scientific reports

OPEN

Artificial intelligence-enhanced diagnosis of degenerative joint disease using temporomandibular joint panoramic radiography and joint noise data

Eunhye Choi[1], Seokwon Shin[2], Kijin Lee[2], Taejin An[2], Richard K. Lee[3], Sunmin Kim[4], Youngdoo Son[2]✉ & Seong Teak Kim[5]✉

This study aimed to develop an artificial intelligence (AI) model for the screening of degenerative joint disease (DJD) using temporomandibular joint (TMJ) panoramic radiography and joint noise data. A total of 2631 TMJ panoramic images were collected, resulting in a final dataset of 3908 images (2127 normal (N) and 1781 DJD (D)) after excluding indeterminate cases and errors. AI models using GoogleNet were evaluated with six different combinations of image data, clinician-detected crepitus, and patient-reported joint noise. The model that integrated all joint noise data with imaging demonstrated the highest performance, achieving an F1-score of 0.72. Another model, which incorporated both imaging and crepitus, also achieved the same F1-score but had lower D recall (0.55 vs. 0.67) and N precision (0.71 vs. 0.74). The AI models outperformed orofacial pain specialists when provided with imaging alone or in combination with all joint noise data. These findings suggest that AI-enhanced DJD diagnosis using TMJ panoramic radiography and joint noise data offers a promising approach for early detection and improved patient care. The results underscore AI's capability to integrate diverse diagnostic factors, providing a comprehensive and accurate assessment that surpasses traditional methods.

Keywords Temporomandibular joint, Degenerative joint disease, Artificial intelligence, Temporomandibular joint panoramic radiography, Joint noise

Degenerative joint disease (DJD) of the temporomandibular joint (TMJ) is a degenerative condition characterized by cartilage breakdown and bone deformation in the mandibular condyle and articular fossa, significantly affecting patients' quality of life by impairing function and aesthetics[1]. Interestingly, while DJD is typically considered female predominant and age-related, the linear correlation between bone change and age observed in other joints is not evident in the TMJ. Temporomandibular disorder (TMD) is most prevalent in young adults aged 20 to 40, and DJD may begin at an early age[2]. Notably, the severity of DJD does not always correspond with pain intensity; DJD changes in young Korean TMD patients are as common as in older individuals, showing no correlation with clinical pain or disc displacement[3]. This makes early detection both challenging and essential, especially since significant degeneration often occurs without overt symptoms.

Cone-beam computed tomography (CBCT) is the reference standard for diagnosing DJD due to its high diagnostic accuracy and detailed anatomical assessment[4]. However, CBCT is costly and exposes patients to higher radiation doses compared to other imaging modalities. While panoramic radiography is more accessible and cost-effective, it lacks comprehensive evaluation capabilities for the TMJ region. Conversely, TMJ panoramic radiography provides a more targeted evaluation, enabling assessment of the mandibular condyle's range of motion without significant projection limitations.

[1]School of Dentistry, Dental Research Institute, Seoul National University, Seoul, Republic of Korea. [2]Department of Industrial and Systems Engineering, Dongguk University, 30 Pildong-ro 1-gil, Jung-gu, Seoul 04620, Republic of Korea. [3]School of Biomedical Engineering, University of Technology Sydney, Sydney, Australia. [4]College of Arts and Sciences, Boston University, Boston, USA. [5]Department of Orofacial Pain and Oral Medicine, Yonsei University College of Dentistry, 50-1, Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea. ✉email: youngdoo@dongguk.edu; K8756050@yuhs.ac

Model	Res18	Res50	Res101	VGG16	VGG19	GoogleNet
1	0.67	0.69	0.68	0.69	0.69	0.70
2	0.71	0.71	0.70	0.71	0.71	0.72
5	0.69	0.68	0.64	0.67	0.69	0.71
6	0.70	0.70	0.66	0.71	0.69	0.72

Table 1. Comparison of F1-scores among various pre-trained models for DJD diagnosis. DJD, degenerative joint disease.

