diagnostics

MDPI

*Editorial*

# Advancements in Artificial Intelligence for Dentomaxillofacial Radiology—2nd Edition: From Proof of Concept to Generalizable and Clinically Trustworthy AI

**Kaan Orhan [1,2,3,*], Andre Luiz Ferreira Costa [4] and Sérgio Lúcio Pereira de Castro Lopes [5]**

[1] Department of Dentomaxillofacial Radiology, Faculty of Dentistry, Ankara University, Ankara 06560, Turkey
[2] Department of Oral Diagnostics, Faculty of Dentistry, Semmelweis University, 1088 Budapest, Hungary
[3] Medical Design Application and Research Center (MEDITAM), Ankara University, Ankara 06000, Turkey
[4] Postgraduate Program in Dentistry, Cruzeiro do Sul University (UNICSUL), São Paulo 08060-070, SP, Brazil; alfcosta@gmail.com
[5] Department of Diagnosis and Surgery, São José dos Campos School of Dentistry, São Paulo State University (UNESP), São José dos Campos 12245-000, SP, Brazil; sergio.lopes@unesp.br
* Correspondence: knorhan@dentistry.ankara.edu.tr

Only a few years ago, the central question in artificial intelligence (AI) for dentomaxillofacial radiology was whether algorithms could match human experts at narrowly defined diagnostic tasks [1–3]. The thirteen contributions gathered in this second edition of the Special Issue "Advancements in Artificial Intelligence for Dentomaxillofacial Radiology" suggest that the field is now confronting a harder set of questions: whether accuracy survives across devices, institutions, and patient populations; whether AI assistance measurably changes what clinicians do; and whether new classes of models, such as radiomics–deep learning hybrids, lightweight edge networks, and large language models (LLMs), can be trusted at the chairside [4]. This closing editorial shows how the contributions to this edition address these questions and sketches the research agenda that follows from them.

A defining characteristic of this edition is its focus on clinically deployed or commercially available systems rather than laboratory prototypes alone. One contribution evaluated the AI-based Diagnocat system for periapical lesion detection on panoramic radiographs, reporting perfect specificity (1.00) alongside moderate sensitivity (0.78), with performance varying markedly by tooth type—a salient reminder that aggregate metrics can conceal clinically important regional weaknesses [5]. A similar profile has been reported outside this Special Issue: in an external validation of two FDA-cleared decision support systems for caries and periodontal bone loss, both showed high specificity and negative predictive value, supporting their role as adjunctive screening aids rather than stand-alone diagnostic tools [6]. AlGhaihab et al. went a step further in the periodontal domain by asking the consequential question of whether AI assistance actually changes clinician performance. In their reader study with ten general dentists, deep learning support yielded no statistically significant improvement in diagnostic accuracy, with area-under-the-curve differences of only 0.031 for periapical and −0.009 for bitewing radiographs [7]. By contrast, a randomized controlled trial of AI-assisted detection of periapical radiolucencies found a modest gain in accuracy, driven mainly by fewer false-positive findings, with the largest benefit among less experienced dentists [8]. In orthodontics, another contribution to this Special Issue evaluated the clinical utility and time efficiency of AI-supported and app-aided cephalometric analysis platforms in comparison with conventional manual tracing. Together, these studies reframe AI tools as adjuncts whose value must be demonstrated at the level of the clinician–machine team, not at the level of the algorithm alone.

Check for updates

Received: 31 October 2026
Accepted: 5 November 2026
Published: 7 November 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

If the first edition of this Special Issue showcased algorithmic capability, this second edition interrogates its portability. Özcan and Allaf trained four deep learning architecture families on two independently sourced cephalometric datasets and showed that in-domain rankings did not predict cross-domain rankings: the most accurate in-domain model suffered among the largest performance drops, cross-dataset errors rose by 155–426%, and clinically decisive landmarks (A-point, B-point, Nasion, and Menton) exceeded the 2 mm acceptability threshold out-of-domain, whereas balanced multi-source training recovered most of the loss [9]. Another contribution provides a complementary, encouraging example: a U2-Net architecture for detecting separated endodontic instruments, developed on panoramic radiographs pooled from multiple imaging systems, achieved robust segmentation performance (Dice coefficient 0.849), although device-stratified external testing has yet to be performed. Two recent external validation studies of commercial CBCT tooth segmentation software point in the same direction: accuracy was largely maintained across five CBCT systems, with little influence of the scanner itself, whereas patient-related factors such as orthodontic brackets, endodontic treatment, and high-density artifacts increased the need for manual refinement [9,10]. At the evidence-synthesis level, Hashim et al. appraised sixteen publications on deep learning for alveolar bone loss on intraoral radiographs and delivered a sobering verdict: only two studies performed external validation, none was prospective, and the certainty of evidence was rated very low [11].

