The Open Dentistry Journal
DOI: 10.2174/01187421063134260810074044, 2026, 20, e187421063134
ISSN: 1874-2106

RESEARCH ARTICLE
OPEN ACCESS

Utilizing Artificial Intelligence to Diagnose Temporomandibular Joint Osteoarthritis *via* Orthopantomographs

Thanh Manh Nguyen[1,*], Thanh Chi Nguyen[2], Hong Thu Thi Le[2], Huyen Thi Phung[3], Hung Trieu Dang[3], Nguyen Manh Truong[3], Quynh Thanh Thi Nguyen[4], Phuong Thu Thi Nguyen[1], Hanh My Bui[1], Anh Ngoc Thi Le[5] and Phong Hoang Nguyen[1]

[1]*Hanoi Medical University, Hanoi, Vietnam*
[2]*Le Quy Don Technical University, Hanoi, Vietnam*
[3]*Hanoi Medical University Hospital, Hanoi, Vietnam*
[4]*Align Dental Clinic, Hanoi, Vietnam*
[5]*Research Department, ClinGen A Chau Laboratory, Hanoi, Vietnam*

Abstract:

Introduction/Background: Temporomandibular Joint Osteoarthritis (TMJOA) is a common temporomandibular disorder that substantially impairs quality of life. Although cone-beam computed tomography is considered the gold standard for detecting osseous changes, its routine use is limited. Orthopantomographs are widely available but have reduced sensitivity for early TMJOA. This study aimed to evaluate the clinical utility of an artificial intelligence model for TMJOA diagnosis using panoramic images and to compare its performance with expert assessment.

Materials and Methods: A total of 651 participants with clinical symptoms suggestive of TMJOA were included. An automated deep learning framework based on YOLOv11 was developed to extract regions of interest encompassing the mandibular condyle, articular fossa, and articular eminence to classify joints as normal or osteoarthritic. Model performance was assessed using five-fold cross-validation, and diagnostic metrics evaluated included accuracy, sensitivity, specificity, area under the curve, Cohen's kappa, and McNemar's test.

Results: The AI model achieved a mean accuracy of 0.76, sensitivity of 0.63, specificity of 0.82, and an AUC of 0.72. Compared to experts, the AI demonstrated higher accuracy and sensitivity, while experts retained higher specificity (0.91 compared to 0.82). The agreement between the AI and expert diagnoses was moderate.

Discussion: The AI model achieved moderate diagnostic performance, consistent with prior dental AI literature, and outperformed expert assessment in sensitivity (0.63 *vs*. 0.51). Conversely, the expert retained higher specificity (0.91), consistent with a conservative interpretive threshold shaped by clinical experience. This sensitivity-specificity trade-off mirrors broader patterns observed in AI-versus-human diagnostic comparisons. These findings support AI as a complementary screening tool, particularly where access to CBCT or specialist expertise is limited.

Conclusion: The proposed AI model improves sensitivity for TMJOA detection on panoramic radiographs and performs comparably to expert clinicians. These findings support the potential role of AI as a cost-effective screening and decision-support tool in routine dental practice, particularly where access to CBCT is limited.

Keywords: Artificial intelligence, Temporomandibular joint osteoarthritis, Orthopantomograph, TMDs, TMJOA, AI.

*Address correspondence to this author at the Hanoi Medical University, Hanoi, Vietnam;
E-mail: manhthanh.hmu@gmail.com

Cite as: Nguyen T, Nguyen T, Le H, Phung H, Dang H, Truong N, Nguyen Q, Nguyen P, Bui H, Le A, Nguyen P. Utilizing Artificial Intelligence to Diagnose Temporomandibular Joint Osteoarthritis *via* Orthopantomographs. Open Dent J, 2026; 20: e187421063134. http://dx.doi.org/10.2174/01187421063134260810074044

Received: February 13, 2026
Revised: July 09, 2026
Accepted: July 20, 2026
Published: August 12, 2026

1. INTRODUCTION

Temporomandibular Joint Osteoarthritis (TMJOA) is one of the most common disorders in the Temporomandibular Joint (TMJ) group, leading to many consequences, including joint pain, trismus, malocclusion, and even permanent deformity [1]. TMJOA is a complex condition characterized by the resorption of articular cartilage and the remodeling and ossification of subchondral bone. Over the world, TMJOA significantly affects the quality of life of 10 - 15% of the population [2].

The pathogenesis of TMJOA involves a complex interplay of mechanical, inflammatory, cellular, and tissue-level processes within the joint [3–5]. Although several contributing factors and signaling pathways have been identified, the underlying mechanisms of TMJOA remain incompletely understood. A central feature of TMJOA is abnormal remodeling of the subchondral bone in the mandibular condyle, driven by multiple interacting factors, including excessive mechanical loading, inflammatory responses, and degeneration of the articular disc [6]. In the early stage of the disease, increased osteoclastic bone resorption occurs, initiating degeneration of the overlying articular cartilage. This is followed by a prolonged and inadequate bone repair process, resulting in subchondral bone sclerosis and increased bone mineral density. Consequently, the osteochondral interface becomes thicker and stiffer, further compromising joint biomechanics and accelerating disease progression. Clinically, patients with TMJOA commonly present with restricted mouth opening, pain during mastication, joint sounds such as clicking or crepitus, mandibular deviation during mouth opening, headache, and otology-related symptoms [7–9].

Panoramic X-ray images are widely used in the evaluation of maxillofacial structures; however, this technique has certain limitations. The most notable limitation is its difficulty in detecting small changes on the surface of the temporomandibular joint due to being overlapped by nearby anatomical structures (Fig. **1**). Cone-Beam Computed Tomography (CBCT) is considered the reference standard in detecting TMJOA due to its ability to display detailed changes in the cortical and subcortical bone layers [10]. Nevertheless, CBCT is still not the first choice for TMJOA examination in routine clinical settings due to its high radiation exposure and high cost [11]. In this context, the application of new methods to enhance the early and accurate diagnosis of TMJOA is essential.

