RESEARCH ARTICLE

Classification of the Relationship Between Mandibular Third Molar and Inferior Alveolar Nerve Based on Generated Mask Images

YUNSANG JOO[1], **SEONG-YONG MOON**[2], **AND CHANG CHOI**[1], (Senior Member, IEEE)

[1]Department of Computer Engineering, Gachon University, Seongnam, Sujeong 13120, Republic of Korea
[2]Department of Oral and Maxillofacial Surgery, School of Dentistry, Chosun University, Gwangju 61452, Republic of Korea

Corresponding authors: Seong-Yong Moon (msygood@chosun.ac.kr) and Chang Choi (changchoi@gachon.ac.kr)

ABSTRACT In recent dentistry research, deep learning techniques have been employed for various tasks, including detecting and segmenting third molars and inferior alveolar nerves, as well as classifying their positional relationships. Prior studies using convolutional neural networks (CNNs) have successfully detected the adjacent area of the third molar and automatically classified the relationship between the inferior alveolar nerves. However, deep learning models have limitations in learning the diverse patterns of teeth and nerves due to variations in their shape, angle, and size across individuals. Moreover, unlike object classification, relationship classification is influenced by the proximity of teeth and nerves, making it challenging to accurately interpret the classified samples. To address these challenges, we propose a masking image-based classification system. The primary goal of this system is to enhance the classification performance of the relationship between the third molar and inferior alveolar nerve while providing diagnostic evidence to support the classification. Our proposed system operates by detecting the adjacent areas of the third molar, including the inferior alveolar nerve, in panoramic radiographs (PR). Subsequently, it generates masked images of the inferior alveolar nerve and third molar within the extracted regions of interest. Finally, it performs the classification of the relationship between the third molar and inferior alveolar nerve using these masked images. The system achieved a mean average precision (mAP) of 0.885 in detecting the region of interest in the third molar. Furthermore, the performance of the existing CNN-based positional relationship classification was evaluated using four classification models, resulting in an average accuracy of 0.795. For the segmentation task, the third molar and inferior alveolar nerve in the detected region of interest exhibited a dice similarity coefficient (DSC) of 0.961 and 0.820, respectively. Regarding the proposed masking image-based classification, it demonstrated an accuracy of 0.832, outperforming the existing method by approximately 3%, thus confirming the superiority of our proposed system.

INDEX TERMS Deep learning, radiography, dentistry, medical diagnostic imaging.

I. INTRODUCTION

Among all teeth, the mandibular third molar (M3) has the highest impaction rate, and the eruption state, position, and angle are individually different [1], [2]. Impacted M3 can cause various symptoms, such as pericoronitis, pain, swelling, root resorption of adjacent teeth, dental cysts, and tumors [3]. Due to these factors, people with impacted teeth undergo tooth extraction, dentistry's most common surgical procedure [4]. However, when extracting the impacted M3, inflammation or infection of the inferior alveolar nerve (IAN) and sensory disturbance occur [5]. IAN damage, which

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico.

causes sensory disturbances in the lower teeth, jaw, and lips, is a representative complication [6], [7]. The leading cause of this IAN corruption is the contact between the M3 root and the IAN [8], [9]. Therefore, to minimize the M3 extraction complications, understanding the relationship between the M3 impaction pattern and adjacent nerves in advance [10] is crucial.

Panoramic radiography (PR) and cone beam computed tomography (CBCT) are imaging modalities used in dentistry [11], [12], [13], [14], [15]. Recently, research has been conducted to directly perform or assist diagnosis using deep learning to detect and segment teeth automatically in PR or CBCT [16], [17], [18], [19], [20]. In particular, research is being conducted to predict the challenges of tooth extraction or the possibility of nerve damage before tooth extraction or to classify the relationship between adjacent structures of impacted teeth [21]. Yoo et al. [22] proposed a system to predict the depth, ramal relation and angle of M3 in PR. M3s were detected using a single-shot multibox detector (SSD) and pre-trained ResNet-34. They demonstrated the accuracy of predicting the position of the M3 with a convolutional neural network (CNN). Lee et al. [23] proposed a system for detecting M3 in PR and diagnosing the difficulty of extraction and the possibility of damage to the inferior alveolar nerve. They detected M3s using Retinanet and classified them using a vision transformer (ViT). Results demonstrated that deep learning could support M3 extraction diagnosis. Vollmer et al. [24] proposed a system for classifying the depth and root shape of M3s in PR. They used pre-trained VggNet16, ResNet50, Inceptionv3, EfficientNet, and MobileNetV2 with manually defined maxillary sinuses. They demonstrated that PR could be used to classify the depth and root morphology of impacted M3s.

CNN is an effective deep learning algorithm for medical image classification tasks [25], [26], [27], [28], [29]. It extracts features from the input image and performs classification. When applying CNNs to M3 pattern or IAN positional relation classification tasks, existing methods used the area around M3 as an input to a deep learning model for classification. However, because M3 and IAN have different shapes, positions, and angles for each object, the classification performance of a model that learns image features is limited. Also, the M3 pattern or position is more influenced by the proximity between adjacent structures in the M3 than by the shape of the teeth. In addition, as CNN models create probability values for each class for images, it is challenging to provide a reasonable explanation for classification. Therefore, applying an interpretable model for the classification task is necessary.

This paper proposes a classification system using masked images to improve the classification performance of M3 and IAN relationships and explain classification results. The system performs descriptive diagnoses and can be a useful diagnostic tool for dentists. The proposed system consists of three steps: region of interest (RoI) detection including M3 region, masked image generation of M3 and IAN, and

TABLE 1. Previous research on segmentation and detection of oral structures using PR and CBCT.

Reference	Detection or Segmentation	Image Type	Performance
[30]	Mandibular Third Molar	PR	0.960
[31]	Inferior Alveolar Nerve Canal	PR	0.854
[32]	Mandibular Third Molar and Inferior Alveolar Nerve	PR	0.947 0.847
[33]	Inferior Alveolar Nerve Canal	PR	0.831
[34]	Inferior Alveolar Nerve Canal	CBCT	0.774
[35]	Mandibular Third Molar	CBCT	0.862

relationship classification. The experimental results verified, the RoI detection performance, including M3, with a mean average precision (mAP) of 0.886. The dental fossa and nerve segmentation performance was verified using the extracted RoI with Dice scores of 0.961 and 0.820, respectively. The proposed system verified the classification performance with an improved accuracy of 0.833 using masked images.

The main contribution of the proposed system is to improve the classification performance of M3 and IAN relationships. It also increases the medical diagnosis accuracy and efficiency by providing explainable diagnostic results to medical professionals. For the healthcare industry, these technological advances are critical and can help improve patient treatment options while reducing healthcare costs. Additionally, these techniques can potentially expand the applications of image classification and segmentation in various domains beyond medicine.

