Siamese network is particularly valued for its ability to learn efficiently from small amounts of data, enabling it to deliver a robust performance even with limited datasets.

C. Backbone Network

In the context of Siamese networks, the backbone network refers to the primary network that extracts embeddings from input images. These embeddings are crucial for appropriately capturing essential features from the input data, which are subsequently used to measure the distance between the vectors of two input data samples within the Siamese network. Typically, convolutional neural networks (CNNs) are adopted as backbone networks because of their effectiveness in image processing tasks and their suitability for extracting prominent features from images. In this study, we evaluated five different CNN backbone models to identify the most efficient and appropriate CNN backbone model; the included models were as follows: MobileNetV2 (Sandler M et al., 2018), ResNet101D (He D et al., 2019), Vision Transformer (ViT) (Dosovitskiy A et al., 2020), Twins-svt (Chu X et al., 2021), and SSL-ResNet50 (Yalniz IZ et al., 2019).

Model Training

A. Transfer Learning

Transfer learning is a method used in machine learning and deep learning to apply or reuse models trained on a specific task for a different task. This technique transfers learned knowledge to a new problem, thereby reducing the training time for new tasks, minimizing the required data, and enhancing overall performance. To effectively train Siamese networks, we utilized a CNN backbone network trained on a large dataset known as IMAGENET (Deng J et al., 2009). Images were resized to a resolution of 224x224 px, from which high-level features were extracted. These high-level features were then processed through an embedding layer and inputted into a contrastive loss function. This approach allows for the reuse of models trained with abundant data to enhance performance when data is scarce for new tasks. Additionally, because the model has already learned basic features, the time required for further training is significantly reduced. Consequently, transfer learning can offer better performance than training a model from the start for new tasks.

B. Model Optimization

In this study, Bayesian optimization (Snoek J et al., 2012) was employed to

optimize the hyperparameters of a Siamese network. Bayesian optimization is an effective method for solving optimization problems involving functions that are relatively expensive to evaluate. This approach is particularly useful when function evaluations are costly or time-consuming, such as in hyperparameter optimization of machine learning models. Bayesian optimization builds a model of the function (typically using a Gaussian process) based on past evaluations. This model probabilistically represents the uncertainty in the function and predicts which input values are likely to return the maximum or minimum of the function. Unlike a traditional grid search or random search, Bayesian optimization uses the model outcomes to predict subsequent search locations and accordingly directs the search process. This allows for automatic adjustment of search positions based on previous results, eliminating the need for manual tuning of hyperparameters.

To achieve this goal, we utilized the Tree-structured Parzen Estimator (TPE) algorithm (Bergstra J et al., 2011) and the Asynchronous Successive Halving Algorithm (ASHA) (Li L et al., 2020) provided by the Ray package in Python (Liaw R et al., 2018) to optimize four hyperparameters.

Model Evaluation of Diagnostic Performance and Statistical Analysis

To overcome the limitation of having a disproportionately large ratio of patients with sensory abnormalities to those without any symptoms, our model utilized qualitative and quantitative data enhancements. The most suitable contrastive learning approach was applied for two-way classification, integrating transfer learning and Bayesian optimization during the training process.

To evaluate the performance of the created model, we used quantitative metrics such as precision, recall, accuracy, and F1 score. Statistical significance was assessed by presenting statistical values as means with 95% confidence intervals. A non-parametric bootstrap method was employed to calculate the 95% confidence intervals of the F1 score and accuracy. This method entailed the random resampling of the cases from the test dataset n times in 1,000 bootstrap samples, using the 2.5th to 97.5th percentiles of the bootstrap distribution to define the 95% confidence interval.

To assess the real-world efficacy of the specially employed contrastive learning and Bayesian optimization, the McNemar test was utilized. We compared the performance of deep learning models with and without the use of five backbone networks, employing contrastive learning and Bayesian optimization. The criterion for hypothesis testing was set at a significance level of 0.001.

Model Evaluation compared dentist prediction

This observer study was approved by the IRB of Yonsei University Dental Hospital. Human observation of the above-mentioned samples of 127 radiographs was performed by two oral and maxillofacial surgeons (OMFS) with over ten years of experience, two general dentists with over ten years of experience (GD-10y), and two general dentists with one year of experience (GD-1y). The image used for evaluation was examined under the same conditions (crop, size, contrast, etc.) as the image tested by the developed model. Additionally, observer evaluation was performed using a 24.1-inch liquid-crystal display (LCD) monitor (LG, Seoul, Korea) with a screen resolution of 1920x1200 px in a dark room without sunlight. Each observer looked at the provided image and predicted whether there was IAN numbness or not (class 0 = no IAN numbness, class 1 = presence of IAN numbness) without adjusting the brightness or contrast of the images or zooming in or out. Four weeks later, the same images were randomly shuffled and observer evaluation was conducted under the same conditions. Interobserver agreement between the observers within each group was also calculated.

The performance of the developed model was compared to that of a human observer in terms of prediction of IAN numbness on validation data. The recall, specificity, accuracy, precision, and F1 score were determined for each developed model and each observer. Intraobserver and interobserver agreement were determined using quadratic weighted kappa (k). Statistical analysis was performed using Excel (Microsoft, WA, USA) and SPSS 25.0 (IBM, NY, USA).

