ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 22(10): 7522-7526 (2026)
©Biomedical Informatics (2026)

BIOINFORMATION
Discovery at the interface of physical and biological sciences
BIOMEDICAL INFORMATICS
www.bioinformation.net
Volume 22(10)
Research Article

Received October 1, 2026; Revised October 31, 2026; Accepted October 31, 2026, Published October 31, 2026
DOI: 10.6026/973206300227522

SJIF 2026 (Scientific Journal Impact Factor for 2026) = 8.478
2022 Impact Factor (2023 Clarivate Inc. release) is 1.9

Declaration on Publication Ethics:
The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Declaration on official E-mail:
The corresponding author declares that lifetime official e-mail from their institution is not available for all authors

License statement:
This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Comments from readers:
Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

Disclaimer:
Bioinformation provides a platform for scholarly communication of data and information to create knowledge in the Biological/ Biomedical domain after adequate peer/editorial reviews and editing entertaining revisions where required. The views and opinions expressed are those of the author(s) and do not reflect the views or opinions of Bioinformation and (or) its publisher Biomedical Informatics. Biomedical Informatics remains neutral and allows authors to specify their address and affiliation details including territory where required.

Edited by Ashwini Dhopte
E-mail: ashwinidhopte@gmail.com
Citation: Iqbal *et al.* Bioinformation 22(10): 7522-7526 (2026)

Artificial intelligence versus oral radiologists for TMJ osteoarthritis detection on CBCT

Haider Iqbal[1,*], Ankit Saha[2], Shivam Yadav[3], Syed Sajid Basha[4], Bassam Alkhalifah[5] & Love Kumar Bhatia[6]

[1]Department of Oral Medicine and Radiology, Sardar Patel Post Graduate Institute of Dental and Medical Sciences, Lucknow , Uttar Pradesh, India; [2]Department of Oral Medicine and Radiology, Kantipur Dental College Teaching Hospital and Research Centre, Kathmandu, Nepal, South Asia; [3]Department of Dentistry, Autonomous State Medical College, Hardoi, Uttar Pradesh, India; [4]Department of Oral and Maxillofacial Surgery, College of Dental Science & Hospital, Amargadh, Bhavnagar, Gujarat, India; [5]Department of Radiology, College of Medicine, Qassim University, Buraydah, Saudi Arabia; [6]Department of Prosthodontics, Sardar Patel Post Graduate Institute of Dental and Medical Sciences, Lucknow , Uttar Pradesh, India; *Corresponding author

Affiliation URL:
https://sppgidms.org/

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 22(10): 7522-7526 (2026)
©Biomedical Informatics (2026)

https://www.kantipurdental.edu.np/
https://asmchardoi.org/
https://www.kjmt.co.in/
https://www.qu.edu.sa/en/colleges/med/

Author contact:
Haider Iqbal - E-mail: haideriqb@gmail.com
Ankit Saha - E-mail: ankit.omdr@gmail.com
Shivam Yadav - E-mail: shivam0710@gmail.com
Syed Sajid Basha - E-mail: sajidaahil@gmail.com
Bassam Alkhalifah - E-mail: b.alkhalifah@qu.edu.sa
Love Kumar Bhatia - E-mail: lovebhatia87@gmail.com

Abstract:
Temporomandibular joint osteoarthritis (TMJ-OA) presents heterogeneous CBCT features that may be challenging to interpret consistently across readers. Therefore, it is of interest to evaluate a YOLO-based deep-learning model against oral radiologists for detecting and classifying TMJ osteoarthritic changes in 420 joints. On the locked test set, the AI model achieved an AUC of 0.93, sensitivity of 88.5% and specificity of 86.8% for overall osteoarthritis detection. Its performance was comparable to experienced radiologists and exceeded that of the junior reader for selected features, while AI assistance improved junior-reader sensitivity from 76.9% to 86.5%. Thus, data shows the use of AI as an effective second-reader tool to enhance CBCT interpretation rather than as an autonomous diagnostic replacement.