II. RELATIVE WORKS

Recently, research has been conducted to automatically detect and segment teeth and nerves in PR or CBCT using deep learning. Table 1 lists research that detects or segments the M3 or IAN.

Celik [30] proposed a system to detect M3s in PR. The M3s were detected using Faster-RCNN and YOLOv3. The detection performance was verified with mAP@0.5 of 0.910 for Faster-RCNN and 0.960 for YOLOv3. Maheswari et al. [31] proposed a model with edge enhancement, candidate classification, and candidate pixel clustering to accurately visualize and detect the inferior alveolar nerve canal (IAC) in PR. The proposed method effectively segmented the IAC with an average dice similarity coefficient (DSC) of 0.854. Vinayahalingam et al. [32] segments M3 and IAN in PR automatically. U-net was used to segment the M3 and IAN. The mean DSC for M3 and IAN were 0.947 and 0.847, respectively, verifying that they can be used for clinical decision-making. Ariji et al. [33] constructed and evaluated a segmentation model to visualize the proximity between the impacted M3 and the IAC in PR. The IAC was segmented using U-net. The mean DSC was 0.831, validating the segmentation performance and demonstrating adequate visualization of the IAC in the PR. Lahoud et al. [34] segmented the IAC using 3D U-Net in CBCT. The mean DSC was 0.774, validating the segmentation performance and

TABLE 2. Previous research on classifying impaction patterns of the third molar or the relationship between the M3 and IAN using PR and CBCT.

Reference	Classification	Image Type	Performance
[36]	Paresthesia after Third Molar Extraction	PR	0.827
[37]	Third Molar Impaction Status	PR	0.850
[38]	Relationship between Third Molar and Inferior Alveolar Nerve	PR	0.861
[39]	Relationship between Third Molar and Inferior Alveolar Nerve	PR	0.881
[40]	Relationship between Third Molar and Inferior Alveolar Nerve Canal	PR	0.900
[41]	Relationship between Third Molar and Inferior Alveolar Nerve Canal	CBCT	0.933

demonstrating the potential of the AI tool in CBCT. Orhan et al. [35] segmented the M3 and IAC using U-net to diagnose an impacted M3 in CBCT. Performance was evaluated using Kappa analysis and validated as 0.862. It can detect impacted M3s and improve the mandible's anatomy and pathology.

Research is also being conducted to classify tooth impaction patterns or relationships between adjacent nerves using deep learning. Table 2 lists research classifying the M3 and IAN relationship.

Kim et al. [36] predicted IAN sensory impairment after M3 extraction, using CNN in PR taken before M3 extraction. They used SSD300 and ResNet18 and achieved a mean accuracy of 0.827. They confirmed that CNN could predict IAN sensory impairment after M3 extraction using PR. Maruta et al. [37] evaluated the classification performance of an automated M3 impact classification system based on machine learning in PR. They performed Pell and Gregory classification and Winter classification using VggNet16 and achieved accuracies of 0.861 and 0.843, respectively. Zhu et al. [38] developed a new detection model, YOLOv4-based MM3-IANnet, that evaluates the contact relationship between the IAN and M3 in PR. They achieved a detection accuracy of 0.830, and the collaborative approach combining dentists and MM3-IANnet showed the highest accuracy of 0.881. Sukegawa et al. [39] evaluated contact between M3 and IAC and bone continuity classification in PR. They utilized ResNet50 and achieved accuracies of 0.860 and 0.766, respectively. Fukuda et al. [40] evaluated the impact relationship between the M3 and IAC in PR and compared the performance of several CNNs. They used AlexNet, GoogLeNet, and VggNet16 and compared performance with different patch sizes, achieving AUCs ranging from 0.88 to 0.93. Liu et al. [41] developed a CNN-based deep learning approach to detect M3 and IAC in CBCT and classify their relationships. They used U-Net for M3 and IAC detection and pixel-by-pixel segmentation and ResNet-34 for relationship classification. The average DSC of M3 and MC was

0.973 and 0.925, respectively, verifying the performance of the detection model. The classification model achieved an average accuracy of 0.933, giving results similar to those of dentists.

Existing research classifies the relationship between teeth and nerves using PR or CBCT and deep learning models. However, this method is challenging to support diagnosis because it cannot provide a basis for classification and only outputs the probability value of each class [42], [43]. This paper proposes a system that classifies the IAN and M3 relationship in PR into three states: ''non-contact,'' ''partial contact,'' and ''complete contact,'' using mask images of them. This verifies the evidence for the classification result and supports the diagnosis. In addition, this paper evaluates and compares the performance of CNN-based and mask image-based classifications. Our research provides a basis for improving classification performance and assisting diagnosis by utilizing mask images to classify the relationship between teeth and nerves.

III. RELATIONSHIP CLASSIFICATION SYSTEM BASED ON GENERATED MASK IMAGES

In previous research, the M3 or IAN segmentation in PR has been used as a diagnostic aid system for dentists. In addition, various diagnostic results were output using the M3 region detected in the PR as input to the model. Classification and segmentation research using PRs have been conducted separately. The main objective of the proposed system is to extract masked images of M3 and IAN from cropped PRs and classify the relationship between them. By defining relationships using segmented masking images, the proposed system aims to provide explainable diagnostic results to dentists and patients. An outline of the proposed system is described in Figure 1.

