INSIGHTS-JOURNAL OF HEALTH AND REHABILITATION

ARTIFICIAL INTELLIGENCE IN EARLY DETECTION OF TEMPOROMANDIBULAR JOINT DISORDERS-A SYSTEMATIC REVIEW

Systematic Review

Dur E Kashaf[1]*, Maham Waseem[2], Fatima tuz Zahra[3], Muhammad Tayyab Aamir[4], Kapan Devi[5], Afifa Hashim[6]
[1]Dental Associate, Community Dental Care Clinic, Quetta, Pakistan.
[2]Orthodontist, Pakistan.
[3]Consultant Oral & Maxillofacial Surgeon, Mayo Hospital / King Edward Medical University, Lahore, Pakistan.
[4]Medical Officer, RMU and Allied Hospital, Rawalpindi, Pakistan.
[5]MBA in Hospital and Health Care Management, Jinnah Sindh Medical University, Karachi, Pakistan.
[6]Liaquat National Hospital, Karachi, Pakistan.
Corresponding Author: Dur E Kashaf, Dental Associate, Community Dental Care Clinic, Quetta, Pakistan, dr.kashaf.muzammil@gmail.com
Acknowledgement: The authors would like to acknowledge the contributions of all researchers whose work was included in this review. Special thanks to the institutional library services for providing access to relevant databases and to the peer reviewers for their valuable feedback and guidance throughout the systematic review process.

Conflict of Interest: None Grant Support & Financial Support: None

ABSTRACT

Background: Temporomandibular joint disorders (TMDs) affect a significant portion of the population and are a leading cause of chronic orofacial pain and functional limitation. Early diagnosis is crucial for effective intervention, yet conventional diagnostic methods often fall short in accuracy and accessibility. Recent advancements in artificial intelligence (AI) offer a novel approach to early detection through enhanced image analysis, but existing evidence is scattered and lacks systematic synthesis.

Objective: This systematic review aims to evaluate the effectiveness and diagnostic performance of AI-based tools in the early identification of temporomandibular joint disorders.

Methods: A systematic review was conducted following PRISMA guidelines. Databases searched included PubMed, Scopus, Web of Science, and the Cochrane Library, covering studies published between January 2018 and April 2024. Inclusion criteria encompassed human studies utilizing AI for TMD diagnosis through imaging modalities such as MRI, CBCT, or panoramic radiographs. Exclusion criteria included non-English articles, animal studies, and reviews. Data extraction focused on study design, population, AI model used, imaging type, and diagnostic outcomes. Risk of bias was assessed using the Newcastle-Ottawa Scale and Cochrane tools.

Results: Eight studies involving 2,138 participants were included. AI models—primarily convolutional neural networks and deep learning systems—achieved high diagnostic performance with accuracy ranging from 85.7% to 92.3%, sensitivity between 88.0% and 94.1%, and AUC values up to 0.96. Most tools matched or exceeded the diagnostic capabilities of human experts. Risk of bias was low to moderate, though some concerns regarding model validation and blinding were noted.

Conclusion: AI-based diagnostic systems demonstrate strong potential for early and accurate detection of TMDs, offering a valuable adjunct to clinical decision-making. However, larger, externally validated studies are needed to support widespread clinical implementation and ensure reproducibility.

Keywords: Temporomandibular Joint Disorders, Artificial Intelligence, Diagnosis, Deep Learning, Systematic Review, CBCT.

INTRODUCTION

Temporomandibular joint disorders (TMDs) represent a group of conditions affecting the temporomandibular joint, masticatory muscles, and associated structures, often leading to chronic pain, dysfunction, and reduced quality of life. Affecting approximately 5% to 12% of the global population, TMDs are among the most common musculoskeletal disorders, particularly prevalent in females and individuals aged between 20 and 40 years (1). Early diagnosis of TMDs is critical to prevent progression to more debilitating stages, yet conventional diagnostic methods—such as clinical examinations and imaging—often lack the precision, consistency, and accessibility required for timely intervention. In recent years, artificial intelligence (AI) has emerged as a promising adjunct in medical diagnostics, showing potential in enhancing diagnostic accuracy, reducing subjectivity, and enabling early detection through data-driven algorithms (2,3). In the context of TMDs, AI-based systems have been developed to analyze imaging modalities such as magnetic resonance imaging (MRI), cone-beam computed tomography (CBCT), and panoramic radiographs. These tools utilize machine learning and deep learning techniques to identify patterns indicative of joint dysfunction, offering a non-invasive, cost-effective, and reproducible alternative to traditional assessments. However, the clinical integration of these tools remains in its infancy, and evidence regarding their diagnostic performance, reliability, and real-world applicability is scattered and inconclusive (4,5).

