Vol. 18, No.3 July-September 2026

# BEYOND AUTOMATED DETECTION: ARTIFICIAL INTELLIGENCE IN MULTIMODAL CRANIOFACIAL IMAGING, MORPHOLOGICAL PHENOTYPING, AND CLINICAL DECISION SUPPORT—A NARRATIVE REVIEW

**Argatu Daniela, Balcos Carina, Norina Forna, Bamboi Irina, Corneaga Andrei, Doriana Agop-Forna**

"Grigore T. Popa" University of Medicine and Pharmacy Iasi, Romania
Correspondinf author: profforna@gmail.com

## Abstract

**Background:** Artificial intelligence (AI) has expanded the analytical capabilities of craniofacial radiology, enabling automated detection, anatomical segmentation, quantitative image analysis, and increasingly integrated computational modeling. Its clinical significance, however, remains heterogeneous and is rarely demonstrated beyond technical performance. **Objective:** To critically synthesize AI applications in panoramic radiography, cephalometric imaging, cone-beam computed tomography (CBCT), conventional computed tomography (CT), and magnetic resonance imaging (MRI), and to organize them within a functional framework that distinguishes detection, quantification, morphological phenotyping, and clinical decision support. **Methods:** Structured narrative review. PubMed/MEDLINE, Scopus, Web of Science, and IEEE Xplore were searched for studies published between January 2018 and June 2026, supplemented by foundational earlier work and methodological guidance. Studies were prioritized according to reference-standard quality, independence of test data, multicenter or external validation, and the presence of clinical or longitudinal endpoints. The proposed Craniofacial AI Diagnostic Continuum was positioned against the hierarchical model of diagnostic imaging efficacy. **Evidence synthesis:** Representative studies show mature performance for cephalometric landmark localization, fracture and lesion detection, and tooth, bone, and mandibular-canal segmentation, including multicenter datasets of several thousand examinations. In contrast, morphological phenotyping and multimodal decision support rest on few cohorts, mostly single center, with prospective longitudinal validation reported only exceptionally (e.g., temporomandibular joint osteoarthritis progression). External validation, calibration, and evaluation of clinical impact remain uncommon. **Conclusions:** The continuum adds a phenotyping level that has no direct counterpart in classical efficacy hierarchies and makes explicit which evidence each level requires. Increasing algorithmic complexity does not confer clinical value; future research should prioritize reproducible morphometry, external and prospective validation, and outcome-based evaluation.

**Keywords:** AI; craniofacial imaging; panoramic radiography; cephalometry; CBCT; MRI; morphological phenotyping.

## 1. Introduction

Craniofacial radiology encompasses imaging techniques used to investigate anatomical, developmental, traumatic, degenerative, and pathological conditions of the craniofacial complex. Panoramic and cephalometric radiography provide accessible two-dimensional representations, whereas CBCT and CT enable three-dimensional assessment of osseous anatomy. MRI complements these modalities by visualizing soft tissues, articular structures, and bone marrow.

Interpretation remains challenging because of anatomical complexity, morphological variability, superimposition, artifacts, and acquisition differences. Many clinical procedures still require manual landmark identification, segmentation, and linear, angular, or volumetric measurements, which are time-consuming and observer-dependent.


DOI : 10.62610/RJOR.2026.3.18.5

Artificial intelligence, particularly machine learning and deep learning, has introduced automated identification, localization, segmentation, and quantitative image analysis. Early work demonstrated convolutional neural network (CNN) methods for cephalometric landmark localization [1] and radiographic detection of periodontal bone loss [2], and more recent CBCT studies illustrate substantial advances in anatomical segmentation and efficiency [3–5]. A systematic review of dental and maxillofacial radiology found that most applications concentrated on cephalometric landmarks, osteoporosis, cysts and tumors, and periodontal or periapical disease, while emphasizing that reliability and applicability required further verification before clinical use [6].