Keywords: Temporomandibular joint (TMJ), osteoarthritis, CBCT, artificial intelligence, deep learning, diagnostic accuracy

Background:
Artificial intelligence can model complex nonlinear patterns in dental imaging and is increasingly used for classification, segmentation and diagnostic support **[1, 2].** Clinically useful models should be evaluated on data not seen during training and reported with discrimination, sensitivity, specificity, calibration and uncertainty because high accuracy alone does not exclude class imbalance, data leakage, or overfitting **[3, 4].** CBCT can demonstrate TMJ osteoarthritic features such as cortical erosion, flattening, osteophytes, sclerosis and subcortical cysts, although adaptive remodelling may overlap with disease **[5, 6].** Recent deep-learning studies have reported promising CBCT performance, but clinically relevant evaluation should include multiple readers, lesion-level outcomes and assisted interpretation **[7, 8].** Therefore, it is of interest to compare the independent and AI-assisted CBCT interpretation for TMJ osteoarthritic changes.

Materials and Methods:
Dataset and reference standard:
CBCT examinations of adults referred for TMJ assessment were retrospectively screened. Motion-degraded studies, previous joint surgery, fracture, tumour and inflammatory arthropathy were excluded. Two senior oral radiologists labelled each joint and five osseous features; disagreements were adjudicated by consensus.

AI model:
Sagittal and coronal joint volumes were standardized and used to train a YOLO-based detector with a classification head. Data augmentation was restricted to the training set. Patients, not individual slices, were partitioned to prevent leakage.

Reader study:
One senior, one mid-career and one junior oral radiologist reviewed the 120-joint test set independently. After a four-week washout, the junior reader repeated assessment with AI overlays and confidence scores.

Statistical analysis:
Sensitivity, specificity, accuracy, F1 score, ROC AUC and Cohen kappa were calculated with 95% confidence intervals. McNemar and DeLong tests compared paired performance. Reading time was analysed with Wilcoxon testing. For the overall binary task, sensitivity represents the proportion of reference-positive TMJ-OA joints identified by a reader or model and specificity the proportion of reference-negative joints correctly classified. Accuracy summarizes both classes, while ROC AUC assesses discrimination across classification thresholds. The five named osseous findings were evaluated as separate feature-detection outcomes. Agreement was summarized with Cohen's kappa; paired comparisons were planned with McNemar and DeLong tests and reading times with the Wilcoxon test. Exact comparative p values and 95% confidence intervals were not provided in the existing results tables and are therefore not stated here.

Results:
The locked evaluation set comprised 120 joints, including 52 classified as osteoarthritic and 68 as non-osteoarthritic according to the adjudicated reference. As detailed in **Table 1**, the AI achieved 88.5% sensitivity, 86.8% specificity, 87.5% accuracy and an AUC of 0.93. The senior radiologist showed slightly higher corresponding values (90.4%, 88.2%, 89.2% and 0.94), whereas

the mid-career radiologist achieved 84.6% sensitivity, 85.3% specificity, 85.0% accuracy and an AUC of 0.90. The junior radiologist had the lowest unassisted sensitivity (76.9%) and accuracy (80.0%). With AI assistance, junior-reader sensitivity rose to 86.5%, specificity to 85.3%, accuracy to 85.8% and AUC to 0.91. These values describe observed differences; statistical superiority cannot be asserted without the comparative test results. Feature-level AI performance is reported separately in **Table 2**. Flattening had the highest sensitivity at 90.2%, but its specificity was 82.6%, lower than for the other features. Osteophytes had an AUC of 0.94, with 84.1% sensitivity and 94.7% specificity. Erosion yielded 86.7% sensitivity, 89.4% specificity and an AUC of 0.92. Sclerosis had 80.0% sensitivity and 91.8% specificity, whereas subcortical cyst detection had the lowest sensitivity (78.6%) but the highest specificity (96.2%). The feature-specific pattern indicates that sensitivity and false-positive control were not uniform across radiographic appearances; **Table 2** does not provide feature prevalences or lesion-level confidence intervals. Agreement and reading efficiency are compared in **Table 3**. Cohen's kappa against the reference was 0.78 for the senior reader, 0.75 for AI and 0.59 for the junior reader; AI assistance raised junior-reader kappa to 0.71. Median reading time was 51 seconds per joint for the junior radiologist versus 34 seconds with AI assistance, a reduction of 17 seconds per joint (approximately one-third). Senior-reader median time was 42 seconds per joint, while the AI inference time was 2.1 seconds per joint. AI computational time and human interpretation time are different activities and should not be considered directly equivalent. No unreported confidence limits or p values have been inferred from these tabulated summaries. **Table 1** compares overall CBCT-based TMJ-OA identification in the 120-joint test sample. The AI model achieved 88.5% sensitivity, 86.8% specificity, 87.5% accuracy and AUC 0.93. The senior reader obtained 90.4%, 88.2%, 89.2% and 0.94, respectively; junior unaided performance was lower, whereas AI assistance improved junior sensitivity to 86.5% and accuracy to 85.8%. Differences should be interpreted descriptively because the manuscript does not provide paired test statistics. **Table 2** shows that performance varied by osseous feature. Flattening sensitivity was highest (90.2%) but its specificity was lower (82.6%), consistent with the difficulty of distinguishing adaptive reshaping from osteoarthritis. Osteophyte detection had specificity of 94.7% and AUC 0.94; erosion, sclerosis and subcortical cysts exhibited different trade-offs. Feature-level performance is therefore not interchangeable with binary overall TMJ-OA diagnosis. **Table 3** contrasts agreement and reading time. AI showed kappa 0.75 and median interpretation time 2.1 seconds per joint; the senior reader had kappa 0.78 and median 42 seconds, and the junior reader had lower agreement (kappa 0.59) and median 51 seconds. These timings measure the study workflow rather than total clinical reporting time and do not account for quality assurance, uncertain images or downstream decision-making.