III . RESULTS

Deep Learning Model Evaluation of Diagnostic performance

Table 2. Performances of deep learning models employing contrastive learning.

	Precision	Recall	Accuracy	F1 score
MobileNetV2	0.857	1.000	0.919(0.878-0.931)	0.923(0.905-0.935)
Resnet101D	0.882	0.833	0.864(0.849-0.878)	0.857(0.845-0.870)
Vision Transformer	0.809	0.944	0.864(0.853-0.876)	0.871(0.857-0.884)
Twins-svt	0.750	1.000	0.838(0.822-0.850)	0.857(0.843-0.872)
SSL-ResNet50	1.000	0.777	0.892(0.879-0.906)	0.875(0.860-0.889)

Note: The data in parentheses are 95% confidence intervals.

To secure the most efficient CNN model, we evaluated the precision, recall, accuracy, and F1 scores of five backbone models: MobileNetV2, Resnet101D, ViT, Twins-svt, and SSL-ResNet50 (Table 2). Overall, the models demonstrated good performance with accuracy ranging from 0.838 to 0.919 and F1 scores between 0.857 and 0.923.

SSL-ResNet50 exhibited the highest precision, whereas MobileNetV2 scored the highest in recall, accuracy, and F1 score. Specifically, MobileNetV2 achieved a precision of 0.857, a recall of 1.000, an accuracy of 0.919 (confidence interval 0.878-0.931), and an F1 score of 0.923 (confidence interval 0.905-0.935). Resnet101D showed a precision of 0.882, a recall of 0.833, an accuracy of 0.864 (confidence interval 0.849-0.878), and an F1 score of 0.857 (confidence interval 0.845-0.870).

The ViT model reached a precision of 0.809, a recall of 0.944, an accuracy of 0.864 (confidence interval 0.853-0.876), and an F1 score of 0.871 (confidence interval 0.857-0.884). Twins-svt demonstrated a precision of 0.750, a recall of 1.000, an accuracy of 0.838 (confidence interval 0.822-0.850), and an F1 score of 0.857 (confidence interval 0.843-0.872). Finally, SSL-ResNet50 achieved a precision of 1.000, a recall of 0.777, an accuracy of 0.892 (confidence interval 0.879-0.906), and an F1 score of 0.875 (confidence interval 0.860-0.889)

.

Comparative Study on the Efficacy of Contrastive Learning

Table 3. F1 score improvements and corresponding P-values for various models utilizing contrastive learning (CL).

	F1 score Improvement	P-value
MobileNetV2	**144% (0.302 → 0.740)**	**P < 0.001**
Resnet101D	**266.8% (0.188 → 0.689)**	**P < 0.001**
Vision Transformer	**156.0% (0.275 → 0.704)**	**P < 0.001**
Twins-svt	**94.6% (0.370→ 0.719)**	**P < 0.001**
SSL-ResNet50	**429.4% (0.109 → 0.576)**	**P < 0.001**

Note: The data in parentheses represent the disparities between the results obtained without CL and those achieved with CL.
Improvement (%) = (With CL - Without CL) * 100/Without CL.

The data in Table 3 demonstrates substantial improvements in F1 scores across various models, suggesting the effectiveness of contrastive learning.

MobileNetV2 showed a remarkable increase of approximately 144.8%, with F1 scores rising from 0.302 to 0.740. Similarly, F1 scores in Resnet101D improved from 0.188 to 0.689, an increase of about 266.8%, and those in the ViT model rose from 0.275 to 0.704, marking an increase of 156.0%. Twins-svt also showed an improvement in F1 scores from 0.370 to 0.719, representing an increase of 94.6%.

SSL-ResNet50 displayed the most dramatic improvement, with F1 scores surging from 0.109 to 0.576, which represents an increase of 429.4%.

These results highlight the potential of contrastive learning to significantly enhance model performance in terms of F1 scores, as corroborated by statistically significant P-values ($P < 0.001$). This outcome underscores not only the effectiveness of contrastive learning techniques but also their robustness in enhancing model accuracy across diverse architectures.

Comparative Study on the Efficacy of Bayesian Optimization

Table 4. F1 score improvements and corresponding P-values for various models utilizing Bayesian optimization (BO).

	F1 score Improvement	P-value
MobileNetV2	**24.7% (0.740 → 0.923)**	**P < 0.001**

Resnet101D	**24.4% (0.689 → 0.857)**	**P < 0.001**
Vision Transformer	**23.7% (0.704 → 0.871)**	**P < 0.001**
Twins-svt	**19.2% (0.719→ 0.857)**	**P < 0.001**
SSL-ResNet50	**51.7% (0.576 → 0.875)**	**P < 0.001**

Note: The data in parentheses represent the disparities between the results obtained without BO and those achieved with BO.
Improvement (%) = (With BO - Without BO) * 100/Without BO.

Table 4 illustrates the fact that Bayesian optimization can significantly enhance the performance of various machine learning models as measured by the harmonic mean of precision and recall, i.e., the F1 score.