A. MANDIBULAR THIRD MOLAR DETECTION

To facilitate the classification of the positional relationship between the M3 and IAN, the first step involves utilizing an object detection model to detect the presence of M3 in the PR images. The object detection model is employed to identify and locate the M3 object within the PR. By accurately detecting the M3, we can proceed with further analysis and classification of its relationship with the IAN. In this case, the model is trained on PR images with bounding box annotations and class information. The bounding box represents a rectangular area where the M3 object exists, while the class information indicates the type of object. The object detection model identifies the M3 object in the PR and outputs the bounding box coordinates and class information. The size of the bounding box is determined based on the location and size of the M3 object. To include the adjacent nerve, the bounding box is set to be 1.5 times larger than the tooth size. The class information corresponds to the 38th tooth, representing the lower left M3, and the 48th tooth, representing the lower right M3.
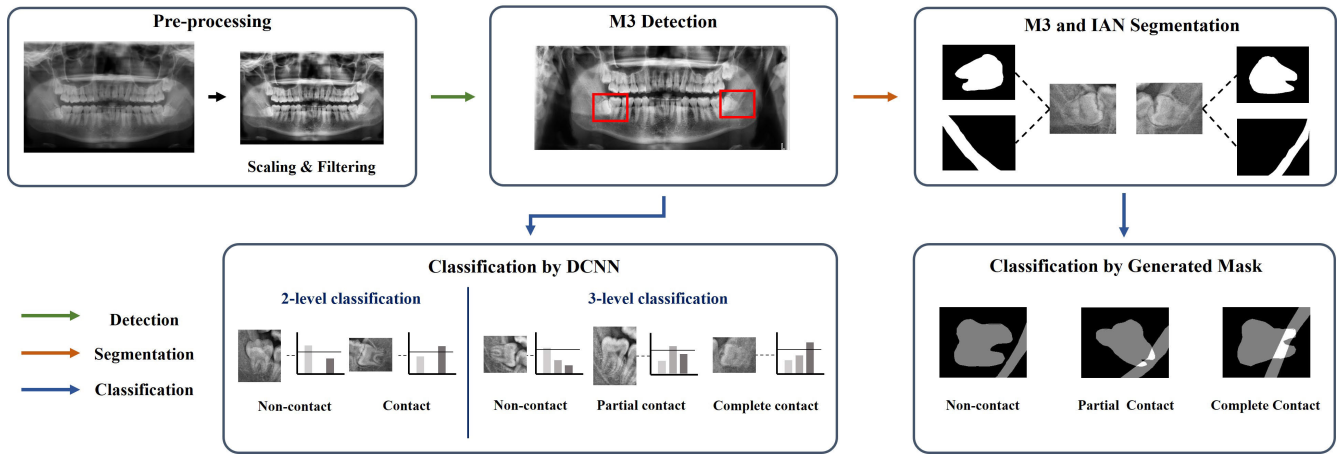
FIGURE 1. Overview of the proposed system. The proposed system detects the adjacent areas of the M3, including the IAN, in PR images. It then extracts regions of interest, generates masked images of the IAN and M3 within these regions, and classifies the relationship between the M3 and IAN using the generated masked images.

B. CLASSIFICATION OF THE RELATIONSHIP BETWEEN THE THIRD MOLAR AND THE INFERIOR ALVEOLAR NERVE

For the purpose of comparison with our proposed system, we employ a CNN-based classification model to classify the contact states between the M3 and IAN. This classification model allows us to determine and categorize the different states of contact between the M3 and IAN. By leveraging the classification model, we can evaluate its performance and compare the results with those obtained from our proposed system, thereby assessing the effectiveness of our approach. Classification involves assigning input data to predefined classes or labels and outputs probabilities for each class based on the features of the input data. In addition to binary contact classification, a 3-level classification is performed to categorize the positional relationship between the M3 and IAN into three groups: partial contact, complete contact, and non-contact. The group with complete contact poses the highest risk for nerve damage after tooth extraction.

C. MANDIBULAR THIRD MOLAR AND INFERIOR ALVEOLAR NERVE SEGMENTATION

Once the M3 region is detected, we employ an image segmentation model to perform the segmentation of the M3 and IAN based on this region. The segmentation model used in our system is a neural network model specifically designed for image segmentation tasks, structured in a U-shaped architecture consisting of encoding and decoding paths. It takes the input image and divides it into individual pixels, analyzing each pixel separately and assigning it to a corresponding class or category. In the context of M3 and IAN segmentation, the segmentation model aims to label each pixel as either belonging to the M3 region or the IAN region. To ensure efficient and accurate segmentation, we specify a Region of Interest (RoI) as the input to the segmentation model. The RoI defines a specific area within the image where the segmentation is performed. In our case, the RoI is set to be within the range

of 100 to 400 pixels, allowing the segmentation model to focus its analysis on this targeted area. During the training phase of the segmentation model, we utilize RoI and mask images obtained from the M3 and IAN as training data. The mask images represent binary masks where the M3 and IAN regions are separated from the rest of the image. By training the segmentation model with these RoI and mask image pairs, it learns to accurately segment the M3 and IAN regions in future input images. In summary, the segmentation model employed in our system utilizes a neural network architecture to perform pixel-level segmentation, dividing the input image into pixels, analyzing them individually, and assigning each pixel to the appropriate class. The model is trained using RoI and mask images, allowing it to accurately segment the M3 and IAN regions based on the detected M3 region.

D. CLASSIFICATION BASED ON MASKED IMAGES

To classify the contact states between the M3 and IAN, we utilize a predefined algorithm that operates on masked images generated by the segmentation model. These masked images are created by retaining only the regions corresponding to the teeth and nerves, while removing all other elements from the original image. The flow chart of the masked image-based classification algorithm, depicting the sequence of steps involved in the classification process, can be referred to in Figure 2. This flow chart provides a visual representation of the algorithm's workflow, illustrating how the masked images are processed and classified to determine the contact states between the M3 and IAN.

Based on the classification algorithm, the contact state between the M3 and IAN is categorized into three distinct groups: partial contact, complete contact, and non-contact. These classifications provide valuable information for dentists to assess the relationship between the M3 and IAN, aiding in the development of an appropriate surgical plan. Partial contact refers to a contact state where there are
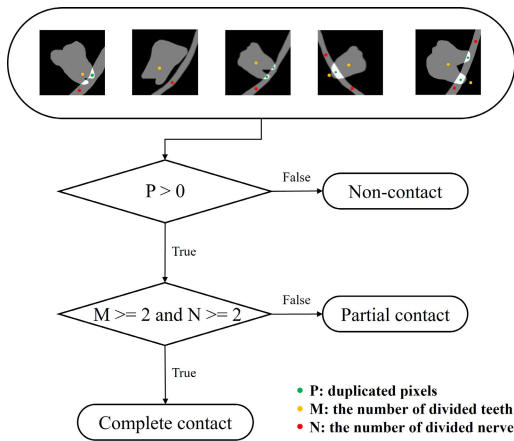
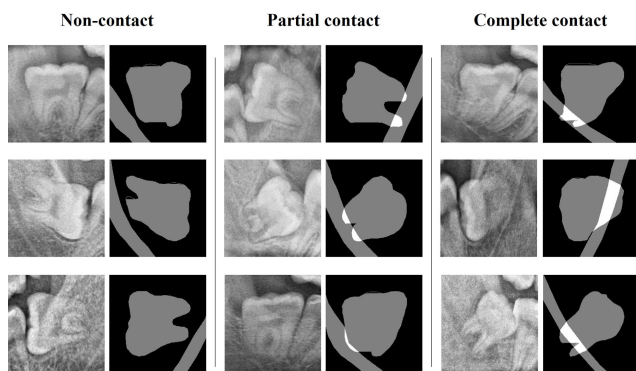**FIGURE 2.** Flow chart of the masked image-based classification algorithm.

FIGURE 3. Definition of the relationship between the M3 and the IAN.

TABLE 3. PR dataset composition.

Task	The Number of Images		
	Train	Validation	Test
Detection	4104	512	515
Segmentation	5385	1794	1797
Classification			

(GT) and the IAN as polygons. Excluding images without M3s, 8976 M3s included in 5131 images were used for training, validation, and evaluation. Contrast-Limited Adaptive Histogram Equalization (CLAHE) was applied to the image to reduce noise, improve contrast, and enable efficient model training [44]. Table 3 shows the composition of the dataset.