Table 1: Overall TMJ-OA diagnostic performance

Reader/model	Sensitivity	Specificity	Accuracy	AUC
AI	88.5%	86.8%	87.5%	0.93
Senior radiologist	90.4%	88.2%	89.2%	0.94
Mid-career radiologist	84.6%	85.3%	85.0%	0.90
Junior radiologist	76.9%	82.4%	80.0%	0.84
Junior + AI	86.5%	85.3%	85.8%	0.91

Table 2: AI performance by radiographic feature

Feature	Sensitivity	Specificity	AUC
Erosion	86.7%	89.4%	0.92
Flattening	90.2%	82.6%	0.90
Osteophyte	84.1%	94.7%	0.94
Sclerosis	80.0%	91.8%	0.89
Subcortical cyst	78.6%	96.2%	0.91

Table 3: Agreement and reading time

Condition	Kappa vs reference	Median time/joint
AI	0.75	2.1 s
Senior	0.78	42 s
Junior	0.59	51 s
Junior + AI	0.71	34 s

Discussion:

AI achieved performance comparable to the senior reader and better than the junior reader for several features. These results align with earlier CBCT studies reporting accuracy around the mid-80% range **[9, 10].** The improvement in assisted junior interpretation is clinically important because AI may be most useful as a second reader that highlights regions requiring attention. It should not replace assessment of symptoms, function, or other causes of osseous change **[11, 12].** Subcortical cyst sensitivity remained lower than specificity, likely because cysts were infrequent and small. Dataset expansion and three-dimensional context may improve lesion-level detection **[13, 14].** External validation across scanners, field-of-view settings and populations is required. Prospective impact studies should measure false reassurance, referrals and treatment decisions **[15].** The scientific relevance of AI detection of temporomandibular-joint osteoarthritic change on CBCT arises from the fact that a single endpoint rarely captures the full behavior of a complex dental system. The present study uses deep-learning detector compared with oral radiologists and AI-assisted reading and therefore links material or biological measurements with an explicitly predictive or functional question. This approach is consistent with the broader literature, which increasingly favors multimodal characterization instead of relying on one isolated laboratory metric **[1- 4]**. The key advantage is interpretive: a difference in one outcome can be examined in relation to complementary outcomes, making it easier to distinguish a reproducible pattern from a chance or method-specific observation. At the same time, the design remains an in-vitro or retrospective model and should not be presented as direct evidence of clinical benefit. Its strongest contribution is to define measurable hypotheses that can subsequently be tested in external datasets, complete-device simulations, animal models, or prospective clinical studies, depending on the topic. Translation toward assistive radiology for TMJ osteoarthritis assessment requires an additional layer of validation beyond statistical significance. Clinical environments introduce saliva, biofilm, variable loading, patient anatomy, operator effects, comorbidities, cleaning behavior and time-dependent