A promising approach is the use of Artificial Intelligence (AI). AI has shown broad potential in dentistry, from analyzing lateral cephalograms to localizing third molars and the inferior alveolar nerve, as well as diagnosing maxillary sinusitis, dental caries, and other conditions [11, 12]. A recent systematic review concluded that Artificial Intelligence (AI) has considerable potential for the imaging-based diagnosis of TMJOA, demonstrating a pooled sensitivity of approximately 80%, a specificity of up to 90%, and an area under the receiver operating characteristic curve as high as 0.92 [13]. Deep learning models employing transfer learning with ResNet architectures have also been applied to panoramic radiographs, achieving diagnostic sensitivities and specificities of approximately 76–79% [11]. More recently, researchers have expanded AI-based diagnostic approaches by integrating multimodal data rather than relying solely on panoramic radiographs. For example, Eunhye Choi and colleagues developed a diagnostic framework that combined orthopantomogram (OPG) image analysis with temporomandibular joint acoustic signals (joint sounds) recorded during mandibular movement, thereby enhancing the comprehensiveness of TMJOA assessment [14]. Most published studies have been conducted in single-center settings with relatively limited datasets, restricting the generalizability and clinical applicability of their findings. Consequently, further investigations are needed to establish their role in routine clinical practice.

Fig. (1). Orthopantomograph results.

Based on the current clinical applications and research background, this study aims to investigate the clinical utility of an AI diagnostic tool developed for TMJOA diagnosis from Orthopantomograph (OPG) images using algorithms that compare the AI diagnosis with that of an expert.

2. METHODOLOGY

2.1. Participants and Sample Size

Radiographic images of the patients with the following selection and exclusion criteria were reviewed.

Selection criteria:

- Visited the Department of Odonto and Stomatology, Hanoi Medical University Hospital with at least 1 of the following clinical symptoms: temporomandibular joint pain, crepitus, and trismus.
- Have panoramic images.
- Accepted to enroll in the study.

Exclusion criteria:

- Had a history of orthopedic surgery.
- Had a history of systemic diseases.

The sample size of the study uses the following formula to evaluate the sensitivity and specificity of a disease diagnosis method:

$$n_{se} = \left(\frac{Z_{\alpha}^{2} \times p_{se} \times (1 - p_{se})}{w^{2}}\right)/p_{dis}$$

With:

n_{se}: Sample size for sensitivity,

Z_{α}^{2}: standard normal deviate (α = 0.05 then Z = 1.96)

p_{se}: Probability of sensitivity. Take p_{se} = 80% [15].

w: Error of sensitivity and specificity.

p_{dis}: Prevalence of the disease in the population. Take p_{dis} = 0.15 [16].

$$n_{se} = \left(\frac{1{,}96^{2} \times 0{,}73 \times (1 - 0{,}73)}{0{,}09^{2}}\right)/0{,}15 = 623$$

In reality, the research sample size was 651 participants.

Figure **2** illustrates the step-by-step schema of the study.

2.2. Orthopantomogram Image Criteria and TMJOA Diagnosis Criteria

All panoramic radiographs were acquired with standardized imaging parameters: 12 mA tube current, 85 kV tube voltage, and a mean exposure time of 17.6 seconds. Acceptable image quality includes the following:

- Alveolar processes with all upper and lower teeth must be clearly represented.
- The image should show the entire mandible, including the TMJ.
- The dimensions of anatomical structures in vertical and horizontal planes must be symmetrical.
- Density image appearance should be uniform, free of air over the tongue, with the appearance of transparent tape (black) over the roots of upper teeth.
- The panoramic image should not have artifacts due to dentures or removable orthodontic appliances, glasses, earrings, and other jewelry on the patient.

Patients were diagnosed with TMJOA based solely on radiographic evidence of temporomandibular joint changes.

2.3. AI Model Development

An automated model was developed to extract regions of interest (ROI), including the mandibular condyle and surrounding anatomical structures, and to classify them into two categories: normal and osteoarthritis (OA), from each OPG image using a YOLOv11-based object detection framework. Once an image is input into the system, the YOLOv11 model performs inference to detect regions suspected of containing temporomandibular joint abnormalities. The results are presented visually as bounding boxes overlaid on the image, accompanied by predicted class labels and corresponding confidence scores (Fig. **3**).

YOLOv11 is a one-stage object detection framework designed for real-time applications by integrating localization and classification into a unified neural network [17]. Unlike two-stage approaches, it eliminates the need for a region proposal mechanism, allowing the model to infer object locations and categories directly from the input image. The detection process begins with preprocessing, where the image is resized and normalized to ensure consistency. During inference, post-processing techniques such as confidence filtering and non-maximum suppression are employed to remove redundant predictions and yield accurate final detections.

Unlike conventional approaches that rely on manual selection of regions of interest (ROIs), the detection-based architecture of YOLOv11 simultaneously localizes the mandibular condyle and performs classification directly on the detected region. This approach is particularly well suited to panoramic radiographs, where the mandibular condyle typically occupies only a small portion of the image and can easily be obscured by surrounding anatomical structures [18, 19]. In this context, whole-image analysis or manual ROI selection may reduce the signal-to-noise ratio and limit the model's ability to focus on diagnostically relevant features. By learning directly from automatically localized regions, YOLOv11 can more effectively capture morphological characteristics associated with TMJOA, including articular surface erosion, subchondral sclerosis, and relative changes in joint space morphology.

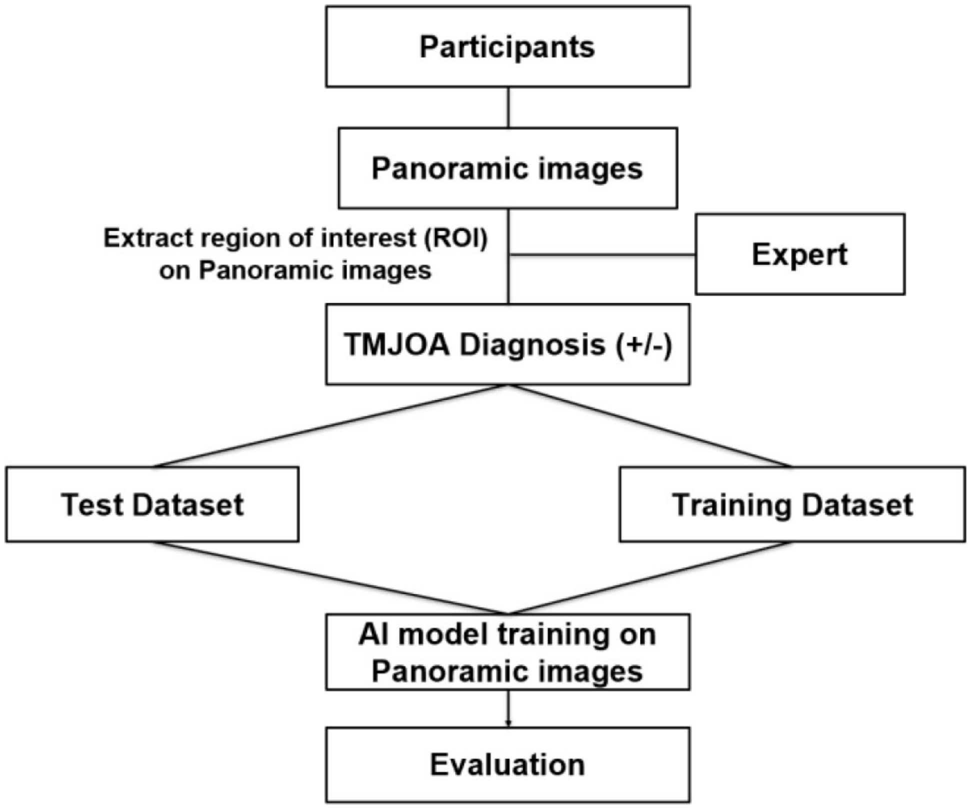

Fig. (2). Study schema.