Model		Precision	Recall	F1-score	Accuracy	Macro average F1-score
1	True = D	0.78	0.53	0.63	0.72	0.70
	True = N	0.69	0.88	0.77		
2	True = D	0.81	0.55	0.66	0.74	0.72
	True = N	0.71	0.89	0.79		
3	True = D	0.69	0.63	0.66	0.70	0.70
	True = N	0.71	0.76	0.74		
4	True = D	0.71	0.67	0.69	0.72	0.72
	True = N	0.74	0.77	0.75		
5	True = D	0.69	0.61	0.65	0.69	0.69
	True = N	0.70	0.77	0.73		
6	True = D	0.74	0.58	0.65	0.72	0.71
	True = N	0.71	0.83	0.76		

Table 2. Detailed performance metrics for different AI models in DJD diagnosis. AI, artificial intelligence; DJD, degenerative joint disease; D, degenerative joint disease; N, normal.

Recent a systematic review and meta-analysis of artificial intelligence (AI) studies using radiography for DJD diagnosis reported a pooled sensitivity of 0.76 and specificity of 0.79, based on a total of 10,077 TMJ images from seven studies. While three of these studies focused on panoramic radiography and various transfer learning models, others investigated the 3D shape of the condyle and disease classification using CBCT images[5]. Most existing AI research on DJD diagnosis has relied primarily on single imaging modalities with little incorporation of clinical data. Although the value of joint noise data in TMD diagnosis is not always definitive, some studies have underscored its importance. For instance, click sounds have been found to be more strongly associated with disk displacement without reduction[6], and coarse crepitus has been identified as a useful clinical indicator for DJD[7]. The Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) also recommends utilizing all joint noise data in diagnosing DJD[8].

This study aimed to develop and evaluate an AI model that integrates TMJ panoramic radiography and joint noise data to improve the screening and diagnosis of DJD. By aligning with current diagnostic criteria and addressing the limitations of previous AI studies that relied solely on imaging data, this approach offers the potential for a more comprehensive and accurate diagnostic tool for DJD. Our study uniquely integrates imaging data with clinical joint noise information, combining both clinician-detected crepitus and patient-reported joint noise with diagnostic labeling based on TMJ panoramic radiography and CBCT analysis. This multimodal approach aims to improve diagnostic accuracy and provide a more comprehensive tool for DJD diagnosis.

Results

Among the six AI models evaluated, GoogleNet consistently demonstrated superior performance, as evidenced by the highest F1-score across all conditions, regardless of the presence or type of joint noise data (Table 1). In the assessment of models incorporating various joint noise data inputs using GoogleNet, all experiments were conducted in triplicate to ensure reliability, with mean values utilized for analysis. The integration of crepitus data resulted in an overall performance enhancement (Model 1 vs. 2). Conversely, the inclusion of subjective joint noise data exhibited a stabilizing effect on label-specific predictions (Model 1 vs. 4). Model 2 and 4 achieved the highest F1-score (0.72) (Table 2, Fig. 1).

In selecting the best model, we focused on DJD (D) recall (sensitivity) and normal (N) precision (positive predictive value) as key metrics, given their importance in medical diagnostics, especially for screening tools. D recall represents the proportion of actual DJD cases correctly identified by the model. High D recall is crucial for a screening tool because it minimizes false negatives, ensuring that fewer cases of DJD are missed. This is particularly important in early detection and intervention of DJD. N precision, on the other hand, indicates the model's accuracy in classifying normal cases. High N precision means that when the model classifies a case as normal, there's a high probability that it is indeed normal. This is important for reducing unnecessary further testing or treatment for patients who don't have DJD. By prioritizing these metrics, we aimed to develop a screening tool that effectively identifies potential DJD cases (high D recall) while also accurately ruling out