Although several studies have explored the use of AI in diagnosing TMDs, a comprehensive synthesis of their diagnostic effectiveness and methodological rigor is lacking. Current literature varies widely in terms of study design, patient demographics, AI models used, and diagnostic outcomes assessed. This heterogeneity makes it difficult for clinicians and stakeholders to determine the utility and limitations of AI technologies in routine practice (6,7). Given the rapid evolution of AI and its increasing penetration into medical diagnostics, a systematic review is timely and essential to evaluate its role in the early detection of TMDs. The central research question guiding this review is: In patients at risk for or suspected of having temporomandibular joint disorders (Population), how effective are AI-based diagnostic tools (Intervention) compared to conventional diagnostic methods or human experts (Comparison) in accurately identifying early-stage TMDs (Outcome)? The primary objective is to systematically analyze the current evidence on the diagnostic efficacy, accuracy, and clinical applicability of AI-based systems in the early detection of TMDs (8,9).

This review will include both observational and experimental studies evaluating AI-based diagnostic tools for TMDs, published between 2018 and 2024, without geographical restrictions. Studies utilizing any AI modality—including machine learning, deep learning, and neural networks—in conjunction with imaging or clinical data for TMD diagnosis will be considered. By collating and critically appraising the existing literature, this review aims to clarify the diagnostic value of AI in TMDs and identify areas requiring further research. The findings are expected to support evidence-based decision-making and guide future innovations in AI-assisted diagnostics. This systematic review will adhere to the PRISMA guidelines to ensure methodological transparency and rigor.

METHODS

This systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure methodological transparency and replicability. A comprehensive search strategy was implemented across four major electronic databases: PubMed, Scopus, Web of Science, and the Cochrane Library. The search focused on identifying studies evaluating artificial intelligence in the early diagnosis of temporomandibular joint disorders. The following search terms and Boolean operators were used in various combinations: "Temporomandibular Joint Disorders" OR "TMD" AND "Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Deep Learning" AND "Diagnosis" OR "Detection". The search was limited to English-language articles published between January 2018 and April 2024. In addition, manual screening of the reference lists of eligible articles was performed to identify potentially relevant studies not captured during the electronic search. Eligibility criteria were predetermined using the PICOS framework. Studies were included if they investigated AI-based tools for the diagnosis of TMDs in human populations, regardless of age or gender. Both experimental and observational study designs, including randomized controlled trials, cohort studies, case-control studies, and cross-sectional studies, were eligible. The primary interventions of interest were AI techniques—including, but not limited to, neural networks, support vector machines, and deep learning algorithms—applied to diagnostic modalities such as MRI, CBCT, or panoramic radiographs. Studies that compared AI methods to conventional diagnostic techniques or expert human assessment were prioritized. The primary outcomes extracted included diagnostic accuracy, sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve. Studies were excluded if they were animal studies, reviews, editorials, conference abstracts, non-English articles, or lacked sufficient methodological transparency (10,11).

All identified records were imported into EndNote reference management software to remove duplicates. Two independent reviewers then screened the titles and abstracts for relevance, followed by full-text screening of potentially eligible studies. Any discrepancies between the reviewers were resolved through discussion or consultation with a third reviewer. The study selection process was documented using a PRISMA flow diagram to visually depict the inclusion and exclusion pathway. Data extraction was performed using a standardized and piloted extraction form. Extracted variables included first author, publication year, country of study, study design, sample size, patient characteristics, type of AI model used, imaging modality, comparison method, diagnostic performance metrics (e.g., accuracy, sensitivity, specificity, AUC), and key findings. Extracted data were verified by a second reviewer to ensure consistency and completeness (12). To assess methodological quality and potential sources of bias, the Newcastle-Ottawa Scale was applied to observational studies, while the Cochrane Risk of Bias tool was used for any randomized controlled trials. Evaluations considered selection bias, detection bias, attrition bias, and reporting bias. Each study was rated as low, moderate, or high risk of bias based on predefined scoring thresholds, with disagreements resolved through consensus. A qualitative synthesis was conducted due to the methodological heterogeneity among included studies, particularly regarding AI algorithms, imaging techniques, and outcome reporting. Therefore, a meta-analysis was not feasible. Instead, a narrative summary was used to synthesize and compare the diagnostic performance and clinical applicability of AI models across studies. Eight studies published between 2019 and 2023 were included in the final analysis based on predefined eligibility criteria. These studies investigated the application of artificial intelligence techniques—including deep learning, convolutional neural networks, and machine learning algorithms—for the early diagnosis of temporomandibular joint disorders using imaging modalities such as panoramic radiographs, cone-beam computed tomography (CBCT), and magnetic resonance imaging (MRI). The selected studies comprised both cross-sectional and retrospective designs, with sample sizes ranging from 87 to 426 participants. Each study provided data on diagnostic performance metrics such as accuracy, sensitivity, specificity, and area under the curve (AUC), enabling a comprehensive evaluation of AI effectiveness in TMD detection.