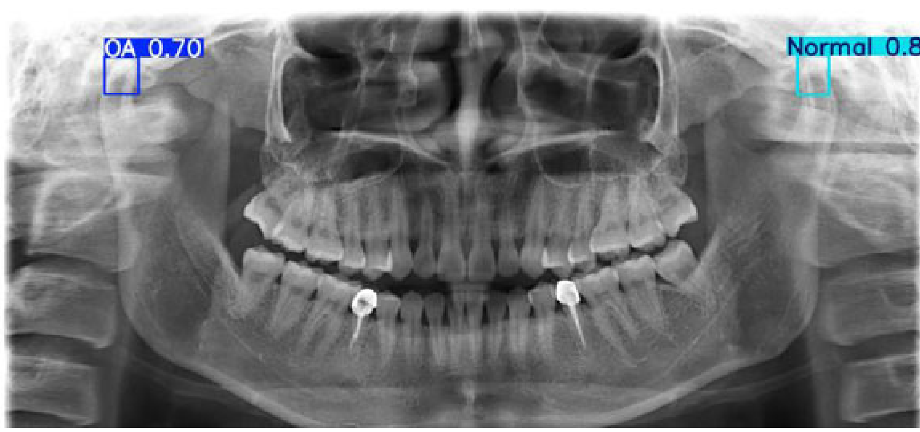

Fig. (3). Results of ROI extraction.

The YOLOv11 model was implemented using the PyTorch deep learning framework, enabling efficient model training, optimization, and inference through its dynamic computational graph and GPU acceleration support. All algorithms were programmed and trained on a PC with a GeForce GTX 1080 Ti GPU. Model training was performed using the Adam optimizer with a learning rate of 1.0×10^{-6}. All the training data is divided into mini-batches, with a mini-batch size is set to 16 during training. Training was conducted for up to 1,000 epochs until convergence, with early stopping applied based on the validation loss to prevent overfitting.

Data augmentation techniques, including image rotation (±5 degrees), horizontal and vertical shifts (±10%), brightness adjustment (±10%), and contrast adjustment (±10%), were applied to enhance data diversity and improve model robustness.

2.4. Model and Statistical Analysis

Accuracy, precision, recall, and F1 score were calculated for model performance. Accuracy is defined as the ratio of correct predictions. Precision is the ratio of true positives to true positives and false positives. Recall is the ratio of true positives to true positives and false negatives. Finally, the F1 score is a harmonic mean of precision and recall: (2 × precision × recall)/(precision + recall). Accuracy, specificity, and sensitivity were calculated for diagnostic performance; Cohen's kappa was used to estimate agreement in TMJOA diagnoses between experts and AI reads; and McNemar's test was used to evaluate the significance of the difference. All p values < 0.05 were considered to be statistically significant.

2.5. Ethics Approval and Consent to Participate

The study was approved by the University Council - Hanoi Medical University (Approval No. 4571/QD-DHYHN – 06/10/2023). All procedures involving human participants were performed in accordance with the ethical standards of the institutional and/or research committees and the 1975 Declaration of Helsinki, as revised in 2013. Agreements for participation were requested, and informed consents were given to all participants.

3. RESULTS

The demographics of participants in the study are shown in Table **1**. A total of 651 participants were included in the analysis, comprising 377 osteoarthritic (OA) joints. For k-fold validation, a five-fold cross-validation strategy (k = 5) was adopted to ensure robust model evaluation and to mitigate bias arising from limited data availability. The dataset is divided into 5 (k=5) equal-sized folds, where each fold is used once as a testing set while the remaining folds are used to train the machine learning model (Supplement **1**).

A total of 651 participants were included in the analysis, comprising 368 individuals in the normal group and 283 individuals who had TMJOA. In the normal group, females accounted for 226 participants (61.4%) and males for 142 participants (38.6%). The mean age was 36.34 years (SD = 14.51), with a 95% confidence interval (CI) of 34.85-37.83.

The model's performance is best in fold 2 (0.79), followed by fold 3 (0.78) and fold 1 (0.76) (Table **2**). The recall result is found best in fold 2 (0.69), and the highest precision is shown in fold 3 (0.77) (Table **3**). The average accuracy, precision, recall, and F1 score were 0.76, 0.73, 0.63, and 0.68, respectively (Table **3**).

The AI in fold 2 has the highest accuracy (0.79) and sensitivity (0.69), but specificity peaks in fold 3, and Cohen's kappa is also found highest in Fold 3 (Table **3**).

Table 1. Clinical and demographic characteristics of the OPG dataset.

Characteristics	Normal			OA		
	Female	Male	Total	Female	Male	Total
Number of participants	226	142	368	199	84	283
Number of joints	452	284	736	268	109	377
Mean age	38.32	33.19	36.34	33.44	31.68	32.92
SD	14.679	13.709	14.511	15.029	17.023	15.638
95% CI	36.40 - 40.25	30.92 - 35.46	34.85 - 37.83	31.34 - 35.54	27.98 - 35.37	31.09 - 34.74

Table 2. Confusion matrix and model performance.