We verified the proposed system using an Inter Core i9 processor and NVIDIA TITAN RTX and performed experiments using Python 3.8. All experiments were trained with the training set in the data configuration table. The model with the best performance in the validation set was selected and evaluated in the test set.

The relationship between M3 and IAN was predicted using VggNet [45], ResNet [46], DenseNet [47], and ViT [48] to compare with the proposed method. These four models achieved state-of-the-art performance in the ImageNet competition. We used RoIs, including M3, and manually collected labels to train these models. For the hyperparameters of model training, we set the image size to 224×224, the batch size to 64, and the epoch to 300. We used the Adam optimizer to update the model weights.

Then, to evaluate the proposed system, we detected M3 using YOLOv7 [49], a one-stage detector mainly used for object detection tasks. The detection target is two tooth numbers 38 and 48, and the bounding box is 1.5 times the size of M3. As for the hyperparameters of model training, we set the image size to 1280×1280, the batch size to 8, and the epoch to 300. We used the adaptive moment estimation (Adam) optimizer to update the model weights.

We use Swin-Unet to perform M3 and IAN segmentation on the RoI extracted from PR [50]. Swin-Unet is a segmentation model based on a U-shaped encoder-decoder architecture with a transformer using a shifted window. It learns local-global semantic features by tokenizing images patch-by-patch. The extracted RoI and masked image are used as the model input. For model training, we set the window size to 7, the image size to 224×224, the batch size to 24, and the epochs to 150. The stochastic gradient descent (SGD) optimizer updated the model weights.

B. EVALUATION METRICS

For the evaluation of the experiment, the following metrics are used: TP represents cases where the model predicted positive and the actual class is positive. TN represents cases where the model predicted negative and the actual class is negative. FP represents cases where the model predicted positive but the actual class is negative. FN represents cases where the model predicted negative but the actual class is positive.

overlapping pixels between the M3 and IAN regions in the masked images. However, in this state, the tooth does not pass through the nerve. It indicates a proximity between the M3 and IAN, but without direct physical contact or risk of nerve damage. Complete contact occurs when the M3 tooth crosses or intersects with the IAN in the masked images, resulting in a split appearance of the tooth. This contact state poses the highest risk for potential nerve damage during tooth extraction or surgical procedures. On the other hand, the non-contact state indicates an absence of overlapping pixels between the M3 and IAN regions when the masked images are overlapped. In this state, the M3 and IAN are visually separated, indicating a safe distance between the tooth and the nerve. To provide visual examples and enhance understanding, Figure 3 showcases representative images for each contact state group.

IV. EXPERIMENTS
A. DATASET AND EXPERIMENT SETTINGS

The dataset used in this experiment consists of 5408 PR images of patients who underwent M3 tooth extraction treatment at Chosun University Dental Hospital. The PR size is about 2000-3000 pixels in width and 1000-1500 pixels in height. Seven dental specialists plotted the M3 ground truth

TABLE 4. Detection performance of the M3.

Class	Precision	Recall	mAP@.5	mAP@.5:.95
All	0.984	0.983	0.992	0.885
#38	0.985	0.982	0.991	0.880
#48	0.982	0.984	0.993	0.889

Accuracy (1), Precision (2), Recall (3), F1-score (4) and AUC (5) are used for evaluating the classification performance. Area Under the Curve (AUC) is an indicator of the area under the Receiver Operating Characteristic curve (ROC curve). The ROC curve is used to evaluate the performance of a binary classification model and represents the change in the false positive rate (FPR) against the true positive rate (TPR) while varying the threshold of the classification model.

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \quad (1)$$

$$Precision_k = \frac{TP_k}{FP_k + TP_k} \quad (2)$$

$$Recall_k = \frac{TP_k}{FN_k + TP_k} \quad (3)$$

$$F1\ score_k = 2 \cdot \frac{precision_k \cdot recall_k}{precision_k + recall_k} \quad (4)$$

$$AUC = \int TPR(FPR)\, dFPR \quad (5)$$

DSC (6), hausdorff distance(HD) (7) are used for evaluating the segmentation performance. A represents the predicted segmentation mask, while B represents the ground truth segmentation mask. In the formula, A and B represent two sets of points or regions. The function d(a, b) represents the distance between point a in set A and point b in set B.

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (6)$$

$$H(A, B) = \max(\sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(b, a)) \quad (7)$$

mAP (8) are used for evaluating the detection performance. AP is a metric that measures the precision of a retrieval system at various recall levels. It is computed for each individual class. mAP is the mean of AP values across all classes. It provides an overall measure of the system's performance.

$$mAP = \frac{1}{n} \sum_{i=1}^{n} AP_i \quad (8)$$

V. RESULTS AND ANALYSIS

We detected M3 in PR and assessed the CNN-based positional relationship between the M3 and IAN using the detected regions. Next, to evaluate the proposed system, we performed M3 and IAN segmentation within the detected regions and sequentially conducted mask-based the positional relationship between the M3 and IAN. Finally, we compared the performance of CNN-based classification with mask-based classification.

TABLE 5. Classification performance of the positional relationship between the M3 and the IAN using CNNs.

Classes	Model	Accuracy	Class	Precision	Recall	F1 score	AUC
3-class	VggNet	0.787	None	0.704	0.500	0.585	-
			Partial	0.801	0.927	0.859	
			Complete	0.771	0.428	0.550	
	ResNet	0.798	None	0.700	0.519	0.596	-
			Partial	0.818	0.921	0.867	
			Complete	0.762	0.517	0.616	
	DenseNet	0.790	None	0.661	0.522	0.584	-
			Partial	0.820	0.905	0.861	
			Complete	0.728	0.534	0.616	
	ViT	0.806	None	0.717	0.516	0.600	-
			Partial	0.822	0.922	0.869	
			Complete	0.791	0.576	0.667	
	Average (Macro)	0.795	-	0.758	0.649	0.689	-
2-class	VggNet	0.869	None	0.705	0.427	0.532	0.694
			Contact	0.888	0.962	0.924	
	ResNet	0.867	None	0.674	0.462	0.548	0.707
			Contact	0.893	0.953	0.922	
	DenseNet	0.878	None	0.725	0.487	0.583	0.724
			Contact	0.898	0.961	0.929	
	ViT	0.874	None	0.668	0.557	0.608	0.749
			Contact	0.909	0.941	0.925	
	Average (Macro)	0.872	-	0.796	0.719	0.746	0.719

Table 4 shows the result of the detection performance verification of M3 in PR. The precision and recall of the lower left M3 #38 were 0.985 and 0.982, respectively, and mAP@0.5:0.95 was confirmed at 0.880. The precision and recall of the lower right M3 #48 were 0.982 and 0.984, respectively, and mAP@0.5:0.95 was confirmed at 0.889. These results confirmed that the RoI extracted through the detection model could be sufficiently used as an input for classification and segmentation models.