For MobileNetV2, the F1 score increased from 0.740 to 0.923, marking an improvement of 24.7%. This highlights the adaptability of MobileNetV2 to optimization techniques. Resnet101D improved the F1 score from 0.689 to 0.857 with an increase of 24.4%, indicating significant benefits from fine-tuning using Bayesian methods. The ViT model saw an increase in F1 score from 0.704 to 0.871, improving by 23.7%. This suggests that even newer architectures like ViT are well-suited to Bayesian optimization, despite employing mechanisms different from those of traditional CNNs. The F1 score in Twins-svt rose from 0.719 to 0.857, an improvement of 19.2%, demonstrating good scalability and potential for enhancement through optimization. SSL-ResNet50 exhibited a dramatic increase in F1 score from 0.576 to 0.875, a rise of 51.7%. This dramatic improvement underscores the hidden potential of semi-supervised learning (SSL) models when properly optimized.

It is crucial to emphasize that the P-values for these improvements are all less than 0.001, confirming that the observed enhancements are not the result of random fluctuations but are genuine effects of the Bayesian optimization process. This underscores the potential of Bayesian optimization not only to fine-tune model parameters effectively but also to significantly enhance model performance across various architectures.

Comparison of model predictions and dentist predictions

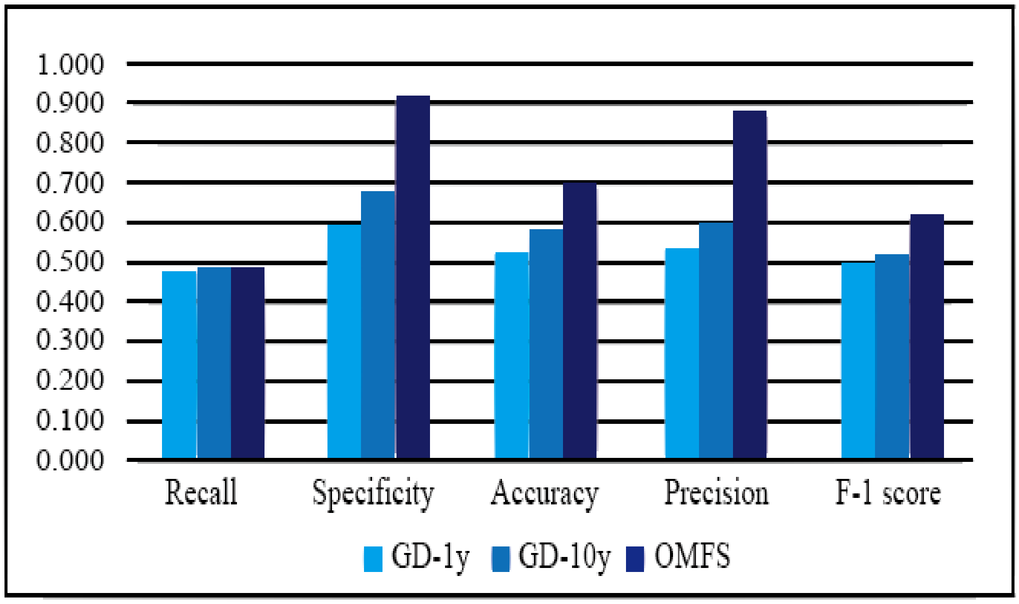


Figure 6. Test results based on dentist's clinical experience
The higher the dentist's clinical experience, the higher the results in all outcomes.

Table 5. Comparison of the prediction accuracy of dentists and AI models.

	Two dentists with one year of experience		Two dentists with over 10 years of experience		Two specialists in oral and maxillofacial surgery		AI model
	1st	2nd	1st	2nd	1st	2nd	
Score	138	145	143	156	179	160	0.838 ~0.919
Accuracy	0.5433	0.5709	0.5630	0.6142	0.7047	0.6299	

Table 5 reveals a progression in accuracy from first-year specialists to experienced specialists to oral and maxillofacial surgeons, with evaluated accuracy rates ranging from a minimum of 0.5433 to a maximum of 0.7047. This demonstrates that in our study, the AI model was significantly more accurate than experienced dentists in assessing the likelihood of sensory abnormalities following wisdom tooth extractions(Figure 6). Human accuracy increases with the experience level of the dentist but nevertheless remains consistently lower than that of our AI model.

IV. DISCUSSION

This research included patients who visited the Department of Integrated Dentistry or Oral and Maxillofacial Surgery at Yonsei University Dental Hospital between January 2018 and December 2020 for the extraction of their third molars and had pre-extraction dental radiographic images taken. The aim was to create a model capable of predicting the potential for damage to the IAN before the extraction of the lower third molar solely based on pre-surgical radiographic images, using deep learning to analyze both panoramic and CBCT radiographic images.

The extraction of impacted mandibular third molars is a common dental surgery, and IAN damage is a typical postoperative complication. Although many cases of IAN damage are temporary, some can cause total or partial permanent damage. Therefore, dentists aim to predict these risks using radiographic images such as CBCT to ensure a safer surgical process and outcome. Recently, to surpass the limitations of dentists in predicting the preoperative risk of IAN damage, studies have been conducted on the diagnosis of impacted mandibular third molars using AI models. Since 2019, sixteen papers have been published regarding the use of radiographic image data and the application of CNN models in this context (Table 6).