Fold	Confusion Matrix			Model Performance					
	Actual	Predicted		Precision	Recall	Accuracy	Weighted Average Precision	Weighted Average Recall	F1 score
		Normal	OA						
1	**Normal**	176	22	0.89	0.78	0.76	0.80	0.68	0.67
	OA	37	27	0.58	0.42				
2	**Normal**	172	19	0.90	0.80	0.79	0.83	0.75	0.72
	OA	42	27	0.58	0.61				
3	**Normal**	168	17	0.91	0.77	0.78	0.83	0.71	0.72
	OA	28	47	0.63	0.57				
4	**Normal**	152	19	0.89	0.68	0.71	0.77	0.63	0.63
	OA	42	47	0.47	0.49				
5	**Normal**	160	20	0.89	0.73	0.73	0.78	0.64	0.65
	OA	37	43	0.54	0.43				
Average	**Normal**	165.6	19.4	0.90	0.75	0.76	0.80	0.69	0.68
	OA	37.2	38.2	0.56	0.50				

Table 3. 5-fold cross-validation.

Fold	Recall	Precision	F1 Score	Accuracy	AUC
1	0.60	0.73	0.67	0.76	0.71
2	0.69	0.76	0.72	0.79	0.75
3	0.67	0.77	0.72	0.78	0.76
4	0.59	0.68	0.63	0.71	0.69
5	0.58	0.71	0.65	0.73	0.70
Average	0.63	0.73	0.68	0.76	0.72

The comparison of the diagnostic performance among folds is shown in Table **4**, and a visual presentation is shown in Fig. (**4**). Table **4** illustrates the diagnostic performance of the model across five cross-validation folds. Accuracy ranged from 0.71 to 0.79, with consistently higher specificity compared with sensitivity across all folds. Cohen's kappa values indicate moderate agreement, ranging from 0.401 to 0.589. The highest agreement is observed in Fold 3 ($\kappa = 0.589$), while the lowest is in Fold 4 ($\kappa = 0.401$). McNemar's test demonstrates no statistically significant difference between paired classifications in most folds ($p > 0.05$), except for Fold 4, which shows a significant difference ($p = 0.002$). This suggests overall stability of model predictions across folds, with some variability in specific iterations. The visual comparison of sensitivities and specificities across folds is shown in Fig. (**4**).

The difference in diagnostic performance is shown in Table **5** and Fig. **5**. Cohen's kappa shows a substantial level of agreement for the expert (0.62) and moderate agreement for the AI (0.50) (Table **5**). On average, the expert diagnostic performance is inferior compared to AI in terms of accuracy and sensitivity. However, the diagnosis of true negatives is better read by experts.

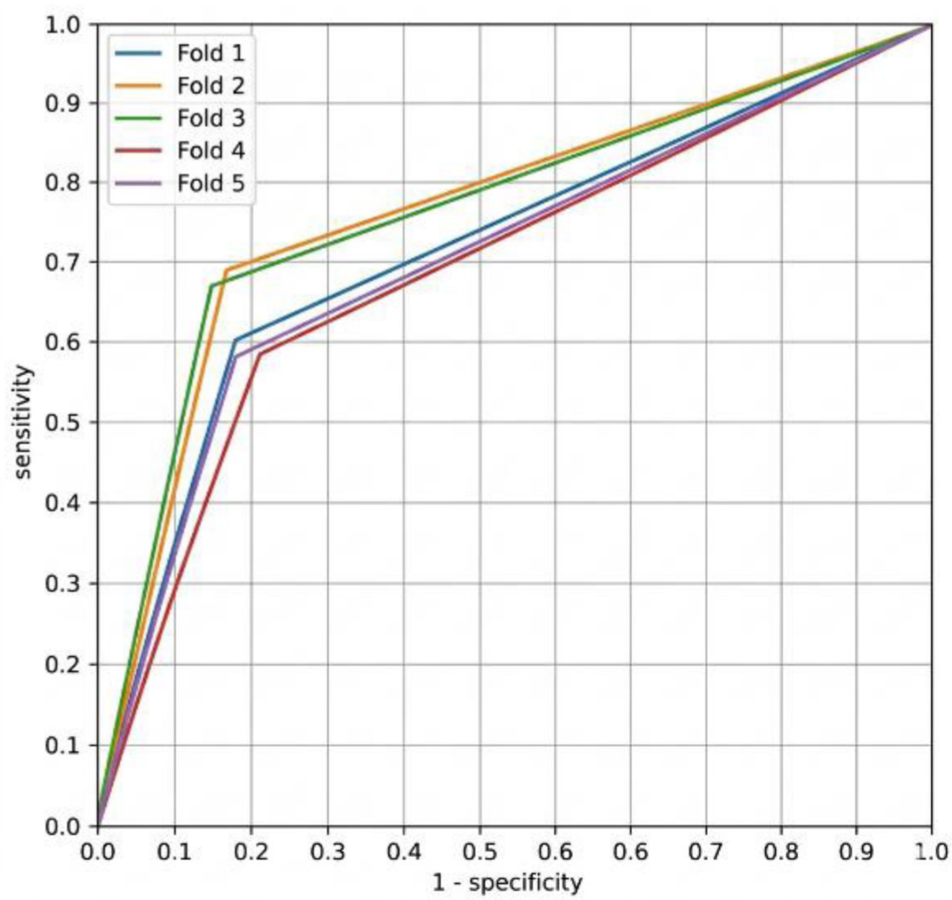

Fig. (4). Comparison of the sensitivities and specificities among folds.