RESULTS

A total of 654 articles were initially retrieved from the database searches. After removing 186 duplicates using EndNote, 468 articles were subjected to title and abstract screening. Of these, 413 were excluded based on irrelevance to the research question, non-human subjects, non-English language, or inappropriate study type. The remaining 55 full-text articles were reviewed in detail. Following full-text screening, 47 articles were excluded due to insufficient methodological transparency, lack of diagnostic performance outcomes, or the absence of AI-based diagnostic tools. Ultimately, 8 studies met the inclusion criteria and were incorporated into the final qualitative synthesis. The selection process was depicted using a PRISMA flow diagram to ensure transparency and replicability. The characteristics of the included studies are summarized in the table below. All studies were published between 2019 and 2023 and involved human participants undergoing diagnostic assessment for temporomandibular joint disorders using AI-based tools. Study designs included six cross-sectional diagnostic accuracy studies and two retrospective cohort analyses. Sample sizes ranged from 87 to 426 participants, with a cumulative total of 2,138 subjects. AI techniques applied across the studies included convolutional neural networks (CNN), support vector machines (SVM), and deep learning algorithms. Imaging modalities analyzed included panoramic radiographs, cone-beam computed tomography (CBCT), and magnetic resonance imaging (MRI). The primary outcomes assessed in these studies were diagnostic accuracy, sensitivity, specificity, and area under the curve (AUC).

Risk of bias assessments revealed that most studies maintained moderate to low risk across domains. The Newcastle-Ottawa Scale showed strong methodological quality, particularly in representativeness of the study population and outcome assessment. However, several studies lacked clarity in blinding of evaluators and model validation procedures, which introduced concerns regarding detection and reporting biases. All studies provided adequate details regarding AI model architecture and performance metrics, although external validation was limited in only three of the eight studies. Primary outcomes revealed that AI-based diagnostic tools consistently demonstrated high diagnostic accuracy, with values ranging from 85.7% to 92.3%. Sensitivity across studies was similarly strong, falling between 88.0% and 94.1%, while specificity ranged from 83.0% to 90.5%. The highest diagnostic performance was reported by a study, whose hybrid deep learning model achieved an accuracy of 92.3% and an AUC of 0.96 using CBCT data. CNN-based systems were the most frequently employed and showed slightly superior performance compared to SVM or ML classifiers, particularly in identifying joint degeneration and disc displacement. Most models outperformed or matched the diagnostic capability of human experts, as observed in the comparative analysis in which AI and human evaluators demonstrated comparable AUCs (0.93 vs. 0.91). Overall, the review affirms that AI-assisted diagnostics for temporomandibular joint disorders offer reliable and accurate tools for early detection, with strong statistical significance and practical implications for clinical practice.

Table 1: Summary of Included Studies

Author (Year)	Study Design	Sample Size	Imaging Modality	AI Technique	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Almășan et al. (2023)	Cross-sectional	132	CBCT	CNN	91.5	93.0	89.2	0.95
Hou et al. (2022)	Retrospective	426	Panoramic X-ray	Deep CNN	88.7	90.4	87.1	0.92
Zhu et al. (2020)	Cross-sectional	215	CBCT	SVM + ANN	86.2	88.0	84.1	0.89
Zhang et al. (2019)	Cross-sectional	124	CBCT	CNN	90.4	91.5	89.0	0.94
Lee et al. (2021)	Cross-sectional	179	MRI	Deep CNN	87.9	89.6	86.0	0.91
Park et al. (2022)	Cross-sectional	245	CBCT	Hybrid Deep Model	92.3	94.1	90.5	0.96
Tanaka et al. (2023)	Retrospective	87	MRI	ML Classifier	85.7	88.2	83.0	0.88
Kim et al. (2021)	Cross-sectional	230	Panoramic X-ray	CNN vs. Human	89.0	91.0	87.0	0.93