The evaluation results for the classification of the M3 and IAN relationships using VggNet, ResNet, DenseNet, and ViT are presented in Table 5. For the three-class classification task, the average classification accuracy, precision, recall, and F1 score of the four classification models are 0.795, 0.758, 0.649, and 0.689, respectively. It is worth noting that the higher precision and relatively lower recall suggest the influence of class imbalance. This indicates that the models tend to predict the majority class more frequently. On the other hand, for the binary classification task, the average classification accuracy, precision, recall, F1 score and AUC of the four classification models are 0.872, 0.796, 0.719, and 0.746, 0.719, respectively. The binary classification models demonstrate higher performance compared to the multi-class classification models. Notably, the recall of 0.719 is relatively high, while the precision of 0.796 is even higher. This indicates that the models effectively identify the contact status but may misclassify some instances of the None status. Therefore, the performance of the models can vary depending on the number of samples in each class, and the values of precision and recall are crucial. It is essential to consider these metrics when evaluating the models and determining their effectiveness. Based on this analysis, there are opportunities to improve the models or implement additional measures to enhance their performance.

Figure 4 illustrates examples of misclassifications made by the CNN-based model, where samples and masks are misclassified. In the left image, a non-contact instance is
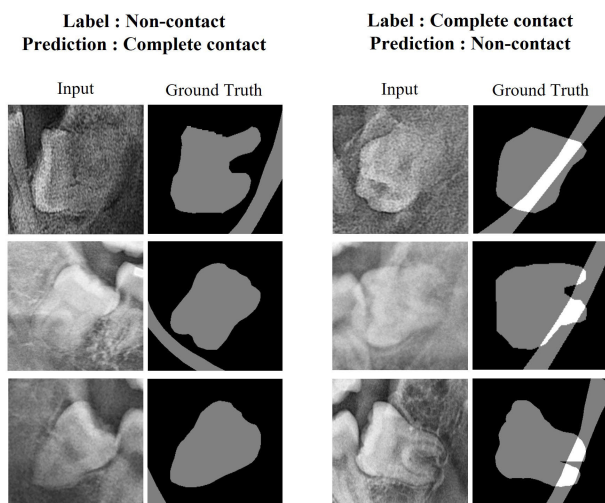
Label : Non-contact
Prediction : Complete contact

Label : Complete contact
Prediction : Non-contact

Input   Ground Truth   Input   Ground Truth

FIGURE 4. Visualization of relationship misclassification results by Convolutional Neural Networks (CNNs).

TABLE 6. Segmentation performance of the M3 and IAN.

Class	Mean DSC	Mean HD
M3	0.961	8.012
IAN	0.820	11.888

Input   Ground Truth   Prediction

FIGURE 5. Example output of Nerve and tooth segmentation from the segmentation model.

incorrectly classified as complete contact, while in the right image, a complete contact instance is erroneously classified as non-contact. The black-box nature of deep learning models makes it challenging to pinpoint the exact cause of misclassifications. There could be several factors contributing to these misclassifications, such as the shape of the tooth or nerve, or their proximity to each other. Even when the teeth and nerves are clearly separated or in contact, CNN-based models may encounter difficulties in accurately extracting the corresponding features. It is important to note that these misclassifications highlight the limitations of the CNN-based model and emphasize the need for further analysis and improvement. Understanding the underlying causes of misclassifications can potentially lead to enhancements in the model's performance and accuracy.

Table 6 presents the performance evaluation results of the M3 and IAN segmentation in the extracted Regions of Interest (RoI). M3 achieved a mean Dice Similarity Coefficient (DSC) of 0.961 and a mean Hausdorff Distance (HD) of 8.012. On the other hand, IAN achieved a mean DSC of 0.820 and a mean HD of 11.888. Since most of the nerve is on the teeth, which often overlap, the segmentation performance of IAN is relatively lower compared to M3. The mask images generated by the segmentation models are used for classification, making them closely related to the classification performance. Although there is a difference in performance between the two models, they are both sufficient for classifying the positional relationship. In this context, the positional relationship refers to whether the nerve passed
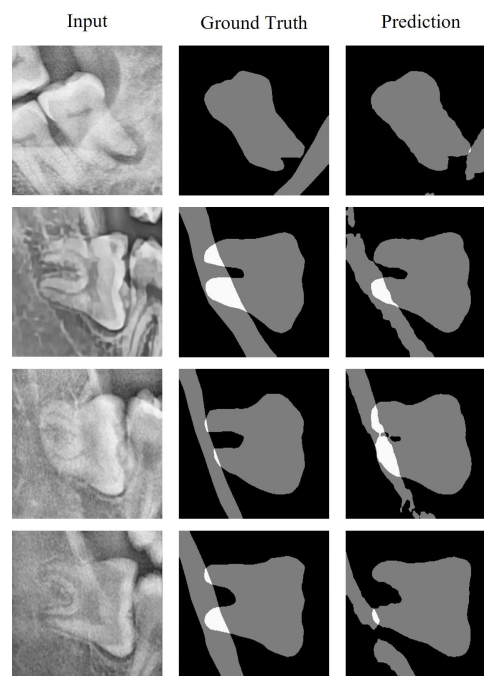
through the tooth, which is more important than achieving perfect segmentation.

Figure 5 illustrates examples of mask images generated by the segmentation models for the M3 and IAN. While the majority of the mask images are accurately generated, there are instances where the nerves appear sparse or interrupted. This can pose challenges in determining the presence of contact between the M3 and IAN.

Table 7 illustrates the performance of M3 and IAN in classifying the spatial relationships based on masked images, taking into account class imbalance. A comparison was made with the conventional CNN-based classification approach. For the 3-class classification in mask-based classification, considering the class imbalance, the accuracy, precision, recall, and F1 score were measured as 0.832, 0.829, 0.832, and 0.825, respectively. For binary classification, also considering the class imbalance, the accuracy, precision, recall, and F1 score were measured as 0.903, 0.898, 0.903, and 0.898, respectively. In both cases, the performance showed an improvement of approximately 3% compared to the CNN-based classification. These findings demonstrate that despite class imbalance and potential imperfections in generating masked images caused by IAN segmentation, the proposed approach effectively predicts the relationship between M3 and IAN.

Figure 6 displays examples of false and true positive images obtained during the classification using masked images. The proximity between objects serves as the criterion for class classification, and utilizing a segmentation model enhances the visual interpretation of the results. Most
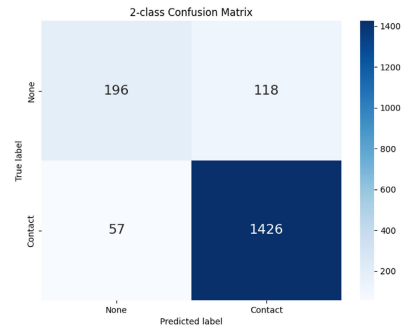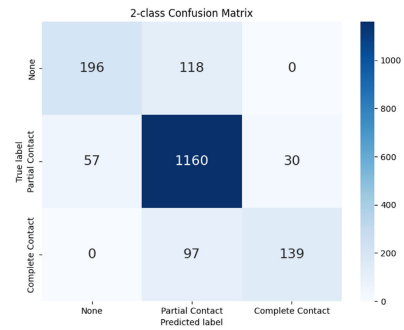
TABLE 7. Classification performance of the M3 and IAN positional relationship using the proposed system.