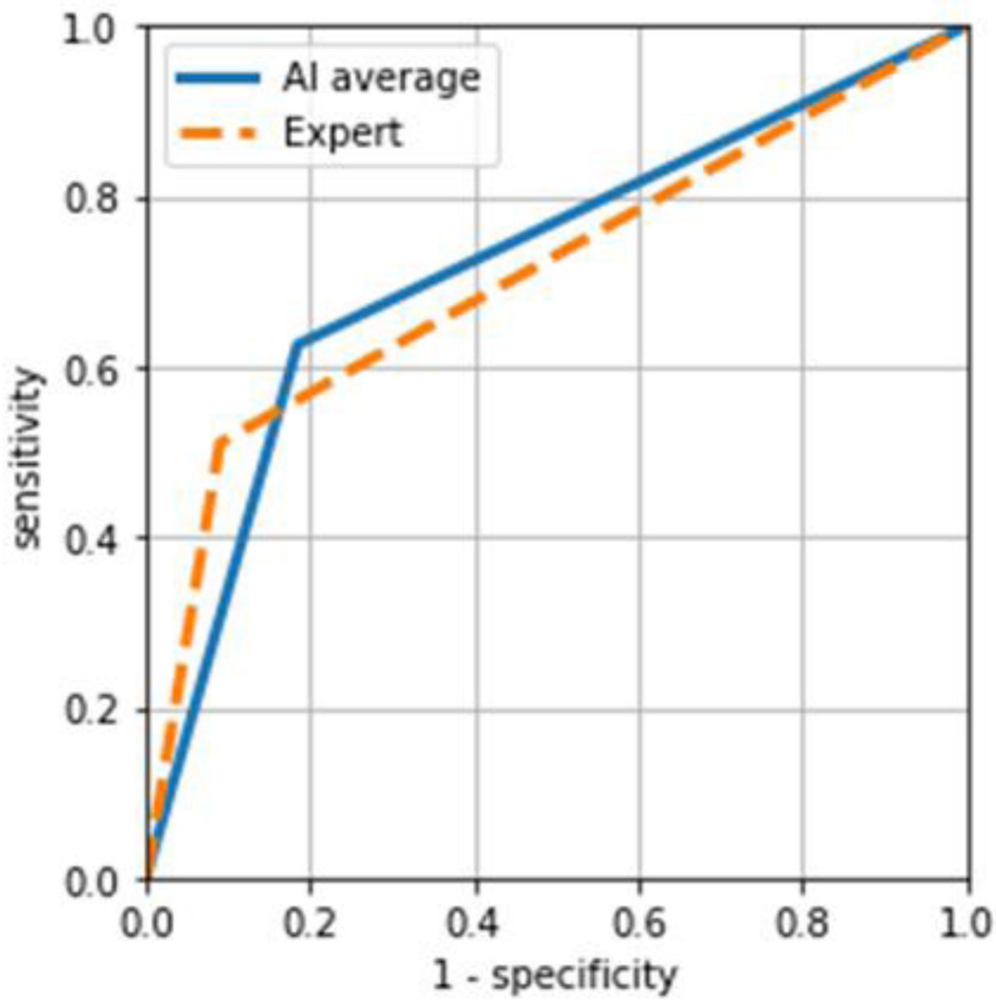

Fig. (5). Comparison of the sensitivities and specificities between AI and Expert.

Table 4. Comparison of the diagnostic performance among folds.

	Diagnostic Performance			Cohen's Kappa	McNemar's test (*p*-Value)
	Accuracy	Sensitivity	Specificity		
Fold 1	0.76	0.60	0.82	0.487	0.666
Fold 2	0.79	0.69	0.82	0.566	0.533
Fold 3	0.78	0.67	0.84	0.589	0.117
Fold 4	0.71	0.59	0.79	0.401	0.002
Fold 5	0.73	0.58	0.82	0.466	0.058

Table 5. Comparison of the diagnostic performance between AI and Expert.

	Diagnostic Performance			Cohen's Kappa
	Accuracy	Sensitivity	Specificity	
AI average	0.76	0.63	0.82	0.502
Expert	0.69	0.51	0.91	0.621

The AI model demonstrates higher overall accuracy and greater sensitivity than the experts, indicating an improved ability to correctly identify osteoarthritis. In contrast, the experts achieve higher specificity, reflecting a stronger performance in correctly identifying normal joints. Regarding agreement with the reference standard, the experts show a higher Cohen's kappa score than the AI model, corresponding to substantial agreement, whereas the AI exhibits moderate agreement. These findings suggest that while the AI model offers superior case detection and overall accuracy, expert assessment remains more conservative, resulting in fewer false-positive classifications.

The curves comparing the average performance of the AI system and expert assessment are shown in Fig. (**5**). Both approaches demonstrated good discriminatory ability, with sensitivity increasing as specificity decreased. The findings suggest that the AI system performs at least comparably to expert evaluation, with a potential advantage in sensitivity.

4. DISCUSSION

The study developed an artificial intelligence model based on the YOLOv11 framework for the automatic detection of Temporomandibular Joint Osteoarthritis (TMJOA) using orthopantomographs (OPG). To our knowledge, this is among the first studies in Vietnam to apply deep learning for this specific diagnostic challenge. The primary finding of this investigation is that the AI model demonstrated superior diagnostic accuracy and sensitivity compared to human expert assessment, although the expert maintained higher specificity.

4.1. Comparison with Previous Deep Learning Models

In this study, our deep learning model demonstrated moderate diagnostic performance in the interpretation of panoramic dental radiographs. While slightly lower than some previous benchmarks, these results fall within the range reported in the literature on AI for dental image analysis. One of the earliest deep learning investigations in dental radiography applied a deep neural transfer network (DeNTNet) for the detection of periodontal bone loss in panoramic radiographs [20]. This model, trained on over 12,000 images, achieved an F1 score of 0.75, indicating robust lesion detection and competitive performance relative to clinicians. The variation in performance across studies may be partly explained by design differences. Models like Mask R-CNN, which perform pixel-level segmentation, often provide more precise boundary delineation, enhancing performance on tasks requiring detailed localization compared to object detection-oriented networks that prioritize speed, such as YOLO variants [21]. A systematic review of deep learning in dental radiographs observed that while standard convolutional and segmentation networks generally yield high pooled diagnostic metrics, their efficacy remains highly sensitive to variations in dataset size, imaging modality, and clinical task [22].

4.2. Performance Comparison between AI and Expert

A prominent finding in our study was the marked difference in diagnostic performance between the AI model and expert clinicians. Expert clinicians exhibited high specificity (0.91) but notably low sensitivity (0.51), consistent with well-recognized limitations of OPGs in detecting subtle osseous changes due to anatomical superimposition and low contrast resolution. This discrepancy likely reflects well-documented limitations of OPGs, where overlapping anatomical structures, low resolution, and geometric distortion reduce the visibility of subtle bone changes, making early disease difficult to detect for human observers. These limitations make early condylar bone changes difficult to visualize, even for experienced interpreters, and echo findings from broader literature showing that standard panoramic imaging has limited reliability compared to volumetric modalities such as CBCT for TMJOA evaluation [23]. Previous research confirms that panoramic imaging often underperforms more advanced modalities such as CBCT for bone pathology, including condylar degeneration and other mandibular changes, largely due to these inherent imaging constraints. CBCT provides multiplanar, three-dimensional visualization without superimposition, resulting in superior diagnostic accuracy for changes that may be indiscernible on 2D OPGs [24, 25].