In 2019, a model was developed using panoramic radiographs to accurately detect the IAN and the impacted mandibular third molar. Efforts in 2020 focused on accurately identifying the IAN and mandibular third molar in CBCT images, along with attempts to create models using panoramic images that could distinguish actual contact between the IAN and the teeth, as observed in CBCT images. Subsequent efforts were made to improve accuracy by applying various methods and different CNN models. In 2021, Kim et al. conducted research to develop a model that could autonomously and preoperatively predict IAN damage; the model used actual patient images (both with and without damage) as training data and achieved an average accuracy rate of 0.827.

Table 6. Review of previous studies

	Author / year	Journal	full name	Modality	Number of data	CNN backbone model	segmentation or labeling (0/X)	object	evaluation
1	Vinayahalingam et al.(2019)	Scientific Reports	Automated detection of third molars and mandibular nerve by deep learning	panoramic radiographs	81	U-net	o	third molar and inferior alveolar nerve	mean dice-coefficients: M3s 0.947+-0.033, IAN 0.847+-0.099
2	Jaskari et al.(2020)	Scientific Reports	Deep learning method for mandibular canal segmentation in dental cone beam computed tomography volumes	CBCT	637	3D fully convolutional neural network	o	mandibular canal	mean curve distance 0.56mm, average symmetric surface distance 0.45mm
3	Orhan et al.(2020)	J Stomatol Oral Maxillofac Surg.	Evaluation of artificial intelligence for detecting impacted third molars on cone-beam computed tomography scans	CBCT	130	U-net-like architecture	o	mandibular third molar	right detection frequecy: impacted tooth number(86.2%), root number(78.6%), canal number(68.1%)
4	Yoo et al.(2021)	Scientific Reports	Deep learning based prediction of extraction difficulty for mandibular third molars	panoramic radiographs	600 patients 1053	ResNet-34	o	prediction of extraction difficulty	accuracy: depth 78.91%, ramal relationship 82.03%, angulation 90.23%
5	Kim et al.(2021)	Diagnostics	Deep learning-based prediction of paresthesia after third molar extraction: A preliminary study	panoramic radiographs	300	SSD300, ResNet-18	x	paresthesia after third molar extraction	average accuracy 0.827
6	Zhu et al.(2021)	Diagnostics	Artificial intelligence model to detect real contact relationship between mandibular third molars and inferior alveolar nerve based on panoramic radiographs	panoramic radiographs, CBCT	503	YOLOv4(MM3-IANnet)	o	contact mandibular third molar with IAN	average precision 83.02%
7	Sukegawa et al.(2022)	Scientific Reports	Evaluation of multi-task learning in deep learning-based positioning classification of mandibular third molars	panoramic radiographs	1330	VGG16	o	classification of mandibular third molar	1) multi-task model: <accuracy> class 0.8487, position 0.8861, Winter's classification 0.8537 <precision> class 0.8541, position 0.8779, Winter's classification 0.8332 2)two-taskmulti-taskmodel:<accuracy>class0.8543,position0.8899<precision>class0.8590,position0.8814
8	Choi et al.(2022)	Scientific Reports	Artificial intelligence in positioning between mandibular third molar and inferior alveolar nerve on panoramic radiography	panoramic images based on CBCT	571	ResNet-50	o	contact mandibular third molar with IAN	accuracy: true contact positon between M3 and IAN (72.32%), bucco-lingual position between M3 and IAN (80.65%)
9	Buyuk et al.(2022)	Diagnostics	A fused deep learning architecture for the detection of the relationship between the mandibular third molar and the mandibular canal	panoramic radiographs	1880	U-net-like architecture, AlexNet	o	mandibular third molar, mandibular canal	1) The segmentation network: global accuracy 0.99, weighted intersection over union score 0.98, average dice score overall images 0.91 2)Theclassficationnetwork:accuracy0.80,perclasssensitivity0.74,0.83,0.86,0.67,perclassspecificity0.92,0.95,0.88,0.96
10	Sukegawa et al.(2022)	Scientific Reports	Deep learning model for analyzing the relationship between mandibular third molar and inferior alveolar nerve in panoramic radiography	panoramic radiographs	1279	ResNet50, ResNet50v2	o	contact mandibular third molar with IAN	1) contact analysis: accuracy 0.860, AUC 0.890 2)continuityanalysis:accuracy0.766,AUC0.843
11	Takebe et al.(2022)	Journal of Dental Science	Deep learning model for the automated evaluation of contact between the lower third molar and inferior alveolar nerve on panoramic radiography	panoramic images based on CBCT	579	YOLOv3	o	contact mandibular third molar with IAN	accuracy: 0.984 in the original data set, 0.927 in the external data set)
12	Celik(2022)	Diagnostics	Deep learning based detection tool for impacted mandibular third molar teeth	panoramic radiographs	558	Faster RCNN with ResNet50, AlexNet, VGG16, YOLOv3	o	classification of impacted third molar	average accuracy: YOLOv3 0.86, Faster RCNN-ResNet50 0.79, Faster RCNN-AlexNet 0.68, Faster RCNN-VGG16 0.7
13	Kempers et al.(2023)	Journal of Dentistry	Positional assessment of lower third molar and mandibular canal using explainable artificial intelligence	panoramic radiographs	863 patients 1444	U-net with MobileNet V2	o	classification of impacted third molar	weighted accuracy 0.951, precision 0.943
14	Casto et al.(2023)	Life	Artificial intelligence for classifying the relationship between impacted third molar and mandibular canal on panoramic radiographs	panoramic images based on CBCT	142)	ResNet-152, VGG-19	o	third molar teeth & IAN true contact	accuracy: ResNet-152: 88.86%. VGG-19: 85.28%
15	Bui et al.(2023)	International Journal of Environmental Research and Public Health	Artificial intelligence as a decision-making tool in forensic dentistry: A pilot study with I3M	panoramic radiographs	456	Mask R-CNN, U-Net	o	mandibular third molar	accuracy 95%
16	Jeon et al.(2023)	Scientific Reports	Automatic diagnosis of true proximity between the mandibular canal and the third molar on panoramic radiographs using deep learning	panoramic images based on CBCT	901	RetinaNet, YOLOv3, EfficientDet-D4	o	contact mandibular third molar with IAN	accuracy: RetinaNet 59.55%, YOLOv3: 73.03%, EfficientDet-D4: 78.65%