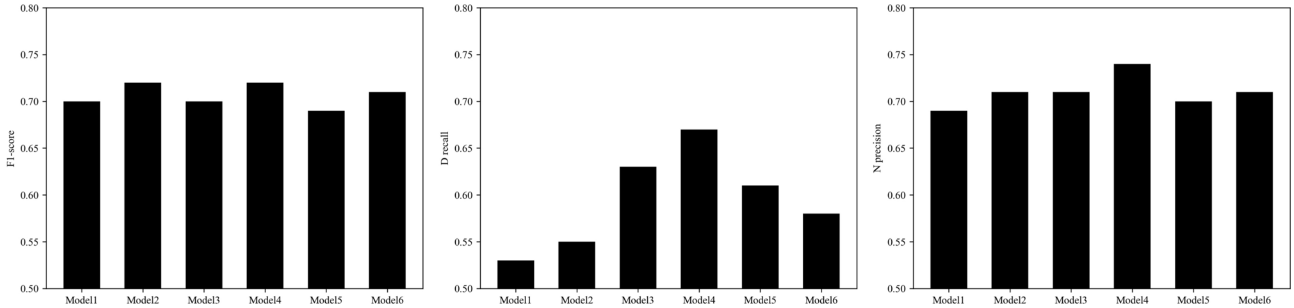

Fig. 1. Performance comparison of six AI models in diagnosing DJD. AI, artificial intelligence; DJD, degenerative joint disease; D, degenerative joint disease; N, normal.

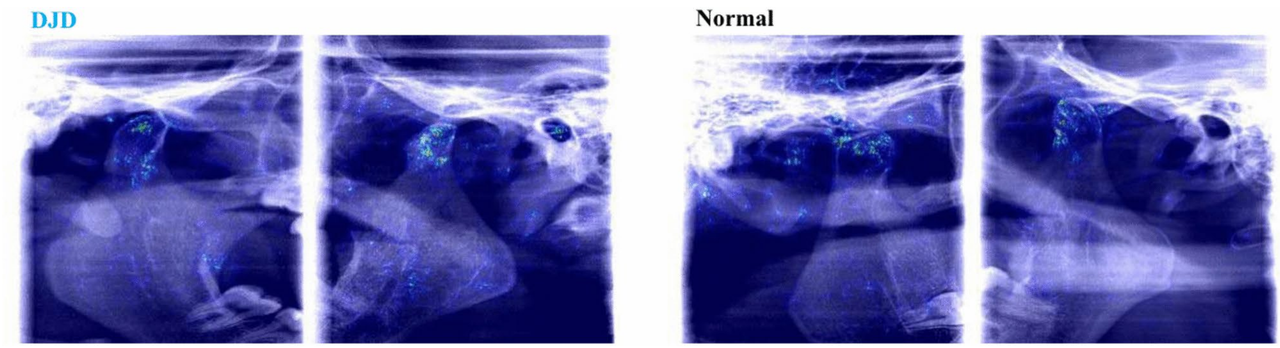

Fig. 2. Saliency map highlighting the region of interest in AI-based DJD diagnosis. AI, artificial intelligence; DJD, degenerative joint disease.

Model		Precision	Recall	F1-score	Accuracy	Macro average F1-score
1	True = D	0.69	0.50	0.58	0.64	0.63
	True = N	0.61	0.77	0.68		
4	True = D	0.65	0.44	0.52	0.60	0.59
	True = N	0.57	0.76	0.66		

Table 3. Average diagnostic performance of eight orofacial pain specialists in DJD diagnosis. DJD, degenerative joint disease; D, degenerative joint disease; N, normal.

normal cases (high N precision). This approach helps to balance the need for early detection with the goal of minimizing unnecessary interventions or further diagnostic procedures. The ranking of model suitability, calculated by summing the ranks of macro average F1, D recall, and N precision, yielded the following order: 4, 2, 6, 1, 3, 5.