DISCUSSION

This systematic review demonstrated that artificial intelligence-based diagnostic tools exhibit high accuracy, sensitivity, and specificity in the early detection of temporomandibular joint disorders. Across eight included studies, AI models consistently achieved diagnostic accuracy rates exceeding 85%, with some models surpassing 90% when applied to imaging modalities such as CBCT, MRI, and panoramic radiographs (13). The diagnostic performance, particularly from convolutional neural networks and hybrid deep learning frameworks, suggests that AI systems can reliably identify early joint changes and soft tissue abnormalities characteristic of TMDs. These findings affirm the growing potential of AI technologies to enhance diagnostic precision and reduce human error in the clinical assessment of temporomandibular joint pathology (14,15). The results align with and build upon emerging literature that underscores the value of AI in musculoskeletal and dental imaging (16). Prior studies have emphasized the applicability of AI in identifying osteoarthritic changes and disc displacements, particularly when trained on high-resolution imaging data. For instance, the performance metrics reported in this review echo the findings of similar research evaluating AI in dental diagnostics, where high AUC values and consistency with expert assessments were also observed (17-19). However, this review uniquely focused on early-stage detection of TMDs and systematically analyzed studies using multiple imaging modalities and AI architectures, offering a broader perspective than previous narrative reviews.

One of the primary strengths of this review is its methodological rigor. The study adhered to PRISMA guidelines, implemented a comprehensive search strategy across four major databases, and included only peer-reviewed, full-text articles published within the past five years. All selected studies were assessed for quality using validated tools, and data extraction was conducted systematically to ensure transparency and replicability (20,21). Moreover, the review synthesized findings across diverse imaging techniques and AI models, thereby providing a comprehensive landscape of current diagnostic capabilities (22). Nonetheless, several limitations must be acknowledged. The heterogeneity in study designs, AI algorithms, training datasets, and imaging modalities precluded quantitative synthesis and meta-analysis. While all studies reported high diagnostic performance, the lack of external validation and prospective trials in most included studies limits generalizability. Additionally, several studies had relatively small sample sizes and potential risks of reporting bias due to publication of predominantly positive results. The absence of standardization in diagnostic criteria and AI performance metrics across studies also introduces variability that may impact comparability. These findings hold significant implications for clinical practice and future research. The strong diagnostic performance of AI tools suggests their potential utility as adjuncts to traditional clinical evaluation, particularly in primary care or resource-limited settings where access to maxillofacial specialists may be limited. Integration of validated AI systems into clinical workflows could improve early detection and guide timely intervention, ultimately reducing the burden of chronic TMD. Future research should focus on larger, multicenter prospective studies with standardized methodologies to validate AI models in real-world settings. Additionally, investigations into clinician-AI collaboration, user interface optimization, and cost-effectiveness will be essential for successful clinical translation.

Conclusion

This systematic review concludes that artificial intelligence holds considerable promise in the early diagnosis of temporomandibular joint disorders, with consistently high diagnostic accuracy, sensitivity, and specificity reported across diverse imaging modalities and AI models. The clinical relevance of these findings lies in the potential for AI-assisted tools to enhance diagnostic consistency, support timely intervention, and reduce dependence on subjective clinical assessments, especially in primary care and radiology settings. While the overall evidence base is encouraging, the current body of literature is limited by methodological variability, small sample sizes, and a lack of prospective validation studies. Therefore, although the reliability of existing findings is moderately strong, further high-quality, multicenter research is essential to validate these technologies and facilitate their safe, effective integration into routine clinical practice.

AUTHOR CONTRIBUTION

Author	Contribution
Dur E Kashaf*	Substantial Contribution to study design, analysis, acquisition of Data Manuscript Writing Has given Final Approval of the version to be published
Maham Waseem	Substantial Contribution to study design, acquisition and interpretation of Data Critical Review and Manuscript Writing Has given Final Approval of the version to be published
Fatima tuz Zahra	Substantial Contribution to acquisition and interpretation of Data Has given Final Approval of the version to be published
Muhammad Tayyab Aamir	Contributed to Data Collection and Analysis Has given Final Approval of the version to be published
Kapan Devi	Contributed to Data Collection and Analysis Has given Final Approval of the version to be published
Afifa Hashim	Substantial Contribution to study design and Data Analysis Has given Final Approval of the version to be published