Previous studies have predominantly focused on marking or labeling pre-training data to facilitate classification learning, thereby enabling the inference of the likelihood of IAN damage through the spatial relationship between the mandibular third molar and the IAN. In contrast, our research aimed to create the most accurate model capable of autonomously predicting IAN damage. This goal was achieved by collecting radiographic images from patients who have experienced IAN damage and those who have not, and subsequently employing various methods to analyze the data.

To overcome the fundamental issue of data scarcity in medical models, efforts have been made to enhance the accuracy of deep learning models through three key processes:

1. Image preparation: The process involved histogram equalization using CLAHE and data augmentation to increase the size of the dataset.

2. Model architecture: The architecture was set to utilize contrastive learning, employing a Siamese network for this purpose. The shared CNN of the Siamese network employed five backbone networks (MobileNetV2, ResNet101D, ViT, Twins-svt, and SSL-ResNet50).

3. Training process: Optimization of various hyperparameters was achieved through the use of transfer learning and Bayesian optimization.

These measures led to satisfactory outcomes across the five backbone networks in terms of precision, recall, accuracy, and F1 score. The precision ranged from 0.750 to 1.000, with SSL-ResNet50 showing the highest precision, followed by ResNet101D, MobileNetV2, ViT, and Twins-svt in that order. The recall scores ranged from 0.777 to 1.000, with MobileNetV2 achieving the highest recall score, followed by Twins-svt, ViT, ResNet101D, and SSL-ResNet50 in that order. Accuracy ranged from 0.838 to 0.919, with MobileNetV2 being the most accurate model, followed by SSL-ResNet50, ResNet101D, ViT, and Twins-svt in that order; similarly, the F1 scores ranged from 0.857 to 0.923 in the same order. MobileNetV2 achieved high precision and recall, ResNet101D showed a balanced performance, ViT exhibited high recall but lower precision, Twins-svt achieved perfect recall with lower precision, and SSL-ResNet50 showed perfect precision but lower recall. The accuracy rates of these models were all much higher than those of the dentists in predicting the possibility of IAN damage.

The effectiveness of contrastive learning and Bayesian optimization was validated using the McNemar test, demonstrating significant improvements in overcoming the limitations of data scarcity. Contrastive learning improved F1 scores across all models, with improvements ranging from 94.6% to 429.4%. Similarly, Bayesian optimization increased F1 scores across all models, with increases ranging from 19.2% to 51.7%.

Unlike previous studies that focused on segmentation between the third molar and the IAN and evaluated their contact potential, this study created a model capable of predicting the possibility of IAN damage after extraction using pre-extraction radiographic images and actual clinical outcomes of IAN damage without the need for segmentation.

To address the challenge of the absolute scarcity of patients with post-extraction sensory abnormalities compared with those without, the study utilized image processing and augmentation followed by contrastive learning and Bayesian optimization to ensure predictive accuracy.

Our study had several limitations. First, when considering the complications of wisdom tooth extraction, our analysis could only account for aspects that can be reproduced through radiographic images, such as the distance to the nerve and the shape of the tooth root. Factors such as the inexperience of the surgeon, nerve damage during the extraction process, and patient characteristics such as age and sex cannot be considered in our analysis. Second, the study sample exclusively utilized data from a single institution known as the Yonsei University Dental Hospital. Therefore, the applicability of our findings and the accuracy of the predictive models in diverse clinical settings must be validated with data from private dental practices and other institutions.

A third limitation of our study was the small proportion of patients with sensory abnormalities relative to the total number of patients who underwent tooth extraction, which constrained the rate of increase in sample size. To overcome this limitation, we utilized rotation and transformation of the training data along with transfer and contrastive learning techniques. We plan to continue augmenting our dataset with additional data from patients experiencing sensory abnormalities to further our research and improve model training.

Finally, another limitation was the unavoidable human intervention in the cropping process and data determination. In the collection of panoramic and CBCT images, the cropping performed by the doctor is essential, as the computer does not perform cropping autonomously. This reliance on human judgment introduces potential variability and subjective bias in the data preparation process.

This study aims to develop a more precise predictive model by utilizing a variety of radiographic data. We plan to implement the modeling techniques derived from this research to construct the most effective model and will expand our dataset by collecting additional data from patients experiencing sensory abnormalities. This expanded dataset will help to enhance the model's accuracy. Furthermore, the ultimate goal is to integrate a capability within the AI to autonomously crop images, thereby minimizing the need for dental practitioner intervention. By achieving these milestones, we intend to create a practical model that enables dentists to proactively manage potential post-extraction complications associated with wisdom teeth.