The study's AI model increased sensitivity from 0.51 of experts to 0.63, identifying true positive cases that were missed by experts. This enhanced ability to detect subtle radiographic patterns mirrored results from previous systematic studies of AI applications in TMJOA detection. A recent meta-analysis of AI models applied to TMJOA radiographic datasets reported pooled sensitivity and specificity values around 0.80 and 0.79, when using both CBCT and panoramic imaging modalities, underscoring that deep learning systems can achieve meaningful diagnostic performance across variable imaging conditions [11]. Furthermore, this review highlighted that architectures such as ResNet exhibit moderate to high sensitivity when trained on adequately annotated condylar osteoarthritis datasets, supporting our finding that AI may overcome some of the human limitations inherent in two-dimensional radiographs. Numerous studies in dental imaging also highlighted the capacity of CNN-based models to extract non-linear, sub-visual features, enabling more sensitive detection of radiographic anomalies such as periapical lesions, bone loss, or other pathologies. In a recent large-scale evaluation of AI performance on panoramic radiographs, deep learning achieved moderate to high sensitivity and specificity for detecting dental conditions, often approaching or exceeding the performance of human readers in certain tasks [26, 27].

Several individual diagnostic accuracy studies further illustrate the performance landscape of AI in TMJOA detection. One study developed deep learning models for condylar osteoarthritis classification using panoramic TMJ projection images and found that models such as GoogLeNet and VGG-16 achieved AUCs up to 0.89, outperforming less experienced human observers, particularly on standardized TMJ projections [28]. Another study focused specifically on OPGs showed that a ResNet-based algorithm yielded sensitivity to 0.73 and specificity to 0.82 when trained and tested in a clinically relevant setting confirmed by CBCT, thus demonstrating performance comparable to experts while achieving a more balanced trade-off [15]. It is also notable that many approaches utilize advanced transfer learning and pre-trained architectures, such as ResNet-152 and EfficientNet-B7, had demonstrated that model selection and training strategy significantly influence diagnostic outcomes [29]. These findings support the notion that AI systems can extract and integrate subtle texture and morphological cues from low-dimensional imaging that may be below the threshold of visual detection by human readers.

Despite these advances, expert clinicians often retain higher specificity, which highlights a key benefit of human judgment in reducing false positives. In our cohort, experts achieved a specificity of 0.91, indicating a strong ability to correctly identify normal temporomandibular joints and avoid overdiagnosis. This finding aligns with multiple radiology and dental imaging studies reporting that experienced clinicians adopt a conservative interpretative threshold, particularly when evaluating panoramic radiographs for subtle degenerative changes. Several factors may explain this pattern.

First, clinicians rely heavily on well-established radiographic hallmarks of TMJOA, such as clear condylar erosion, flattening, osteophyte formation, or sclerosis, and are often reluctant to label early or ambiguous findings as pathological in the absence of unequivocal structural changes [30]. As a result, borderline radiographic variations, age-related remodeling, or projection artifacts are frequently classified as normal by human observers, thereby increasing specificity at the expense of sensitivity.

Second, expert judgment is often influenced by implicit clinical context, even when formal clinical data are not provided. Years of exposure to normal anatomical variability allow clinicians to recognize patterns that are unlikely to represent clinically meaningful TMJOA [31]. This experiential knowledge, while valuable for ruling out disease, may also contribute to underrecognition of early osteoarthritic changes, particularly when structural alterations are mild or atypical [32]. Consequently, expert readers may prioritize diagnostic certainty over early detection, reinforcing high specificity. Several investigations demonstrate that while AI models frequently achieve higher sensitivity, expert clinicians consistently outperform AI in confirming normal joints. This suggests that specificity may represent a strength of human expertise, rooted in holistic interpretation and risk-averse decision-making, whereas AI excels at identifying subtle deviations from learned patterns, regardless of their clinical salience.

From a clinical perspective, this trade-off has meaningful implications. In definitive diagnostic settings or specialist care, high specificity is essential to prevent misclassification and unnecessary escalation of care. However, in screening or primary care, the cost of missed TMJOA diagnoses, particularly in the early stages, may outweigh the drawbacks of increased false positives. In such scenarios, AI-assisted screening could complement expert interpretation by flagging suspicious cases for further review, while clinicians retain ultimate authority over diagnosis. Overall, given the clinical implications of missing an early TMJOA diagnosis, the improved sensitivity of AI models suggests a valuable role for AI as a decision-support tool, particularly in settings where access to CBCT or highly experienced specialists is limited.

Artificial intelligence has emerged as a promising approach in radiology, a field characterized by large volumes of relatively standardized imaging data. Although the current literature remains limited in both quantity and methodological rigor, existing evidence consistently suggests that AI can achieve diagnostic performance comparable to that of healthcare professionals [33–36]. Notably, several studies have reported improved model performance when the training set consisted of ROI images rather than entire radiographs [37]. However, in most published work, ROIs are manually delineated, which reduces the scalability and clinical practicality of AI-based workflows. To address this limitation, the present study implemented an object-detection algorithm to automatically extract ROIs from orthopantomogram (OPG) images.

Previous research using similar techniques has reported exceptionally high precision in mandibular condyle detection, with average precision values of 99.4% on the right side and 100% on the left side [38]. Importantly, the automatically generated ROIs in our study encompassed not only the mandibular condyle but also the articular fossa and articular eminence. This broader anatomical coverage is clinically relevant, as TMJOA is not confined to condylar changes alone. Degenerative alterations are known to involve additional joint components, including the fossa and the eminence, underscoring the need for AI diagnostic models to analyze the entire joint rather than focusing solely on the condyle [39]. Consistent with this rationale, large-scale reviews have reported pooled diagnostic performance with AUC values exceeding 0.90, underscoring the strong clinical potential of AI-based approaches, although variability related to imaging modality and dataset characteristics remains an important factor [13].

5. LIMITATIONS

This study has several limitations. First, the study's ground truth was established based on OPG features. OPGs have inherent geometric distortions; therefore, some "false positives" identified by the AI might actually be true positives if validated against a CBCT gold standard, or *vice versa*. Second, the study was conducted at a single center in Vietnam. While the sample size (N=651) is quite robust, external validation on datasets from different OPG machines and different races is necessary to ensure the model generalizes well to other populations

CONCLUSION

Overall, the combination of the YOLOv11 model, a context-aware ROI selection strategy, and an appropriate data preprocessing pipeline resulted in a diagnostic model well adapted to the characteristics of panoramic radiographs for TMJOA detection. This integrated approach was designed not only to improve diagnostic performance but also to establish a practical and clinically applicable training framework with the potential to facilitate AI-assisted TMJOA diagnosis in routine clinical settings. Although further validation is required to confirm the model's robustness, its performance on the data suggests it is a viable, cost-effective screening aid in general dental practice.