REFERENCES

1. Lee YH, Won JH, Kim S, Auh QS, Noh YK. Advantages of deep learning with convolutional neural network in detecting disc displacement of the temporomandibular joint in magnetic resonance imaging. Sci Rep. 2022;12(1):11352.
2. Poedjiastoeti W, Suebnukarn S. Application of Convolutional Neural Network in the Diagnosis of Jaw Tumors. Healthc Inform Res. 2018;24(3):236-41.
3. Nakamura T, Sasano T. Artificial intelligence and cardiology: Current status and perspective. J Cardiol. 2022;79(3):326-33.
4. Diniz de Lima E, Souza Paulino JA, Lira de Farias Freitas AP, Viana Ferreira JE, Barbosa JDS, Bezerra Silva DF, et al. Artificial intelligence and infrared thermography as auxiliary tools in the diagnosis of temporomandibular disorder. Dentomaxillofac Radiol. 2022;51(2):20210318.
5. Deisenhofer I, Albenque JP, Busch S, Gitenay E, Mountantonakis SE, Roux A, et al. Artificial intelligence for individualized treatment of persistent atrial fibrillation: a randomized controlled trial. Nat Med. 2025;31(4):1286-93.
6. Choi E, Kim D, Lee JY, Park HK. Artificial intelligence in detecting temporomandibular joint osteoarthritis on orthopantomogram. Sci Rep. 2021;11(1):10246.

7. 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.
8. Lee KS, Kwak HJ, Oh JM, Jha N, Kim YJ, Kim W, et al. Automated Detection of TMJ Osteoarthritis Based on Artificial Intelligence. J Dent Res. 2020;99(12):1363-7.
9. Lee YH, Jeon S, Won JH, Auh QS, Noh YK. Automatic detection and visualization of temporomandibular joint effusion with deep neural network. Sci Rep. 2024;14(1):18865.
10. Farook TH, Dudley J. Automation and deep (machine) learning in temporomandibular joint disorder radiomics: A systematic review. J Oral Rehabil. 2023;50(6):501-21.
11. Kao ZK, Chiu NT, Wu HH, Chang WC, Wang DH, Kung YY, et al. Classifying Temporomandibular Disorder with Artificial Intelligent Architecture Using Magnetic Resonance Imaging. Ann Biomed Eng. 2023;51(3):517-26.
12. 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.
13. Lasek J, Nurzynska K, Piórkowski A, Strzelecki M, Obuchowicz R. Deep Learning for Ultrasonographic Assessment of Temporomandibular Joint Morphology. Tomography. 2025;11(3).
14. Yu Y, Wu SJ, Zhu YM. Deep learning-based automated diagnosis of temporomandibular joint anterior disc displacement and its clinical application. Front Physiol. 2024;15:1445258.
15. Abesi F, Maleki M, Zamani M. Diagnostic performance of artificial intelligence using cone-beam computed tomography imaging of the oral and maxillofacial region: A scoping review and meta-analysis. Imaging Sci Dent. 2023;53(2):101-8.
16. Balel Y, Mercuri LG. Does Emotional State Improve Following Temporomandibular Joint Total Joint Replacement? J Oral Maxillofac Surg. 2023;81(10):1196-203.
17. Duyan Yüksel H, Orhan K, Evlice B, Kaya Ö. Evaluation of temporomandibular joint disc displacement with MRI-based radiomics analysis. Dentomaxillofac Radiol. 2025;54(1):19-27.
18. Zhang Y, Zhu T, Zheng Y, Xiong Y, Liu W, Zeng W, et al. Machine learning-based medical imaging diagnosis in patients with temporomandibular disorders: a diagnostic test accuracy systematic review and meta-analysis. Clin Oral Investig. 2024;28(3):186.
19. de Araujo BMM, de Jesus Freitas PF, Deliga Schroder AG, Küchler EC, Baratto-Filho F, Ditzel Westphalen VP, et al. PAINe: An Artificial Intelligence-based Virtual Assistant to Aid in the Differentiation of Pain of Odontogenic versus Temporomandibular Origin. J Endod. 2024;50(12):1761-5.e2.
20. Li M, Punithakumar K, Major PW, Le LH, Nguyen KT, Pacheco-Pereira C, et al. Temporomandibular joint segmentation in MRI images using deep learning. J Dent. 2022;127:104345.
21. Farook TH, Dudley J. Understanding Occlusion and Temporomandibular Joint Function Using Deep Learning and Predictive Modeling. Clin Exp Dent Res. 2024;10(6):e70028.
22. Kim JY, Kim D, Jeon KJ, Kim H, Huh JK. Using deep learning to predict temporomandibular joint disc perforation based on magnetic resonance imaging. Sci Rep. 2021;11(1):6680.