The methodological and clinical breadth of the field also continues to expand. Çetin et al. fused handcrafted radiomic features with a purpose-built convolutional architecture (RadC-CNN) to differentiate radicular cysts from periapical granulomas on CBCT—entities that are notoriously difficult to separate preoperatively—achieving 90% accuracy against histopathologically confirmed diagnoses [12]. Across jawbone lesions more broadly, a recent meta-analysis reported consistently high specificity but widely varying sensitivity, and noted that limited external validation and transparent reporting still restrict clinical interpretability [13]. Another contribution tested whether general-purpose LLM-based chatbots (GPT-4o, GPT-o4-mini-high, and GPT-o1-pro) could estimate bone age and growth stage directly from hand–wrist radiographs without domain-specific training; the predictions correlated significantly with expert reference values, although agreement remained modest, positioning such models as accessible adjuncts rather than replacements. Exploring the frontier of generative and language-driven artificial intelligence, emerging investigations in dentistry and radiology have begun evaluating general-purpose foundation models and multimodal vision–language architectures for clinical triage, automated interpretation, and decision support [14,15]. However, as emphasized in recent radiological best practices, clinical use of such generative models requires rigorous safeguards against confabulation, limited reproducibility, stochastic drift, and data privacy risks [16]. Addressing daily diagnostic, prosthetic, and surgical challenges, further contributions to this Special Issue expanded automated analysis across modalities—demonstrating deep learning pipelines for dental implant system identification, automated segmentation of delicate anatomical structures such as the infraorbital canal, AI-assisted volumetric planning for maxillary sinus floor augmentation, and lightweight explainable frameworks for pediatric infraocclusion detection.

Future research should therefore prioritize the following directions:

- Prospective, multicenter, multi-device external validation, with local acceptance testing on the receiving institution's own images before clinical reliance [9,11,17].
- Clinically anchored reference standards—clinical attachment levels, surgical confirmation, or histopathology—to replace annotation-only ground truth [6,11].

- Human-in-the-loop reader studies that measure the performance of clinician-plus-AI teams, explicitly addressing automation bias, workflow effects, and patient outcomes [7,8,18].
- Open, multi-device benchmark datasets with device-stratified reporting, enabling fair comparison beyond single-dataset leaderboards [9,19,20].
- Responsible exploration of foundation and vision–language models, with explicit calibration, hallucination control, and privacy-preserving training strategies such as federated learning [16].
- Full adoption of AI-specific reporting and risk-of-bias standards (CLAIM, TRIPOD-AI, CONSORT-AI, and PROBAST-AI) and early engagement with regulatory approval pathways [21–27].

Like its predecessor, this Special Issue was enriched by contributions from multiple continents and disciplines, spanning periodontology, endodontics, orthodontics, implant dentistry, oral and maxillofacial surgery, and pediatric dentistry. We sincerely thank all authors, reviewers, and the editorial team of *Diagnostics* for their commitment. Taken together, these thirteen papers show that the question is no longer whether AI can interpret dentomaxillofacial images, but how to ensure that it does so reliably, transparently, and equitably in everyday clinical practice. We hope this collection accelerates the transition from proof of concept to validated, trustworthy clinical integration.