AUTHORS' CONTRIBUTIONS

The authors confirm contributions to the paper as follows: T.M.N., T.C.N., H.T.T.L., P.T.T.N., H.M.B.: Study conception and design; N.M.T., Q.T.T.N.: Data collection; T.M.N., T.C.N., H.T.T.L., A.N.T.L., P.H.N.: Analysis and interpretation of results; T.M.N., H.T.P., H.T.D.: Draft manuscript. All authors reviewed the results and approved the final version of the manuscript.

LIST OF ABBREVIATIONS

AI	=	Artificial Intelligence
CBCT	=	Cone-Beam Computed Tomograph
DeNTNet	=	Deep Neural Transfer Network
OPG	=	Orthopantomogram
ROI	=	Region of Interest (ROI)
TMJOA	=	Temporomandibular Joint Osteoarthritis
TMDs	=	Temporomandibular Disorders

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study was approved by the University Council - Hanoi Medical University (Approval No. 4571/QD-DHYHN – 06/10/2023).

HUMAN AND ANIMAL RIGHTS

All procedures involving human participants were performed in accordance with the ethical standards of the institutional and/or research committees and the 1975 Declaration of Helsinki, as revised in 2013.

CONSENT FOR PUBLICATION

Agreements for participation were requested, and informed consents were obtained from all participants.

STANDARDS OF REPORTING

STROBE guidelines were followed.

AVAILABILITY OF DATA AND MATERIALS

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

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

[1] Jha N, Lee K, Kim YJ. Diagnosis of temporomandibular disorders using artificial intelligence technologies: A systematic review and meta-analysis. PLoS One 2022; 17(8): e0272715. http://dx.doi.org/10.1371/journal.pone.0272715 PMID: 35980894

[2] Lee KS, Kwak HJ, Oh JM, *et al.* Automated detection of TMJ osteoarthritis based on artificial intelligence. J Dent Res 2020; 99(12): 1363-7. http://dx.doi.org/10.1177/0022034520936950 PMID: 32609562

[3] Li B, Guan G, Mei L, Jiao K, Li H. Pathological mechanism of chondrocytes and the surrounding environment during osteoarthritis of temporomandibular joint. J Cell Mol Med 2021; 25(11): 4902-11. http://dx.doi.org/10.1111/jcmm.16514 PMID: 33949768

[4] de Souza RF, Lovato da Silva CH, Nasser M, Fedorowicz Z, Al-Muharraqi MA. Interventions for managing temporomandibular joint osteoarthritis. Cochrane Libr 2012; 2018(6): CD007261. http://dx.doi.org/10.1002/14651858.CD007261.pub2 PMID: 22513948

[5] Cevidanes LHS, Walker D, Schilling J, *et al.* 3D osteoarthritic changes in TMJ condylar morphology correlates with specific systemic and local biomarkers of disease. Osteoarthritis Cartilage 2014; 22(10): 1657-67. http://dx.doi.org/10.1016/j.joca.2014.06.014 PMID: 25278075

[6] Vos L M, Kuijer R, Huddleston Slater J J, Stegenga B. Alteration of cartilage degeneration and inflammation markers in temporomandibular joint osteoarthritis occurs proportionally. J Oral Maxillofac Surg 2013; 71(10): 1659-64. http://dx.doi.org/10.1016/j.joms.2013.06.201 PMID: 23932112

[7] Pergamalian A, Rudy TE, Zaki HS, Greco CM. The association between wear facets, bruxism, and severity of facial pain in patients with temporomandibular disorders. J Prosthet Dent 2003; 90(2): 194-200. http://dx.doi.org/10.1016/S0022-3913(03)00332-9 PMID: 12886214

[8] Dıraçoğlu D, Alptekin K, Çifter ED, Güçlü B, Karan A, Aksoy C. Relationship between maximal bite force and tooth wear in bruxist and non-bruxist individuals. Arch Oral Biol 2011; 56(12): 1569-75. http://dx.doi.org/10.1016/j.archoralbio.2011.06.019 PMID: 21783173

[9] Zhao J, Wang R, Zhu S, Li Z, Ren R, Jiang N. Comparative review: Clinical and pathological heterogeneity in knee *versus* temporomandibular joint osteoarthritis. Front Bioeng Biotechnol 2025; 13: 1684481. http://dx.doi.org/10.3389/fbioe.2025.1684481 PMID: 41000472

[10] Larheim TA, Abrahamsson A-K, Kristensen M, Arvidsson LZ. Temporomandibular joint diagnostics using CBCT. Dentomaxillofac Radiol 2015; 44(1): 20140235. http://dx.doi.org/10.1259/dmfr.20140235 PMID: 25369205

[11] Almășan O, Leucuța DC, Hedeșiu M, Mureșanu S, Popa ȘL. Temporomandibular joint osteoarthritis diagnosis employing

artificial intelligence: Systematic review and meta-analysis. J Clin Med 2023; 12(3): 942.
http://dx.doi.org/10.3390/jcm12030942 PMID: 36769590

[12] Mackie T, Al Turkestani N, Bianchi J, *et al.* Quantitative bone imaging biomarkers and joint space analysis of the articular fossa in temporomandibular joint osteoarthritis using artificial intelligence models. Front Dent Med 2022; 3: 1007011.
http://dx.doi.org/10.3389/fdmed.2022.1007011 PMID: 36404987

[13] Xu L, Chen J, Qiu K, Yang F, Wu W. Artificial intelligence for detecting temporomandibular joint osteoarthritis using radiographic image data: A systematic review and meta-analysis of diagnostic test accuracy. PLoS One 2023; 18(7): e0288631.
http://dx.doi.org/10.1371/journal.pone.0288631 PMID: 37450501

[14] Choi E, Shin S, Lee K, *et al.* Artificial intelligence-enhanced diagnosis of degenerative joint disease using temporomandibular joint panoramic radiography and joint noise data. Sci Rep 2025; 15(1): 1823.
http://dx.doi.org/10.1038/s41598-024-83750-4 PMID: 39805862

[15] Choi E, Kim D, Lee JY, Park HK. Artificial intelligence in detecting temporomandibular joint osteoarthritis on orthopantomogram. Sci Rep 2021; 11(1): 10246.
http://dx.doi.org/10.1038/s41598-021-89742-y PMID: 33986459

[16] Egloff C, Hügle T, Valderrabano V. Biomechanics and pathomechanisms of osteoarthritis. Swiss Med Wkly 2012; 142(2930): w13583-3.
http://dx.doi.org/10.4414/smw.2012.13583 PMID: 22815119

[17] Khanam R, Hussain M. YOLOv11: An overview of the key architectural enhancements. arXiv 2024.
http://dx.doi.org/10.48550/arXiv.2410.17725