To identify the image areas that the AI model relied on for its judgment, we utilized a saliency map, as shown in Fig. 2.

Comparative analysis between the AI models and orofacial pain specialists was shown in Table 3 and Fig. 3. AI models outperformed specialists across all metrics (precision, recall, F1-score) for both Model 1 (image-only) and Model 4 (image and whole joint noise data). The AI's diagnostic capability improved with the addition of joint noise data (Model 4), as evidenced by the increase in macro average F1-score from 0.70 (Model 1) to 0.72 (Model 4). Conversely, specialists showed a decline in performance when presented with both image and joint noise data (Model 4) compared to image-only assessment (Model 1), with macro average F1-scores decreasing from 0.63 to 0.59.

Discussion

Our study reveals important findings regarding the diagnosis of DJD using AI models and the significance of various diagnostic factors, particularly joint noise.

The most notable finding is that combining both subjective and objective noise data (Model 4) yielded the highest performance, emphasizing the importance of using all available auditory information. This aligns with the DC/TMD for diagnosing DJD, which now includes both fine and coarse crepitus, as well as patient-reported

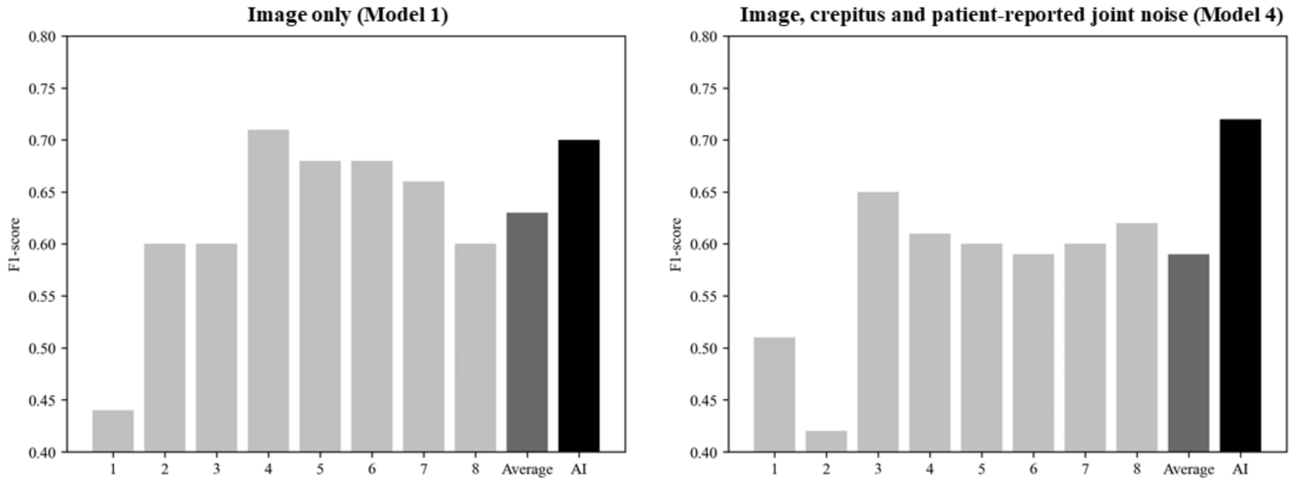

Fig. 3. Diagnostic performance comparison: specialists versus AI models. AI, artificial intelligence.

joint noises[8]. However, for DJD diagnosis, imaging such as CBCT remains essential due to the low sensitivity (0.55) and specificity (0.61) when relying solely on clinical exams without imaging.

While integrating all types of noise data proved most effective, using only crepitus also enhanced model performance. Crepitus, a well-established marker in DJD diagnostics, has shown high specificity in previous studies[9]. Our findings corroborate this, as models incorporating crepitus (Models 2, 4, and 6) outperformed those relying solely on image data (Model 1).