V. CONCLUSION

The principal advantage of this study lies in the fact that it applies a single-step process to overcome the limitation of lack of data, which is a common limitation of information in the medical field.

The first key aspect of our approach was the application of qualitative improvement through histogram equalization using CLAHE in the image preparation process and quantitative data amplification using data augmentation.

Second, the model was set to contrastive learning and a Siamese network was used for this purpose, and five backbone networks were used as joint CNNs.

Third, transfer learning was used in the learning process and several hyperparameters were optimized using Bayesian optimization.

Through these efforts, the precision, recall, accuracy, and F1 scores were measured satisfactorily at a reliable level. Upon using the McNemar test to compare the results of our model to those obtained when the above-mentioned second and third steps were not performed,, an improvement of more than 100% was observed.

The accuracy of predicting IAN injury after the extraction of impacted mandibular third molars can be determined through preoperative radiographic images. According to research, the precision of using radiographic images to predict nerve damage is significantly higher than that of relying on the assessment of dental practitioners.

By applying the results of this study to a larger dataset of radiographic images and incorporating AI in data processing, a clinically applicable prediction model for sensory disturbances following mandibular third molar extraction can be developed based on preoperative radiographic images.

REFERENCES

Bergstra J, Bardenet R, Bengio Y, Kégl B. 2011. Algorithms for hyper-parameter optimization. Advances in neural information processing systems. 24.

Bui R, Iozzino R, Richert R, Roy P, Boussel L, Tafrount C, Ducret M. 2023. Artificial intelligence as a decision-making tool in forensic dentistry: a pilot study with I3M. International Journal of Environmental Research and Public Health. 20(5):4620-4632.

Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. 2020. Albumentations: Fast and flexible image augmentations. Information. 11(2):125.

Buyuk C, Akkaya N, Arsan B Unsal G, Aksoy S, Orhan K. 2022. A fused deep learning architecture for the detection of the relationship between the mandibular third molar and the mandibular canal. Diagnostics. 12(8):2018-2030.

Casalegno F, Newton T, Daher R, Abdelaziz M, Lodi-Rizzini A, Schürmann F, Krejci I, Markram H. 2019. Caries detection with near-infrared transillumination using deep learning. Journal of dental research. 98(11):1227-1233.

Celik M. 2022. Deep learning based detection tool for impacted mandibular third molar teeth. Diagnostics. 12(4):942-954.

Chen H, Zhang K, Lyu P, Li H, Zhang L, Wu J, Lee C-H. 2019. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. Scientific reports. 9(1):1-11.

Choi E, Lee S, Jeong E, Shin S, Park H, Youm S, Son Y, Pang K. 2022. Artificial intelligence in positioning between mandibular third molar and inferior alveolar nerve on panoramic radiography. Scientific Reports, 12(1):2456-2463.

Chu X, Tian Z, Wang Y, Zhang B, Ren H, Wei X, Xia H, Shen C. 2021. Twins: Revisiting the design of spatial attention in vision transformers. Advances in Neural Information Processing Systems. 34:9355-9366.

Improving segmentation of the inferior alveolar nerve through deep label propagation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022.

Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition; 2009: IEEE.

Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L. 2009. Imagenet: A large-scale hierarchical image database. IEEE conference on computer vision and pattern recognition;2009; IEEE.

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929.

Goodfellow I, Bengio Y, Courville A. 2016. Deep learning. MIT press.

Bag of tricks for image classification with convolutional neural networks. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019.

Siamese neural networks for one-shot image recognition. ICML deep learning workshop; 2015: Lille.

He D, Zhou Z, Gan C, Li F, Liu X, Li Y, Wang L, Wen S. 2019. Stnet: Local and global spatial-temporal modeling for action recognition. Proceedings of the AAAI conference on artificial intelligence;2019. 33(1):8401-8408.

Jaskari J, Sahlsten J, Järnstedt J, Mehtonen H, Karhu K, Sundqvist O, Hietanen A, Varjonen V, Mattila, V, Kaski, K. 2020. Deep learning method for mandibular canal segmentation in dental cone beam computed tomography volumes. Scientific reports. 10(1):5842-5849.

Jeon K, Choi H, Lee C, Han S. 2023. Automatic diagnosis of true proximity between the mandibular canal and the third molar on panoramic radiographs using deep learning. Scientific Reports, 13(1):22022-22029.

Jung Y, Cho B. 2013. Prevalence of missing and impacted third molars in adults aged 25 years and above. Imaging science in dentistry. 43(4):219-225.

Kempers S, van Lierop P, Hsu T, Moin D, Bergé S, Ghaeminia H, Xi T, Vinayahalingam S. 2023. Positional assessment of lower third molar and mandibular canal using explainable artificial intelligence. Journal of Dentistry.133:104519-104524.

Kim B, Yeom H, Lee J, Shin W, Yun J, Jeong S, Kang J, Kim S, Kim B. 2021. Deep learning-based prediction of paresthesia after third molar extraction: A preliminary study. Diagnostics. 11(9):1572-1582.

Kim J, Gu H, An J, Kook M, Park H, Oh H. 2006. The study of evaluation to relationship between the inferior alveolar nerve and the mandibular third molar by using radiographic image. Journal of the Korean Association of Oral and Maxillofacial Surgeons. 32(5):464-473.