[18] Sadat-Khonsari R, Fenske C, Behfar L, Bauss O. Panoramic radiography: Effects of head alignment on the vertical dimension of the mandibular ramus and condyle region. Eur J Orthod 2012; 34(2): 164-9.
http://dx.doi.org/10.1093/ejo/cjq175 PMID: 21467122

[19] Dahlström L, Lindvall AM. Assessment of temporomandibular joint disease by panoramic radiography: Reliability and validity in relation to tomography. Dentomaxillofac Radiol 1996; 25(4): 197-201.
http://dx.doi.org/10.1259/dmfr.25.4.9084273 PMID: 9084273

[20] Kim J, Lee HS, Song IS, Jung KH. DeNTNet: Deep neural transfer network for the detection of periodontal bone loss using panoramic dental radiographs. Sci Rep 2019; 9(1): 17615.
http://dx.doi.org/10.1038/s41598-019-53758-2 PMID: 31772195

[21] Kim H. Deep learning in dental radiographic imaging. J Korean Acad Pediatr Dent 2024; 51(1): 1-10.
http://dx.doi.org/10.5933/JKAPD.2024.51.1.1

[22] Bonfanti-Gris M, Herrera A, Salido Rodríguez-Manzaneque MP, Martínez-Rus F, Pradíes G. Deep learning for tooth detection and segmentation in panoramic radiographs: A systematic review and meta-analysis. BMC Oral Health 2025; 25(1): 1280.
http://dx.doi.org/10.1186/s12903-025-06349-9 PMID: 40739210

[23] Mehta V, Tripathy S, Noor T, Mathur A. Artificial intelligence in temporomandibular joint disorders: An umbrella review. Clin Exp Dent Res 2025; 11(1): e70115.
http://dx.doi.org/10.1002/cre2.70115 PMID: 40066511

[24] Kazimierczak W, Wajer R, Wajer A, *et al.* Periapical lesions in panoramic radiography and CBCT imaging — Assessment of AI's diagnostic accuracy. J Clin Med 2024; 13(9): 2709.
http://dx.doi.org/10.3390/jcm13092709 PMID: 38731237

[25] AlKhairAllah HA, Mohan MP, AlSagri MS. A CBCT based study evaluating the degenerative changes in TMJs among patients with loss of posterior tooth support visiting Qassim University Dental Clinics, KSA: A retrospective observational study. Saudi Dent J 2022; 34(8): 744-50.
http://dx.doi.org/10.1016/j.sdentj.2022.09.002 PMID: 36570571

[26] Ba-Hattab R, Barhom N, Osman S, *et al.* Detection of periapical lesions on panoramic radiographs using deep learning. Appl Sci 2023; 13(3): 1516.
http://dx.doi.org/10.3390/app13031516

[27] Turosz N, Chęcińska K, Chęciński M, Brzozowska A, Nowak Z, Sikora M. Applications of artificial intelligence in the analysis of dental panoramic radiographs: An overview of systematic reviews. Dentomaxillofac Radiol 2023; 52(7): 20230284.
http://dx.doi.org/10.1259/dmfr.20230284 PMID: 37665008

[28] Iwase Y, Sugiki T, Kise Y, *et al.* Deep learning classification performance for diagnosing condylar osteoarthritis in patients with dentofacial deformities using panoramic temporomandibular joint projection images. Oral Radiol 2024; 40(4): 538-45.
http://dx.doi.org/10.1007/s11282-024-00768-0 PMID: 38990220

[29] Jung W, Lee KE, Suh BJ, Seok H, Lee DW. Deep learning for osteoarthritis classification in temporomandibular joint. Oral Dis 2023; 29(3): 1050-9.
http://dx.doi.org/10.1111/odi.14056 PMID: 34689379

[30] Oliveira RS, Oliveira S, Rodrigues E, Junqueira JL, Panzarella F. Accuracy of panoramic radiography for degenerative changes of the temporomandibular joint. J Int Soc Prev Community Dent 2020; 10(1): 96-100.
http://dx.doi.org/10.4103/jispcd.JISPCD_411_19 PMID: 32181226

[31] Maita I, Dhillon A, Friesen R, Almeida FT. Diagnostic value of panoramic radiographs in the assessment of degenerative joint disease: A retrospective study. Int Dent J 2025; 75(5): 100910.
http://dx.doi.org/10.1016/j.identj.2025.100910 PMID: 40684678

[32] Lee YH, Auh QS, Chun YH, An JS. Age-related radiomorphometric changes on panoramic radiographs. Clin Exp Dent Res 2021; 7(4): 539-51.
http://dx.doi.org/10.1002/cre2.375 PMID: 33305888

[33] Hayat H, Kudrautsau M, Makarov E, *et al.* Toward the autonomous AI doctor: Quantitative benchmarking of an autonomous agentic AI *versus* board-certified clinicians in a real world setting. medRxiv 2025.
http://dx.doi.org/10.1101/2025.07.14.25331406

[34] Shen J, Zhang C J P, Jiang B, *et al.* Artificial intelligence *versus* clinicians in disease diagnosis: Systematic review. JMIR Med Inform 2019; 7(3): e10010.
http://dx.doi.org/10.2196/10010 PMID: 31420959

[35] Razzaki S, Baker A, Perov Y, *et al.* A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. arXiv 2018.
http://dx.doi.org/10.48550/arXiv.1806.10698

[36] Takita H, Kabata D, Walston SL, *et al.* A systematic review and meta-analysis of diagnostic performance comparison between generative AI and physicians. NPJ Digit Med 2025; 8(1): 175.
http://dx.doi.org/10.1038/s41746-025-01543-z PMID: 40121370

[37] Do S, Song KD, Chung JW. Basics of deep learning: A radiologist's guide to understanding published radiology articles on deep learning. Korean J Radiol 2020; 21(1): 33-41.
http://dx.doi.org/10.3348/kjr.2019.0312 PMID: 31920027

[38] Kim D, Choi E, Jeong HG, Chang J, Youm S. Expert system for mandibular condyle detection and osteoarthritis classification in panoramic imaging using R-CNN and CNN. Appl Sci 2020; 10(21): 7464.
http://dx.doi.org/10.3390/app10217464

[39] O'Ryan F, Epker BN. Temporomandibular joint function and morphology: Observations on the spectra of normalcy. Oral Surg Oral Med Oral Pathol 1984; 58(3): 272-9.
http://dx.doi.org/10.1016/0030-4220(84)90052-5 PMID: 6592523