Classes	Method	Accuracy	Class	Precision	Recall	F1 score	AUC
3-class	CNN-based Classification	0.795	None	0.696	0.514	0.591	-
			Partial	0.815	0.919	0.864	
			Complete	0.763	0.514	0.612	
			Average	0.758	0.649	0.689	
	Mask-based Classification (proposed)	**0.832**	None	0.775	0.624	0.691	-
			Partial	0.844	0.930	0.885	
			Complete	0.822	0.589	0.686	
			Average	**0.814**	**0.714**	**0.754**	
2-class	CNN-based Classification	0.872	None	0.693	0.483	0.568	0.719
			Contact	0.897	0.954	0.925	
			Average	0.796	0.719	0.746	
	Mask-based Classification (proposed)	**0.903**	None	0.775	0.624	0.691	**0.793**
			Contact	0.924	0.962	0.942	
			Average	**0.849**	**0.793**	**0.817**	

True Positive

False Positive
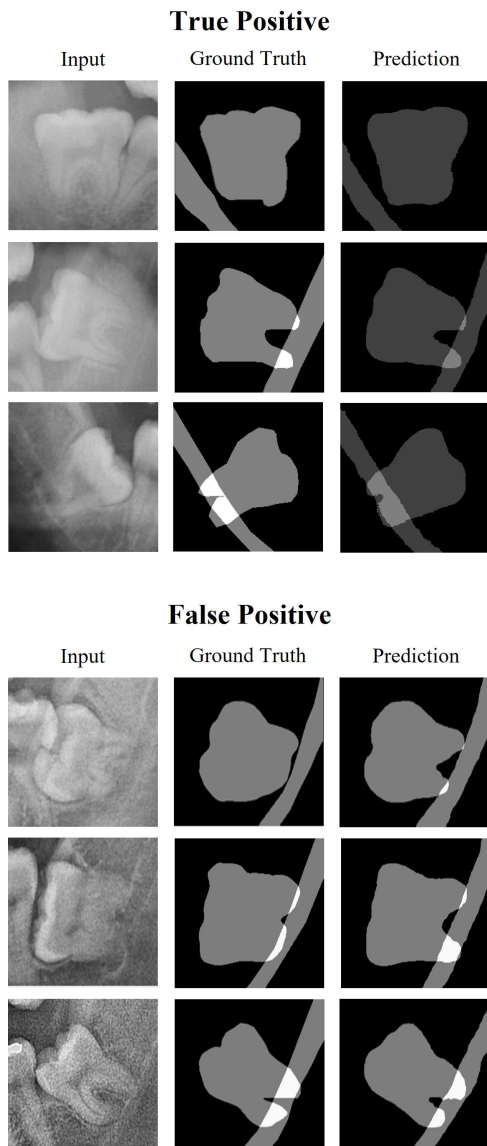
FIGURE 6. Visualization of classification results by the proposed system.

FIGURE 7. Confusion matrix of classification results by the proposed system.

misclassifications result from blurry images or incorrect segmentation of teeth and nerves. This indicates that by improving the segmentation model to generate more precise masks, it is possible to achieve a substantial enhancement in the overall classification accuracy. Thus, focusing on refining the segmentation model is crucial for attaining higher accuracy in the classification process.

The proposed system demonstrates promising performance in classifying the positional relationships between teeth and nerves based on masked images. By outperforming the conventional CNN-based classification approach by approximately 3%, the system showcases its ability to accurately predict the complex relationship between these structures. Additionally, the integration of a segmentation model allows for visual interpretation of the results, facilitating a better understanding of the proximity between objects. However, the accuracy of the segmentation model plays a critical role in generating precise masks, thereby influencing the overall classification accuracy. Figure 7 depicts the confusion matrix for the classification results of the proposed system. From the matrix, it can be observed that there is a deficiency in the number of instances classified as the non-contact class. Addressing these limitations, such as through techniques for handling class imbalance and improving the segmentation model's accuracy, will be essential for further enhancing the system's capabilities.

VI. CONCLUSION

Unlike object classification, the criterion of relationship classification is the M3 and the IAN proximity. Hence, proximity is more important than the tooth shape. Because the tooth shape and size are different for each person, the possibility of incorrect classification increases when an unfamiliar

tooth shape is entered. Deep learning models have black box characteristics making it challenging to process inputs and make predictions; hence, the causes of incorrect samples cannot be explained. We proposed a proximity classification system using M3 and IAN masked images generated by the segmentation model to solve this problem. The proposed system showed adequate detection and segmentation performance for classifying the M3 and IAN relationship in the original PR. In addition, compared to the existing classification model, the classification performance was improved by approximately 3%. Thus, even if the segmentation model does not produce a perfectly masked image, it can accurately classify proximity. However, noises in the mask and ambiguous images may lead to incorrect classification. Therefore, the segmentation model performance needs to be improved in the future, which will improve classification performance.

REFERENCES

[1] A. Hassan, "Pattern of third molar impaction in a Saudi population," *Clin., Cosmetic Investigational Dentistry*, vol. 2, pp. 109–113, Oct. 2010.

[2] M. Eshghpour, A. Nezadi, A. Moradi, R. M. Shamsabadi, N. Rezaer, and A. Nejat, "Pattern of mandibular third molar impaction: A cross-sectional study in northeast of Iran," *Nigerian J. Clin. Pract.*, vol. 17, no. 6, pp. 673–677, 2014.

[3] H. B. Polat, F. Özan, I. Kara, H. Özdemir, and S. Ay, "Prevalence of commonly found pathoses associated with mandibular impacted third molars based on panoramic radiographs in Turkish population," *Oral Surg., Oral Med., Oral Pathol., Oral Radiol., Endodontol.*, vol. 105, no. 6, pp. e41–e47, Jun. 2008.

[4] A. M. T. Oncul, D. Yazicioglu, Z. Alanoglu, S. Demiralp, A. Ozturk, and C. Ucok, "Postoperative analgesia in impacted third molar surgery: The role of preoperative diclofenac sodium, paracetamol and lornoxicam," *Med. Princ. Pract.*, vol. 20, no. 5, pp. 470–476, 2011.

[5] F. Blondeau and N. G. Daniel, "Extraction of impacted mandibular third molars: Postoperative complications and their risk factors," *J. Can. Dental Assoc.*, vol. 73, no. 4, pp. 1–4, 2007.