Technical performance does not by itself establish clinical utility. Across medical imaging, only 6% of 516 AI diagnostic accuracy studies published in 2018 performed external validation, and none combined a diagnostic cohort design, multiple institutions, and prospective data collection [7]. Quantitative outputs may be accurate without defining clinically meaningful phenotypes or improving decisions. This review therefore (i) proposes a functional distinction between detection, quantification, phenotyping, and decision support; (ii) relates this distinction to the established hierarchy of diagnostic imaging efficacy [8]; and (iii) appraises representative evidence across modalities, using the temporomandibular joint (TMJ) as an illustrative case.

## 2. Material and Methods

This manuscript is a structured narrative review. A narrative design was chosen because imaging modalities, target conditions, algorithms, reference standards, and performance measures are too heterogeneous for quantitative pooling. The manuscript was prepared with reference to the six items of the Scale for the Assessment of Narrative Review Articles (SANRA): justification of importance, explicit aims, description of the literature search, referencing, scientific reasoning, and presentation of relevant endpoint data [9].

PubMed/MEDLINE, Scopus, Web of Science, and IEEE Xplore were searched for records published between January 2018 and June 2026. The search combined three concept groups with Boolean operators: (1) artificial intelligence, machine learning, deep learning, convolutional neural network, transformer, radiomics; (2) craniofacial, maxillofacial, dentomaxillofacial, panoramic, cephalometric, cone-beam computed tomography, computed tomography, magnetic resonance imaging, temporomandibular joint; and (3) segmentation, landmark, detection, classification, morphology, shape analysis, phenotype, prediction, decision support. Reference lists of included systematic reviews were screened, and foundational studies published before 2018 were added when they introduced methods or benchmarks still in use (table 1).


DOI : 10.62610/RJOR.2026.3.18.5

**Table 1.The web search results.**

| Aim | Records identified by review authors | Studies included |
|---|---|---|
| Generative AI in dental and maxillofacial radiology | 2,060 | 43 |
| AI in TMJ osteoarthritis on CBCT | 2,543 | 9 |
| AI in the diagnosis of TMJ disorders | 1,497 | 5 reviews |
| AI in cephalometry | Not reported here | 20 |

Original studies were prioritized when they reported (a) a clearly defined reference standard; (b) a test set separated from training data at the patient level; (c) multicenter, multi-device, or external validation; and (d) clinical, longitudinal, or reader-study endpoints. Methodological guidance on reporting and validation was included where directly relevant: CLAIM 2024 [10], the dental AI checklist [11], TRIPOD+AI [12], and DECIDE-AI [13]. Evidence was appraised in relation to dataset characteristics, imaging modality, reference standard, validation approach, metrics, and clinical relevance. The proposed continuum is an interpretive framework, not a validated classification or risk-of-bias instrument.

## 3. Conceptual Framework: The Craniofacial AI Diagnostic Continuum

### 3.1. Four functional levels

The continuum describes four analytical functions: (1) **detection**—localization or classification of structures and abnormalities; (2) **quantification**—segmentation, dimensions, volumes, surface shape, and radiomic features; (3) **phenotyping**—reproducible multidimensional morphological patterns that summarize anatomical variation beyond isolated measurements; and (4) **clinical decision support**—integration of imaging information with clinical data for a defined diagnostic, prognostic, or therapeutic decision (Figure 1, Table 2). The levels can overlap and do not imply that every system must progress sequentially; a well-validated detection model may be more useful than a poorly validated phenotyping model.

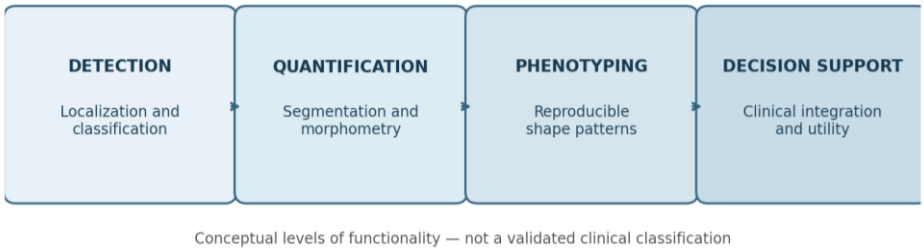



**Figure 1. Proposed Craniofacial AI Diagnostic Continuum**


DOI : 10.62610/RJOR.2026.3.18.5

**Table 2. Functional levels of the Craniofacial AI Diagnostic Continuum and their closest correspondence in the hierarchical model of diagnostic imaging efficacy [8].**