Kim J, Kim H, Hoe J, Kim J, Yang H, Kim H, Hong J, Moon C, Ha T, Seo Y. 2023. Competitive study on dental fees impacted teech extraction in OECD countries and case study of domestic and international litigation. Korean dental association helath policy institute.

Koch G, Zemel R, Salakhutdinov R. 2015. Siamese neural networks for one-shot image recognition. ICML deep learning workshop;2015. 2(1).

Krizhevsky A, Sutskever I, Hinton GE. 2017. Imagenet classification with deep convolutional neural networks. Communications of the ACM. 60(6):84-90.

Krois J, Ekert T, Meinhold L, Golla T, Kharbot B, Wittemeier A, Dörfer C, Schwendicke F. 2019. Deep learning for the radiographic detection of periodontal bone loss. Scientific reports. 9(1):8495.

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. nature. 521(7553):436-444.

Lee J, Park J, Moon SY, Lee K. 2022. Automated prediction of extraction difficulty and inferior alveolar nerve injury for mandibular third molar using a deep neural network. Applied Sciences. 12(1):475.

Lee J-H, Kim D-H, Jeong S-N, Choi S-H. 2018. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. Journal of dentistry. 77:106-111.

Lee K, Kwak J, Oh J, Jha N, Kim Y, Kim W, Baik U, Ryu J. 2020. Automated detection of TMJ osteoarthritis based on artificial intelligence. Journal of dental research. 99(12):1363-1367.

Li L, Jamieson K, Rostamizadeh A, Gonina E, Ben-Tzur J, Hardt M, Recht B, Talwalkar A. 2020. A system for massively parallel hyperparameter tuning. Proceedings of Machine Learning and Systems. 2:230-246.

Liaw R, Liang E, Nishihara R, Moritz P, Gonzalez JE, Stoica I. 2018. Tune: A research platform for distributed model selection and training. arXiv preprint arXiv:180705118.

Liu M-Q, Xu Z-N, Mao W-Y, Li Y, Zhang X-H, Bai H-L, Ding P, Fu K-Y. 2022. Deep learning-based evaluation of the relationship between mandibular third molar and mandibular canal on cbct. Clinical Oral Investigations.1-11.

Lo Casto A, Spartivento G, Benfante V, Di Raimondo R, Ali M, Di Raimondo D, Tuttolomondo A, Stefano A, Yezzi A, Comelli A. 2023. Artificial Intelligence for Classifying the Relationship between Impacted Third Molar and Mandibular Canal on Panoramic Radiographs. Life. 13(7):1441-1450.

Orhan K, Bilgir E, Bayrakdar I, Ezhov M, Gusarev M, Shumilov E. 2021. Evaluation of artificial intelligence for detecting impacted third molars on cone-beam computed tomography scans. Journal of Stomatology. Oral and Maxillofacial Surgery. 122(4):333-337.

Patel PS, Shah JS, Dudhia BB, Butala PB, Jani YV, Macwan RS. 2020. Comparison of panoramic radiograph and cone beam computed tomography findings for impacted mandibular third molar root and inferior alveolar nerve canal relation. Indian Journal of Dental Research. 31(1):91.

Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.

Rood J, Shehab B. 1990. The radiological prediction of inferior alveolar nerve injury during third molar surgery. British Journal of Oral and Maxillofacial Surgery. 28(1):20-25.

Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE conference on computer vision and pattern recognition;2018. 4510-4520.

Snoek J, Larochelle H, Adams RP. 2012. Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems. 25.

Sukegawa S, Matsuyama T, Tanaka F, Hara T, Yoshii K, Yamashita K, Nakano K, Takabatake K, Kawai H, Nagatsuka H, Furuki Y. 2022. Evaluation of multi-task learning in deep learning-based positioning classification of mandibular third molars. Scientific reports. 12(1):684-693.

Sukegawa S, Tanaka F, Hara T, Yoshii K, Yamashita K, Nakano K, Takabatake K, Kawai H, Nagatsuka H, Furuki Y. 2022. Deep learning model for analyzing the relationship between mandibular third molar and inferior alveolar nerve in panoramic radiography. Scientific reports. 12(1):16925.

Suomalainen A, Ventä I, Mattila M, Turtola L, Vehmas T, Peltola J. 2010. Reliability of CBCT and other radiographic methods in preoperative evaluation of lower third molars. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology. 109(2):276-284.

Takebe K, Imai T, Kubota S, Nishimoto A, Amekawa S, Uzawa N. 2022. Deep learning model for the automated evaluation of contact between the lower third molar and inferior alveolar nerve on panoramic radiography. Journal of Dental Sciences.

Tantanapornkul W, Okouchi K, Fujiwara Y, Yamashiro M, Maruoka Y, Ohbayashi N, Kurabayashi T. 2007. A comparative study of cone-beam computed tomography and conventional panoramic radiography in assessing the topographic relationship between the mandibular canal and impacted third molars. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology. 103(2):253-259.

Vinayahalingam S, Xi T, Bergé S, Maal T, de Jong G. 2019. Automated detection of third molars and mandibular nerve by deep learning. Scientific reports. 9(1):9007.