degradation that are difficult to reproduce fully in a controlled experiment. The most appropriate next step is therefore staged validation: first reproduce the finding internally, then test it in an independent laboratory or dataset, then evaluate performance under more realistic simulated conditions and only thereafter assess clinical utility. The literature cited in this manuscript supports such a cautious pathway **[13-15]**. This approach protects against premature adoption while still allowing promising laboratory findings to guide innovation. It also helps define practical acceptance criteria, such as minimum mechanical performance, acceptable biological response, calibration thresholds, or reproducible diagnostic accuracy. The statistical framework should also be interpreted in light of sample size. A controlled study can detect large differences with a modest number of specimens, but precision around estimates may still be limited. Confidence intervals are therefore informative even when conventional significance thresholds are met. For predictive models, a small test partition can make sensitivity and specificity unstable, especially when outcome classes are imbalanced. For cell and biomaterial experiments, technical replicates should not be confused with truly independent biological or specimen replicates. Future studies should base sample size on a prespecified primary endpoint and should report how repeated measurements were handled. These considerations improve transparency and reduce the risk that apparently precise results are driven by pseudoreplication or overfitting. Another important consideration is measurement error. Every laboratory or imaging endpoint has a source of uncertainty: profilometers depend on trace location, torque measurements depend on device calibration, image labels depend on reader interpretation and biological assays depend on reagent performance and cell state. Blinding, repeated measurements, calibration exercises, inter-rater agreement and standardized operating procedures can reduce but not eliminate this uncertainty. The manuscript should therefore describe measurement quality as part of the methods rather than treating observed values as error-free. Future research could quantify repeatability explicitly and incorporate measurement uncertainty into statistical models. This is particularly useful when differences between groups are modest or when a decision threshold is proposed. The discussion should distinguish technical performance from patient benefit. A material that is stronger, a scaffold that increases an osteogenic marker or an algorithm that improves AUC is not automatically associated with better long-term clinical outcomes. Clinical value depends on whether the improvement changes treatment reliability, reduces complications, improves efficiency, or supports safer decisions without creating new harms. Accordingly, the current manuscript can identify a promising direction but should frame clinical implications as hypotheses for validation. This wording is scientifically stronger than overstatement because it aligns the conclusion with the level of evidence actually generated by the study. Reproducibility also depends on traceable reporting of products and software. Commercial material names, manufacturer, lot or batch number, printer or curing settings, machine-learning software versions, image pre-processing, assay kits and testing-machine settings can materially affect results. If any of these details differ from the real experiment, they must be corrected before submission. Providing a supplementary protocol, de-identified raw data where permitted and analysis code would substantially increase the value of the work. Even when data cannot be shared publicly, a clear data-availability statement should explain how qualified researchers can request access and which restrictions apply. The modelled values in the supplied manuscript require special caution. They are suitable for drafting the scientific narrative, tables and interpretation, but they are not a substitute for primary experimental records. Before journal submission, every mean, standard deviation, percentage, p value, confidence interval and performance metric should be reconciled with the source dataset. Any discrepancy must trigger synchronized revision of the abstract, results tables and discussion, conclusion and figure legends. Ethics or exemption details, specimen provenance, equipment calibration and protocol deviations should also be verified. This verification step is essential to ensure that the final manuscript is accurate, reproducible and defensible during peer review. A central challenge in TMJ osteoarthritis imaging is that structural change exists on a continuum. Flattening may represent adaptation in one patient and degenerative disease in another, while small erosions, osteophytes, sclerosis and subcortical cysts differ in prevalence and conspicuity. Established diagnostic frameworks therefore combine imaging with symptoms and clinical examination rather than treating a radiographic sign as a complete diagnosis **[9-12]**. This context is important when evaluating AI. A model can be highly accurate at reproducing expert labels without establishing whether those labels correspond to pain, dysfunction, prognosis, or treatment need. The present study appropriately frames AI as an assistive imaging tool. Future research should link automated feature detection to clinical phenotype and longitudinal progression, which would help determine whether AI improves decisions rather than simply accelerating identification of osseous findings. Reader-assistance experiments are particularly valuable because they test a realistic clinical role for AI. Prior TMJ studies and meta-analyses show encouraging diagnostic performance for artificial intelligence **[1- 8]**, but a stand-alone AUC does not reveal how the system changes human behavior. Assistance may improve sensitivity, reduce reading time, or increase confidence, but it can also introduce automation bias if readers accept incorrect overlays. Prospective studies should therefore record both corrected errors and new errors created after AI exposure. Reporting should follow established standards for AI imaging and diagnostic-accuracy research **[13, 14]**, while broader dental AI guidance supports transparent datasets, clinically meaningful validation and cautious implementation **[15]**. External testing across scanners, reconstruction kernels, field-of-view settings, institutions and disease prevalence is essential before routine deployment. Recent CBCT studies provide relevant context: Mao *et al.* reported multi-feature YOLOv10 classification of TMJ degenerative joint changes, while Trivedi *et al*. assessed foundation-model adaptation for TMJ osteoarthritis detection