**Author Contributions:** Conceptualization, K.O., A.L.F.C. and S.L.P.d.C.L.; writing—original draft preparation, K.O.; writing—review and editing, A.L.F.C. and S.L.P.d.C.L.; supervision, K.O.; project administration, K.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Schwendicke, F.; Golla, T.; Dreher, M.; Krois, J. Convolutional neural networks for dental image diagnostics: A scoping review. *J. Dent*. **2019**, *91*, 103226. [CrossRef] [PubMed]
2. Hung, K.; Montalvao, C.; Tanaka, R.; Kawai, T.; Bornstein, M.M. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. *Dentomaxillofac. Radiol*. **2020**, *49*, 20190107. [CrossRef] [PubMed]
3. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [CrossRef] [PubMed]
4. Schwendicke, F.; Samek, W.; Krois, J. Artificial Intelligence in Dentistry: Chances and Challenges. *J. Dent. Res*. **2020**, *99*, 769–774. [CrossRef] [PubMed]
5. Szabó, V.; Orhan, K.; Dobó-Nagy, C.; Veres, D.S.; Manulis, D.; Ezhov, M.; Sanders, A.; Szabó, B.T. Deep Learning-Based Periapical Lesion Detection on Panoramic Radiographs. *Diagnostics* **2025**, *15*, 510. [CrossRef] [PubMed]
6. Farooqi, O.A.; Fru, G.A.; Gong, Y.M.; Oswald, L.E.; DeNucci, D.J. Evaluation of artificial intelligence-based clinical decision support systems for caries and periodontal bone loss: An external validation study. *J. Am. Dent. Assoc*. **2026**, *157*, 611–618. [CrossRef] [PubMed]
7. AlGhaihab, A.; Moretti, A.J.; Reside, J.; Tuzova, L.; Tyndall, D.A. An Assessment of Deep Learning's Impact on General Dentists' Ability to Detect Alveolar Bone Loss in 2D Intraoral Radiographs. *Diagnostics* **2025**, *15*, 467. [CrossRef] [PubMed]
8. Pul, U.; Tichy, A.; Pitchika, V.; Schwendicke, F. Impact of artificial intelligence assistance on diagnosing periapical radiolucencies: A randomized controlled trial. *J. Dent*. **2025**, *160*, 105868. [CrossRef] [PubMed]
9. Özcan, M.; Allaf, F. Cross-Domain Generalization of Deep Learning Architectures for Cephalometric Landmark Detection: A Dual-Dataset and Multi-Device Benchmark. *Diagnostics* **2026**, *16*, 2726. [CrossRef] [PubMed]
10. Julião, E.L.D.; Adiverci, G.C.; Fagundes, F.B.; Leite, A.F.; Neves, F.S.; Queiroz de Freitas, D.; Fontenele, R.C.; Jacobs, R.; Haiter-Neto, F.; de-Azevedo-Vaz, S.L. Analysis of the generalizability of an artificial intelligence-based software for tomographic segmentation of posterior teeth—An external validation study. *Dentomaxillofac. Radiol*. **2026**, *55*, 508–519. [CrossRef] [PubMed]
11. Hashim, N.T.; Gismalla, B.G.; Rahman, M.M.; Mohammed, R.; Padmanabhan, V.; Islam, M.S.; Babiker, R.; Elsheikh, M.; Ahmed, A.; Kukreja, B.J. Deep Learning for the Assessment of Alveolar Bone Loss on Intraoral Radiographs: A Systematic Review. *Diagnostics* **2026**, *16*, 2575. [CrossRef] [PubMed]