Weiss K, Khoshgoftaar TM, Wang D. 2016. A survey of transfer learning. Journal of Big data. 3(1):1-40.

Unsupervised feature learning via non-parametric instance discrimination. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.

Yalniz IZ, Jégou H, Chen K, Paluri M, Mahajan D. 2019. Billion-scale semi-supervised learning for image classification. arXiv preprint arXiv:190500546.

Yoo J, Yeom H, Shin W, Yun J, Lee J, Jeong S, Lim H, Lee J, Kim B. 2021. Deep learning based prediction of extraction difficulty for mandibular third molars. Scientific Reports. 11(1):1954-1962.

Yu H, Cho S, Kim M, Kim W, Kim J, Choi J. 2020. Automated skeletal classification with lateral cephalometry based on artificial intelligence. Journal of dental research. 99(3):249-256.

Zhu T, Chen D, Wu F, Zhu F, Zhu H. 2021. Artificial intelligence model to detect real contact relationship between mandibular third molars and inferior alveolar nerve based on panoramic radiographs. Diagnostics. 11(9):1664-1672.

Zuiderveld K. 1994. Contrast limited adaptive histogram equalization. Graphics gems.474-485.

국문요약

하악 제3대구치 발치후 대조 학습과 Bayesian 최적화를 이용한 단일단계의 하치조신경 손상의 예상 모델 개발

연세대학교 대학원 치의학과
이 명 호
지도 교수 : 김 기 덕

최근 몇 년 동안 인공 지능(AI)과 딥 러닝은 다양한 과학 기술 분야에 혁명을 일으키며, 복잡한 문제를 해결하는 데 엄청난 잠재력을 보여주었다. 수많은 딥 러닝 애플리케이션 중에서 이미지 분류는 인상적인 정확도로 이미지 내의 개체를 식별하고 분류할 수 있는 강력한 기술로 부상했다.

구강악안면외과에서 가장 많이 시행되는 수술은 하악 제 3 대구치 발치이며, 하치조신경 손상은 일시적 또는 영구적으로 발생할 수 있는 가장 심각한 합병증 중 하나이다. 따라서 발치 전 방사선 사진 분석을 통해 발치 후 신경손상 발생을 예측하기 위한 많은 연구가 진행되고 있다. 그러나 영상을 이용한 감각이상의 가능성 평가에는 한계가 존재한다. 임상의의 경험이 적거나 지식이 부족하면 위험요인을 찾기 어려워 관찰자간 오차가 발생할 수 있다. 또한 임상의나 구강 방사선과 전문의가 아무리 경험이 많더라도 방사선 사진이 불분명하거나 여러 소견이 중복되면 오류가 발생할 가능성이 있다.

이러한 한계를 극복하기 위해 최근 인공지능을 이용한 하악 제 3 대구치와 하치조신경의 관계를 평가하는 연구가 진행되고 있다. 지금까지 대부분의 연구는 하치조신경과 하악 제 3 대구치를 검출하거나 동정하는 단계로, 발치 후 실제 신경손상의 임상자료와의 상관관계를 분석하거나 CBCT 를 이용하여 신경손상을 예측한 연구는 수행된 바가 없다. 또한 대부분의 딥러닝 연구는 감각이상의 환자와 그렇지 않은 환자의 CBCT 데이터 수의 불균형의 한계가 존재한다.

이러한 한계에 의해 딥 러닝 모델에 가장 적합한 하이퍼파라미터를 수동으로 선택하는 작업은 제한된 적은 양의 데이터로 작업할 때

광범위하고 시간이 많이 걸리게 되며, 과접합과 같은 시행착오가 생길 수 있어 정확한 모델을 생성하기 어렵다.

따라서 본 논문에서는 이러한 한계를 극복하여, 하악 사랑니 발치 시 하치조 신경의 손상 의 가능성을 평가하는 모델을 만들기 위해, *CLAHE* 를 이용한 이미지의 질적 증강과 Data augmentation 을 이용한 양적 증강을 시행하였으며, *Siamese Network* 을 이용한 대조 학습방식으로 모델을 디자인 하고, 이를 전이학습과 *Bayesian Optimization* 을 이용한 하이파라미터의 auto-tuning 을 통해 정확성을 향상시켰다.

이전까지의 연구들이 하악 제 3 대구치와 하치조신경의 segmentation 에 집중하고 그 접촉 가능성을 평가하였던 반면, 발치 전 방사선 사진과 실제 임상의 감각이상 결과를 학습하여, segmentation 의 과정 없이 단일 단계만으로 발치 술전 방사선 사진을 통한 감각 이상의 가능성을 예측 할 수 있는 모델을 제작 하였다.

우리는 현재의 연구를 확장하여, 여러가지 종류와 방법으로 수집한 다양한 방사선 사진을 이용하여 하악 사랑니 발치 전 감각이상의

가능성을 평가하는 보다 정확한 모델을 찾을 것이다. 또한 방사선 사진의 편집 과정에 인공지능의 기술을 적용하여 실제 임상에서 실용화 할 수 있는 사랑니 발치 전 하치조신경 손상의 가능성 평가 모델을 만들 수 있을 것이다.

핵심되는 말: 제 3 대구치 발치; 하치조신경; 감각이상; deep learning; CLAHE; data augmentation; 대조학습; *Siamese network;* 전이학습; Bayesian optimization;