[6] S. Yilmaz, M. Z. Adisen, M. Misirlioglu, and S. Yorubulut, "Assessment of third molar impaction pattern and associated clinical symptoms in a central Anatolian Turkish population," *Med. Princ. Pract.*, vol. 25, no. 2, pp. 169–175, 2016.

[7] H. J. Kim, Y. J. Jo, J. S. Choi, H. J. Kim, J. Kim, and S. Y. Moon, "Anatomical risk factors of inferior alveolar nerve injury association with surgical extraction of mandibular third molar in Korean population," *Appl. Sci.*, vol. 11, no. 2, p. 816, Jan. 2021.

[8] J.-W. Kim, I.-H. Cha, S.-J. Kim, and M.-R. Kim, "Which risk factors are associated with neurosensory deficits of inferior alveolar nerve after mandibular third molar extraction?" *J. Oral Maxillofacial Surg.*, vol. 70, no. 11, pp. 2508–2514, Nov. 2012.

[9] Y. Y. Leung and L. K. Cheung, "Risk factors of neurosensory deficits in lower third molar surgery: A literature review of prospective studies," *Int. J. Oral Maxillofacial Surg.*, vol. 40, no. 1, pp. 1–10, Jan. 2011.

[10] V. R. Kumar, P. Yadav, E. Kahsu, F. Girkar, and R. Chakraborty, "Prevalence and pattern of mandibular third molar impaction in Eritrean population: A retrospective study," *J. Contemp. Dental Pract.*, vol. 18, no. 2, pp. 100–106, 2017.

[11] R. Izzetti, M. Nisi, G. Aringhieri, L. Crocetti, F. Graziani, and C. Nardi, "Basic knowledge and new advances in panoramic radiography imaging techniques: A narrative review on what dentists and radiologists should know," *Appl. Sci.*, vol. 11, no. 17, p. 7858, Aug. 2021.

[12] M. Murata, Y. Ariji, Y. Ohashi, T. Kawai, M. Fukuda, T. Funakoshi, Y. Kise, M. Nozawa, A. Katsumata, H. Fujita, and E. Ariji, "Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography," *Oral Radiol.*, vol. 35, no. 3, pp. 301–307, Sep. 2019.

[13] S. Vinayahalingam, S. Kempers, L. Limon, D. Deibel, T. Maal, M. Hanisch, S. Bergé, and T. Xi, "Classification of caries in third molars on panoramic radiographs using deep learning," *Sci. Rep.*, vol. 11, no. 1, Jun. 2021, Art. no. 12609.

[14] D.-M. Son, Y.-A. Yoon, H.-J. Kwon, C.-H. An, and S.-H. Lee, "Automatic detection of mandibular fractures in panoramic radiographs using deep learning," *Diagnostics*, vol. 11, no. 6, p. 933, May 2021.

[15] F. Umer, S. Habib, and N. Adnan, "Application of deep learning in teeth identification tasks on panoramic radiographs," *Dentomaxillofacial Radiol.*, vol. 51, no. 5, Jul. 2022, Art. no. 20210504.

[16] A. Rehman, M. Usman, R. Jawaid, A. M. Saleem, S. S. Byon, S. H. Kim, B. D. Lee, B. I. Lee, and Y. G. Shin, "Dual-stage deeply supervised attention-based convolutional neural networks for mandibular canal segmentation in CBCT scans," 2022, *arXiv:2210.03739*.

[17] Z. Jiang, Y. Chen, Y. Zhang, Y. Ge, F.-F. Yin, and L. Ren, "Augmentation of CBCT reconstructed from under-sampled projections using deep learning," *IEEE Trans. Med. Imag.*, vol. 38, no. 11, pp. 2705–2715, Nov. 2019.

[18] Z. Zheng, H. Yan, F. C. Setzer, K. J. Shi, M. Mupparapu, and J. Li, "Anatomically constrained deep learning for automating dental CBCT segmentation and lesion detection," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 603–614, Apr. 2021.

[19] V. C. B. Reia, G. de Toledo Telles-Araujo, M. Peralta-Mamani, M. R. Biancardi, C. M. F. Rubira, and I. R. F. Rubira-Bullen, "Diagnostic accuracy of CBCT compared to panoramic radiography in predicting IAN exposure: A systematic review and meta-analysis," *Clin. Oral Investigations*, vol. 25, no. 8, pp. 4721–4733, Aug. 2021.

[20] H. Wang, J. Minnema, K. J. Batenburg, T. Forouzanfar, F. J. Hu, and G. Wu, "Multiclass CBCT image segmentation for orthodontics with deep learning," *J. Dental Res.*, vol. 100, no. 9, pp. 943–949, 2021.

[21] A. F. Leite, A. V. Gerven, H. Willems, T. Beznik, P. Lahoud, H. Gaêta-Araujo, M. Vranckx, and R. Jacobs, "Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs," *Clin. Oral Invest.*, vol. 25, no. 4, pp. 2257–2267, Apr. 2021.

[22] J.-H. Yoo, H.-G. Yeom, W. Shin, J. P. Yun, J. H. Lee, S. H. Jeong, H. J. Lim, J. Lee, and B. C. Kim, "Deep learning based prediction of extraction difficulty for mandibular third molars," *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, Jan. 2021.

[23] J. Lee, J. Park, S. Y. Moon, and K. Lee, "Automated prediction of extraction difficulty and inferior alveolar nerve injury for mandibular third molar using a deep neural network," *Appl. Sci.*, vol. 12, no. 1, p. 475, Jan. 2022.

[24] A. Vollmer, M. Vollmer, G. Lang, A. Straub, A. Kübler, S. Gubik, R. C. Brands, S. Hartmann, and B. Saravi, "Performance analysis of supervised machine learning algorithms for automatized radiographical classification of maxillary third molar impaction," *Appl. Sci.*, vol. 12, no. 13, p. 6740, Jul. 2022.

[25] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *J. Big Data*, vol. 6, no. 1, pp. 1–18, Dec. 2019.

[26] M. Xin and Y. Wang, "Research on image classification model based on deep convolution neural network," *EURASIP J. Image Video Process.*, vol. 2019, no. 1, pp. 1–11, Dec. 2019.

[27] W. Wang and Y. Yang, "Development of convolutional neural network and its application in image classification: A survey," *Opt. Eng.*, vol. 58, no. 4, Apr. 2019, Art. no. 040901.

[28] P. R. Jeyaraj and E. R. S. Nadar, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," *J. Cancer Res. Clin. Oncol.*, vol. 145, no. 4, pp. 829–837, Apr. 2019.

[29] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: A survey," *Evol. Intell.*, vol. 15, no. 1, pp. 1–22, Mar. 2022.

[30] M. E. Celik, "Deep learning based detection tool for impacted mandibular third molar teeth," *Diagnostics*, vol. 12, no. 4, p. 942, Apr. 2022.