Interestingly, while our AI models, particularly those integrating both imaging and noise data, outperformed orofacial pain specialists, it is important to consider the context of this comparison. Specialists typically rely on a broader range of diagnostic information during clinical evaluations, including patient history, mandibular movements, muscle and joint palpation, joint sounds, and occlusion, which were not available in this study. The AI models, on the other hand, were limited to analyzing only the panoramic images and noise data. This limitation may explain why the AI models performed better under these conditions.

In diagnosing DJD, research findings are mixed regarding the effectiveness of TMJ panoramic radiography compared to panoramic radiography. TMJ panoramic radiography has been shown to have limited but improved diagnostic accuracy over conventional panoramic radiography, particularly in detecting bony lesions like flattening, erosion, and osteophytes on the mandibular condyle[10]. The combination of lateral and frontal TMJ projections often shows the highest sensitivity for detecting these lesions, though with varying levels of specificity and overall accuracy. In contrast, other studies have reported that general panoramic radiography tends to have better diagnostic accuracy than TMJ panoramic projections, especially when evaluating condylar cortical erosion[11]. The broader context provided by general panoramic radiography may aid in more precise assessments. Despite the promising results from our AI models, their performance was lower than expected compared to studies using general panoramic radiography for DJD diagnosis. However, our study focused on detecting deformation due to subcortical cysts, surface erosion, osteophytes, and generalized sclerosis, which may account for some differences in findings. Also, the challenges associated with TMJ panoramic radiography—such as movement artifacts due to open-mouth positioning and the potential for obscuring key joint structures—likely contributed to the reduced effectiveness observed in our results.

Despite the promising results, this study has several limitations that should be considered for future research. Although our dataset was labeled based on CBCT interpretations by multiple dental radiology specialists, relying on the CBCT images themselves for labeling might improve the precision and consistency of the AI models. Additionally, the incomplete collection of subjective noise data, as not all cases included this information, may have limited the accuracy and generalizability of our findings. Finally, the study's limited sample size and single-center design suggests the need for further research that includes more comprehensive datasets and multi-center data to validate and refine the AI models for broader clinical application.

In conclusion, our study demonstrates that AI-enhanced DJD diagnosis, particularly through the integration of panoramic radiography and joint noise data, offers a promising approach for early detection and improved patient care. The development of this AI-driven diagnostic tool has significant implications for public health and general practice. By combining these modalities, the model provides a practical solution for non-specialist clinicians, enabling more accurate screening and timely referrals of TMD patients. This tool has the potential to reduce diagnostic delays, improve patient outcomes, and ease the burden on specialized healthcare services. However, further research is needed to optimize this approach, particularly with respect to evolving diagnostic criteria and the specific challenges of TMJ imaging.

Methods

Study design and data collection

This study was approved by the Institutional Review Board (IRB) of Yonsei University College of Dentistry (IRB 2-2024-0011), and all methods were performed in accordance with relevant guidelines and regulations. The IRB waived the requirement for informed consent documentation.

Radiographic images were reviewed from patients who visited the Orofacial Pain Clinic at Yonsei University Dental Hospital between January 2019 and February 2021, reported TMD-related symptoms, and had both TMJ panoramic radiography (Rayscan Alpha Plus, Ray Co. Ltd., Hwaseong-si, Korea) and TMJ CBCT (Alphard 3030 device, Asahi Roentgen Ind., Co. Ltd, Kyoto, Japan) evaluated by oral and maxillofacial radiology specialists. Exclusion criteria included patients under 18 years of age, those with a history of orthognathic surgery, macro trauma, systemic diseases causing joint deformity, or a time gap of more than three months between the TMJ panorama and CBCT imaging.