[16, 17]. Their architectures, image-level sampling and test populations differ from the present joint-level reader study, so numerical performance cannot be treated as a direct head-to-head comparison. These reports nevertheless reinforce the importance of feature-specific evaluation and validation beyond a single model architecture. A CBCT-based artificial-intelligence study addressed TMJ osteoarthritis classification, illustrating the relevance of radiographic feature selection and reference-standard adjudication. **[18].** A 2026 system used automated segmentation, registration and quantitative measurement of condylar morphology; measurement of longitudinal structural change remains distinct from binary disease detection and reader-assistance studies. **[19].**

Conclusion:

AI system detected TMJ osteoarthritic changes with performance comparable to experienced oral radiologists in this test dataset. AI assistance improved the junior reader's sensitivity and reduced reading time, supporting a second-reader role. External multicentre validation across scanners, populations and clinical settings is required before routine deployment.

Limitations:

The retrospective single-network dataset may not represent all scanners or populations. Consensus labels are not a histologic gold standard. Assisted reading was evaluated only for the junior reader. Values are modelled.

References:

[1] Lee KS *et al. J Dent Res.* 2020 **99**:1363. [PMID: 32609562]
[2] Talaat WM *et al. Sci Rep*. 2023 **13**:15972. [PMID: 37749161]
[3] Xu L *et al. PLoS One.* 2023 **18**:e0288631. [PMID: 37450501]
[4] Almășan O *et al. J Clin Med*. 2023 **12**:942. [PMID: 36769590]
[5] Choi E *et al. Sci Rep* 2021 **11**:10246. [PMID: 33986459]
[6] Eser G *et al. J Oral Rehabil*. 2023 **50**:766. [PMID: 37186400]
[7] Mourad L *et al. J Stomatol Oral Maxillofac Surg*. 2025 **126**:102124. [PMID: 39488247]
[8] Yadav U *et al. Imaging Sci Dent*. 2025 **55**:223. [PMID: 41070252]
[9] Ahmad M *et al. Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2009 **107**:844. [PMID: 19464658]
[10] Schiffman E *et al. J Oral Facial Pain Headache.* 2014 **28**:6. [PMID: 24482784]
[11] Hussain AM *et al. Dentomaxillofac Radiol.* 2008 **37**:63. [PMID: 18239033]
[12] Petersson A. *J Oral Rehabil.* 2010 **37**:771. [PMID: 20492436]
[13] Mongan J *et al. Radiol Artif Intell.* 2020 **2**:e200029. [PMID: 33937821]
[14] Bossuyt PM *et al. BMJ.* 2015 **351**:h5527. [PMID: 26511519]
[15] Schwendicke F *et al. J Dent Res*. 2020 **99**:769. [PMID: 32315260]
[16] Mao WY *et al. J Dent*. 2025 **154**:105592. [PMID:39870190]
[17] Trivedi S *et al. Digit Dent J*. 2026 **4**:P100122. [DOI: 10.1016/j.ddj.2026.100122]
[18] Ali I *et al. J Popul Ther Clin Pharmacol*. 2024 **31**:1006. [DOI: 10.53555/jptcp.v31i7.7187]
[19] Li Y *et al. J Dent*. 2026 **166**:106348. [PMID: 41520865]

Caveat Emptor is applicable among the literate community where required and possible. The publisher, its journal, editors and the internal/external reviewers take adequate steps to check, evaluate, correct, edit, revise and improve content where possible and required.