12. Çetin, B.; İçöz, D.; Dinç, K.; Kayadibi, İ. Radiomics-Based Differential Diagnosis of Radicular Cysts and Apical Granulomas on CBCT Images Using RadC-CNN Architecture. *Diagnostics* **2026**, *16*, 1428. [CrossRef] [PubMed]
13. Ver Berne, J.; Ton That, M.; Jacobs, R. Performance and clinical applicability of AI models for jawbone lesion classification: A systematic review with meta-analysis and introduction of a clinical interpretation score. *Dentomaxillofac. Radiol.* **2026**, *55*, 130–143. [CrossRef] [PubMed]
14. Martín-Noguerol, T.; López-Úbeda, P.; Luna, A. Agentic AI in Radiology: Fundamentals, Drivers, and Gaps. *Acad. Radiol.* **2026**. [CrossRef] [PubMed]
15. Titu, M.F.S.; Pavel, M.A.; Apurba, A.Z.; Ahmed, S.; Rahman, S.; Dudley, J.; Farook, T.H.; Troiano, G. Error-Tolerant Multimodal Vision-Language Models for Endodontic Triaging: A Cross-Sectional Study. *Int. J. Dent.* **2026**, *2026*, 4148741. [CrossRef] [PubMed]
16. Yi, P.H.; Haver, H.L.; Jeudy, J.J.; Kim, W.; Kitamura, F.C.; Oluyemi, E.T.; Smith, A.D.; Moy, L.; Parekh, V.S. Best Practices for the Safe Use of Large Language Models and Other Generative AI in Radiology. *Radiology* **2025**, *316*, e241516. [CrossRef] [PubMed]
17. Park, S.H.; Han, K.; Jang, H.Y.; Park, J.E.; Lee, J.G.; Kim, D.W.; Choi, J. Methods for Clinical Evaluation of Artificial Intelligence Algorithms for Medical Diagnosis. *Radiology* **2023**, *306*, 20–31. [CrossRef] [PubMed]
18. Goddard, K.; Roudsari, A.; Wyatt, J.C. Automation bias: A systematic review of frequency, effect mediators, and mitigators. *J. Am. Med. Inform. Assoc.* **2012**, *19*, 121–127. [CrossRef] [PubMed]
19. Brinz, J.; Eslamiamirabadi, N.; Salamati, A.; Tresp, V.; Schwendicke, F.; Tichy, A. Data sharing for responsible artificial intelligence in dentistry: A narrative review of legal frameworks and privacy-preserving techniques. *J. Dent.* **2025**, *163*, 106130. [CrossRef] [PubMed]
20. Lee, T.; Lee, J.H.; Yoon, S.H.; Park, S.H.; Kim, H. Availability and transparency of artificial intelligence models in radiology: A meta-research study. *Eur. Radiol.* **2025**, *35*, 5287–5298. [CrossRef] [PubMed]
21. Sultan, Z.; Osathanon, T.; Abdullahi Shire, M.; Schwendicke, F.; Samaranayake, L. Artificial Intelligence in Dentistry: A Concise Review of Reporting Checklists and Guidelines. *Int. Dent. J.* **2025**, *76*, 109322. [CrossRef] [PubMed]
22. Zhong, J.; Xing, Y.; Lu, J.; Zhang, G.; Mao, S.; Chen, H.; Yin, Q.; Cen, Q.; Jiang, R.; Hu, Y.; et al. The endorsement of general and artificial intelligence reporting guidelines in radiological journals: A meta-research study. *BMC Med. Res. Methodol.* **2023**, *23*, 292. [CrossRef] [PubMed]
23. Mongan, J.; Moy, L.; Kahn, C.E. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiol. Artif. Intell.* **2020**, *2*, e200029. [CrossRef] [PubMed]
24. Sounderajah, V.; Guni, A.; Liu, X.; Collins, G.S.; Karthikesalingam, A.; Markar, S.R.; Golub, R.M.; Denniston, A.K.; Shetty, S.; Moher, D.; et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat. Med.* **2025**, *31*, 3283–3289. [CrossRef] [PubMed]
25. Liu, X.; Cruz Rivera, S.; Moher, D.; Calvert, M.J.; Denniston, A.K.; Chan, A.-W.; Darzi, A.; Holmes, C.; Yau, C.; Ashrafian, H.; et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *Nat. Med.* **2020**, *26*, 1364–1374. [CrossRef] [PubMed]
26. Moons, K.G.M.; Damen, J.A.A.; Kaul, T.; Hooft, L.; Andaur Navarro, C.; Dhiman, P.; Beam, A.L.; Van Calster, B.; Celi, L.A.; Denaxas, S.; et al. PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ* **2025**, *388*, e082505. [CrossRef] [PubMed]
27. Uribe, S.E.; Hamdan, M.H.; Valente, N.A.; Yamaguchi, S.; Umer, F.; Tichy, A.; Pauwels, R.; Schwendicke, F. Evaluating dental AI research papers: Key considerations for editors and reviewers. *J. Dent.* **2025**, *160*, 105867. [CrossRef] [PubMed]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.