For the collection of joint noise data, both objective (clinician-detected crepitus) and subjective noise reported by patients were reviewed retrospectively through electronic medical records (EMR). Crepitus was defined as a continuous or multiple coarse friction sound/sensation felt by the examiner's finger placed anterior to the external auditory meatus during three repetitions of opening and closing movements, as well as during lateral and protrusive movements of the mandible. This objective assessment was performed by trained clinicians as part of the standard TMJ examination protocol. Patient-reported joint noise were categorized into four types:

P: clicking or popping joint noise
R: crepitus or grating joint noise like gravel or sand moving
PR: both joint noises present
I: unspecified joint noise (patient reports a joint noise but cannot accurately describe its nature)

Data processing

A total of 2631 TMJ panoramic images were collected for this study. While there was only a slight difference in AI model performance between open and closed TMJ panoramic images, with the open images showing a 2–3% higher performance, this study utilized only open ones (Supplemental Table). Each image was re-labeled by an orofacial pain specialist as either "normal (N)" or "DJD (D)" based on corresponding CBCT analysis criteria for temporomandibular disorder diagnosis[8,12]. These criteria included deformation due to subcortical cyst, surface erosion, osteophyte, or generalized sclerosis. CBCT was chosen due to its high-resolution capability in detecting bony changes, which were used as the reference standard for labeling TMJ panoramic images. Following the exclusion of 1354 indeterminate DJD images and errors[13], the final dataset included 2127 images labeled as "N" and 1781 images labeled as "D".

Crepitus data was collected for the entire dataset, while subjective noise data was collected for a subset of samples (48.13% of the total).

AI model development

The AI model employed pre-trained CNN architectures to extract feature vectors, which were then combined with joint noise information for final diagnosis. We compared the performance of several pre-trained models commonly used for image classification tasks, including ResNet, VGGNet, and GoogleNet.

Data preprocessing involved a comprehensive approach to optimize image quality and enhance the robustness of the prediction model. Initially, images were resized while maintaining the original aspect ratio, using three different scales: 1/4, 1/2, and the full original size. To further improve the model's performance, a two-stage preprocessing pipeline was implemented, consisting of auto-cropping and data augmentation. In the first stage, auto-cropping was applied to eliminate unnecessary background, segmenting each image into four distinct sub-images: right-close, right-open, left-open, and left-close views. Only the open images were selected, ensuring that the analysis focused solely on the relevant regions of interest. Subsequently, various data augmentation techniques were utilized to increase the diversity of the training set. These included random rotation (± 15 degrees) to account for positional variability, color jittering to simulate changes in contrast and brightness, random sharpness adjustments (with a probability of 0.5) to handle differences in image clarity, and horizontal flipping (with a probability of 0.5) to increase the model's robustness to left–right orientation changes. By introducing controlled variability, these augmentations aimed to enhance the model's generalization capabilities, reducing the risk of overfitting. The effects of each preprocessing technique are illustrated in Fig. 4.

To evaluate the effectiveness of sound information in the AI model, we incorporated joint noise information into the feature vector. We recorded objective noise (clinician-detected crepitus) and subjective noise reported by patients for each image. While crepitus data was collected for the entire dataset, subjective noise data was collected for 48.13% of the samples. Both crepitus and three types of patient-reported subjective sounds were binarized and encoded as four-dimensional vectors, with 1 indicating the presence of a sound and 0 indicating its absence. The integration of feature vectors and sound information was achieved by merging binarized sound information with the feature vectors derived from the TMJ panoramic images through the pre-trained models. A single-layer neural network used the merged vectors to perform the final diagnosis. To improve the performance of the AI model, experiments were conducted on six models by varying their inputs. The six models were:

Model 1: image only
Model 2: image and crepitus
Model 3: image and patient-reported joint noise (P, R, PR, I)
Model 4: image, crepitus andpatient-reported joint noise (P, R, PR, I)
Model 5: image and certain patient-reported joint noise (R, PR)