[31] P. U. Maheswari, A. Banumathi, G. Ulaganathan, and R. Yoganandha, "Inferior alveolar nerve canal segmentation by local features based neural network model," *IET Image Process.*, vol. 16, no. 3, pp. 703–716, Feb. 2022.

[32] S. Vinayahalingam, T. Xi, S. Bergé, T. Maal, and G. de Jong, "Automated detection of third molars and mandibular nerve by deep learning," *Sci. Rep.*, vol. 9, no. 1, pp. 1–7, Jun. 2019.

[33] Y. Ariji, M. Mori, M. Fukuda, A. Katsumata, and E. Ariji, "Automatic visualization of the mandibular canal in relation to an impacted mandibular third molar on panoramic radiographs using deep learning segmentation and transfer learning techniques," *Oral Surg., Oral Med., Oral Pathol. Oral Radiol.*, vol. 134, no. 6, pp. 749–757, Dec. 2022.

[34] P. Lahoud, S. Diels, L. Niclaes, S. Van Aelst, H. Willems, A. Van Gerven, M. Quirynen, and R. Jacobs, "Development and validation of a novel artificial intelligence driven tool for accurate mandibular canal segmentation on CBCT," *J. Dentistry*, vol. 116, Jan. 2022, Art. no. 103891.

[35] K. Orhan, E. Bilgir, I. S. Bayrakdar, M. Ezhov, M. Gusarev, and E. Shumilov, "Evaluation of artificial intelligence for detecting impacted third molars on cone-beam computed tomography scans," *J. Stomatol., Oral Maxillofacial Surg.*, vol. 122, no. 4, pp. 333–337, Sep. 2021.

[36] B. S. Kim, H. G. Yeom, J. H. Lee, W. S. Shin, J. P. Yun, S. H. Jeong, J. H. Kang, S. W. Kim, and B. C. Kim, "Deep learning-based prediction of paresthesia after third molar extraction: A preliminary study," *Diagnostics*, vol. 11, no. 9, p. 1572, Aug. 2021.

[37] N. Maruta, K.-I. Morita, Y. Harazono, E. Anzai, Y. Akaike, K. Yamazaki, E. Tonouchi, and T. Yoda, "Automatic machine learning-based classification of mandibular third molar impaction status," *J. Oral Maxillofacial Surg., Med., Pathol.*, vol. 35, no. 4, pp. 327–334, Jul. 2023.

[38] T. Zhu, D. Chen, F. Wu, F. Zhu, and H. Zhu, "Artificial intelligence model to detect real contact relationship between mandibular third molars and inferior alveolar nerve based on panoramic radiographs," *Diagnostics*, vol. 11, no. 9, p. 1664, Sep. 2021.

[39] S. Sukegawa, F. Tanaka, T. Hara, K. Yoshii, K. Yamashita, K. Nakano, K. Takabatake, H. Kawai, H. Nagatsuka, and Y. Furuki, "Deep learning model for analyzing the relationship between mandibular third molar and inferior alveolar nerve in panoramic radiography," *Sci. Rep.*, vol. 12, no. 1, p. 16925, 2022.

[40] M. Fukuda, Y. Ariji, Y. Kise, M. Nozawa, C. Kuwada, T. Funakoshi, C. Muramatsu, H. Fujita, A. Katsumata, and E. Ariji, "Comparison of 3 deep learning neural networks for classifying the relationship between the mandibular third molar and the mandibular canal on panoramic radiographs," *Oral Surg., Oral Med., Oral Pathol. Oral Radiol.*, vol. 130, no. 3, pp. 336–343, Sep. 2020.

[41] M.-Q. Liu, Z.-N. Xu, W.-Y. Mao, Y. Li, X.-H. Zhang, H.-L. Bai, P. Ding, and K.-Y. Fu, "Deep learning-based evaluation of the relationship between mandibular third molar and mandibular canal on CBCT," *Clin. Oral Invest.*, vol. 26, no. 1, pp. 981–991, Jan. 2022.

[42] D. T. Huff, A. J. Weisman, and R. Jeraj, "Interpretation and visualization techniques for deep learning models in medical imaging," *Phys. Med. Biol.*, vol. 66, no. 4, Feb. 2021, Art. no. 04TR01.

[43] A. M. Groen, R. Kraan, S. F. Amirkhan, J. G. Daams, and M. Maas, "A systematic review on the use of explainability in deep learning systems for computer aided diagnosis in radiology: Limited use of explainable AI?" *Eur. J. Radiol.*, vol. 157, Dec. 2022, Art. no. 110592.

[44] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *J. VLSI Signal Process.-Syst. Signal, Image, Video Technol.*, vol. 38, no. 1, pp. 35–44, Aug. 2004.

[45] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[47] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[48] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[49] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.

[50] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-UNet: UNet-like pure transformer for medical image segmentation," 2021, *arXiv:2105.05537*.
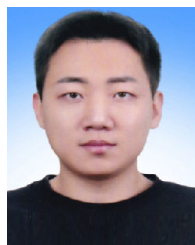
YUNSANG JOO received the B.S. degree in computer engineering from Gachon University, Seongnam, South Korea, in 2022, where he is currently pursuing the M.S. degree. His current research interests include machine learning, artificial intelligence, computer vision, and medical imaging analysis.

SEONG-YONG MOON received the B.S. and M.S. degrees in dentistry from Chosun University, in 2002 and 2006, respectively, and the Ph.D. degree from Chonnam National University, in 2013. He has been a Professor with Chosun University, since 2007. His current research interests include oral and maxillofacial surgery, oral cancer and reconstruction, AI diagnosis and simulation, and VR/AR in the medical field. He has served as a Guest Editor for *Applied Sciences*.

CHANG CHOI (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer engineering from Chosun University, in 2005, 2007, and 2012, respectively. He is currently an Assistant Professor with Gachon University. He serves with the organizing or program committees of international conferences and workshops, such as ACM RACS, EAI BDTA, IE, ACM SAC, and the IEEE CCNC/SeCHID. He has authored more than 50 publications, including articles in prestigious journal/conferences, such as the *IEEE Communications Magazine*, the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING, the IEEE INTERNET OF THINGS JOURNAL, *Information Sciences*, and *Future Generation Computer Systems*. His current research interests include intelligent information processing, semantic web, the smart IoT systems, and intelligent system security. He received the academic awards with the Graduate School, Chosun University, in 2012, and the Korean Government Scholarship for Graduate Students (Ph.D. course), in 2008. He has served as a Guest Editor for high-profile journals, such as the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, *Future Generation Computer Systems*, *Applied Soft Computing*, *Multimedia Tools and Applications*, *Journal of Ambient Intelligence and Humanized Computing*, *Concurrency and Computation: Practice and Experience*, *Sensors*, and *Autosoft*.

・・・