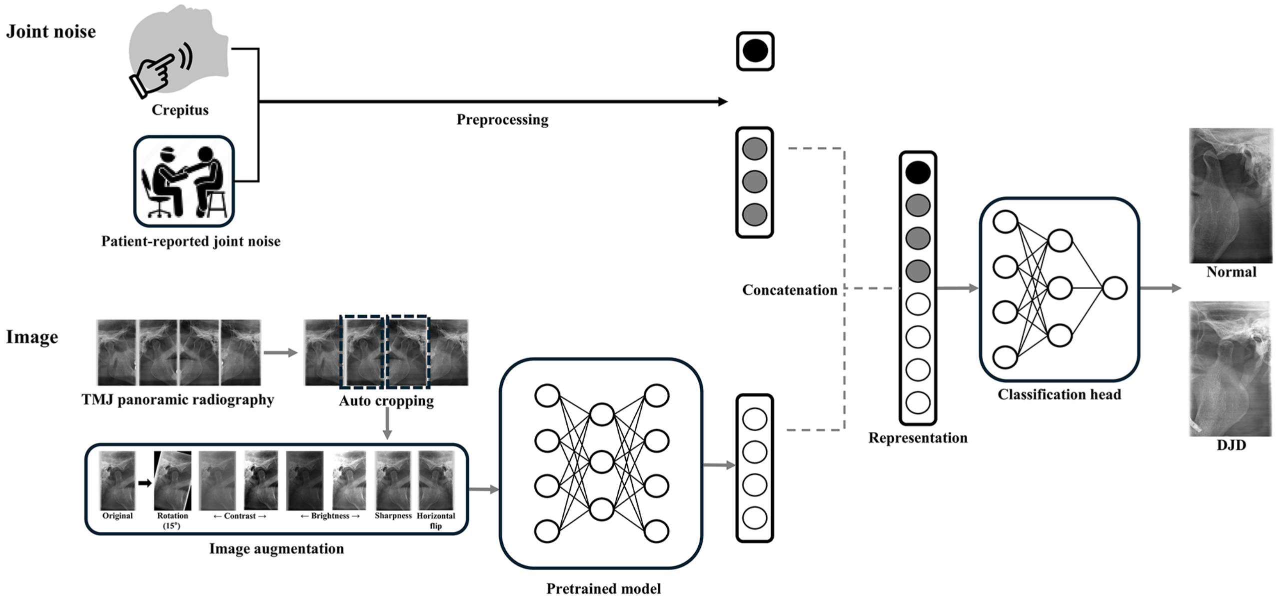

Fig. 4. Architecture of the integrated AI model for DJD diagnosis. AI, artificial intelligence; DJD, degenerative joint disease, TMJ; temporomandibular joint.

Model 6: image, crepitus and certain patient-reported joint noise (R, PR)

The structure of integrated AI model architecture was shown in Fig. 4

Evaluation of AI models' performance and clinical usability

The diagnostic performance of the AI model was evaluated by splitting the dataset into 60% training data and 40% test data. We conducted three repetitions of experiments with different training/test samplings. Due to class imbalance in the dataset, we used the F1-score instead of accuracy for performance evaluation. To train the AI model, we used the stochastic gradient descent (SGD) optimizer with a learning rate of 1.0×10^{-3}. We trained AI models for a total of 300 epochs. The design, training, and evaluation of the diagnostic model were implemented using Python, with the PyTorch library utilized for designing and training the deep learning models. The saliency map was used to interpret and visualize the regions of the image considered by the AI while making predictions.

For the clinical usability evaluation, 100 images labeled as "D" and 100 images labeled as "N" were randomly selected and mixed. These images were then reviewed by eight orofacial pain specialists. Each specialist evaluated the images twice, with and without noise joint information, at one-week intervals.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 14 September 2024; Accepted: 17 December 2024
Published online: 13 January 2025

References

1. Wang, X., Zhang, J., Gan, Y. & Zhou, Y. Current understanding of pathogenesis and treatment of TMJ osteoarthritis. *J. Dent. Res.* **94**, 666–673 (2015).
2. Song, H., Lee, J. Y., Huh, K.-H. & Park, J. W. Long-term changes of temporomandibular joint osteoarthritis on computed tomography. *Sci. Rep.* **10**, 6731. https://doi.org/10.1038/s41598-020-63493-8 (2020).
3. Kim, K., Wojczyńska, A. & Lee, J.-Y. The incidence of osteoarthritic change on computed tomography of Korean temporomandibular disorder patients diagnosed by RDC/TMD; a retrospective study. *Acta Odontol. Scand.* **74**, 337–342. https://doi.org/10.3109/00016357.2015.1136678 (2016).
4. Schiffman, E. L. et al. The research diagnostic criteria for temporomandibular disorders. I: overview and methodology for assessment of validity. *J. Orofac. Pain* **24**, 7 (2010).
5. Almăşan, O., Leucuţa, D. C., Hedeşiu, M., Mureşanu, S. & Popa, ŞL. Temporomandibular joint osteoarthritis diagnosis employing artificial intelligence: systematic review and meta-analysis. *J. Clin. Med.* **12**, 942. https://doi.org/10.3390/jcm12030942 (2023).
6. Manfredini, D., Basso, D., Salmaso, L. & Guarda-Nardini, L. Temporomandibular joint click sound and magnetic resonance-depicted disk position: which relationship?. *J. Dent.* **36**, 256–260. https://doi.org/10.1016/j.jdent.2008.01.002 (2008).
7. Wiese, M. et al. Association between temporomandibular joint symptoms, signs, and clinical diagnosis using the RDC/TMD and radiographic findings in temporomandibular joint tomograms. *J. Orofac. Pain* **22**, 239–251 (2008).
8. Schiffman, E. et al. Diagnostic criteria for temporomandibular disorders (DC/TMD) for clinical and research applications: recommendations of the international RDC/TMD consortium network and orofacial pain special interest group. *J. Oral Facial Pain Headache* **28**, 6–27. https://doi.org/10.11607/jop.1151 (2014).

9. Abrahamsson, A.-K. et al. Frequency of temporomandibular joint osteoarthritis and related symptoms in a hand osteoarthritis cohort. *Osteoarthr. Cartil.* **25**, 654–657 (2017).
10. Im, Y.-G. et al. Diagnostic accuracy and reliability of panoramic temporomandibular joint (TMJ) radiography to detect bony lesions in patients with TMJ osteoarthritis. *J. Dent. Sci.* **13**, 396–404 (2018).
11. Honey, O. B. et al. Accuracy of cone-beam computed tomography imaging of the temporomandibular joint: comparisons with panoramic radiology and linear tomography. *Am. J. Orthod. Dentofacial Orthop.* **132**, 429–438. https://doi.org/10.1016/j.ajodo.2005.10.032 (2007).
12. Ahmad, M. et al. Research diagnostic criteria for temporomandibular disorders (RDC/TMD): development of image analysis criteria and examiner reliability for image analysis. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* **107**, 844–860 (2009).
13. Choi, E., Kim, D., Lee, J.-Y. & Park, H.-K. Artificial intelligence in detecting temporomandibular joint osteoarthritis on orthopantomogram. *Sci. Rep.* **11**, 10246. https://doi.org/10.1038/s41598-021-89742-y (2021).

Author contributions
E.C. contributed to the conception, design, data acquisition, analysis, and interpretation, drafted and critically revised the manuscript. Y.S. contributed to the design, data analysis, and interpretation, and critically revised the manuscript. S.T.K. contributed to the conception, design, critically revised the manuscript. S.S., K.L. and T.A. contributed to data analysis, interpretation, and original draft. R.L. and S.K. contributed to data acquisition. All authors gave final approval and agreed to be accountable for all aspects of the work.

Funding
This study was supported only by the Yonsei University College of Dentistry Fund (6-2024-0005).

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-83750-4.

Correspondence and requests for materials should be addressed to Y.S. or S.T.K.