| Level | Question | AI output | Required validation | Closest Fryback–Thornbury level |
|---|---|---|---|---|
| Detection | Is a structure or abnormality present, and where? | Localization, classification | Sensitivity, specificity, calibration, external testing | Level 2 (diagnostic accuracy) |
| Quantification | What are its anatomical properties? | Volumes, dimensions, shape, radiomic features | Dice with boundary metrics, ICC, measurement error | Level 1–2 (technical and measurement accuracy) |
| Phenotyping | Are there stable, distinct morphological patterns? | Clusters, shape modes, phenotype scores | Resampling stability, replication in independent cohorts, construct validity | No direct equivalent; precedes Levels 3–4 |
| Decision support | How should diagnosis or care change? | Risk estimates, recommendations | Reader studies, decision-curve analysis, prospective impact | Levels 3–5 (diagnostic thinking, management, patient outcome) |

### 3.2. Relation to the hierarchical model of diagnostic efficacy

Fryback and Thornbury described six levels of diagnostic imaging efficacy, from technical image quality and diagnostic accuracy to effects on diagnostic thinking, patient management, patient outcomes, and societal costs and benefits, and argued that efficacy at a lower level is necessary but not sufficient for efficacy at a higher one [8]. The present continuum is compatible with this logic but answers a different question. The efficacy hierarchy classifies the type of evidence available for an imaging test, whereas the continuum classifies the type of information an AI system produces. The two are therefore complementary: Table 1 indicates, for each functional level, which efficacy levels must be demonstrated before clinical claims are justified.

The main addition of the continuum is the phenotyping level. Classical efficacy models assume a predefined target condition with a reference standard. Morphological phenotyping instead derives candidate categories from the data themselves, so its first validation task is not diagnostic accuracy but construct validity: whether the patterns are stable, reproducible across samples and scanners, and associated with independently measured clinical variables. Making this step explicit prevents computationally derived groupings from being presented as diagnostic entities before such evidence exists.

## 4. Artificial Intelligence Across Craniofacial Imaging Modalities

Modality-specific evaluation is necessary because an AI prediction is constrained by the physical information present in its source image. Panoramic and cephalometric radiographs provide projected anatomy, CBCT emphasizes high-resolution osseous geometry, conventional CT offers standardized attenuation values and broader anatomical coverage, and MRI provides complementary soft-tissue contrast. A multimodal model must therefore demonstrate added diagnostic value over its best single-modality comparator, rather than


DOI : 10.62610/RJOR.2026.3.18.5

merely aggregate inputs (Figure 2). Representative studies are summarized in Table 2.

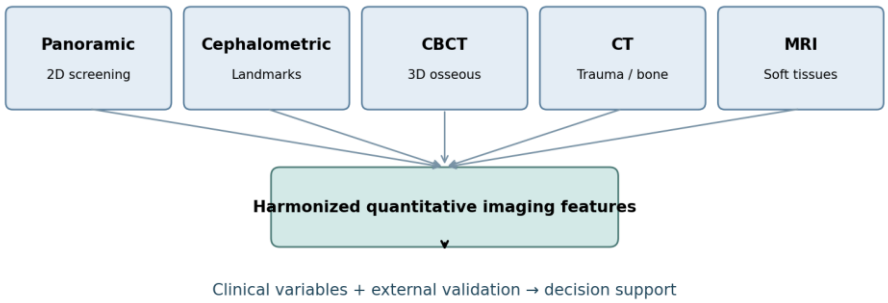



**Figure 2. Conceptual multimodal integration pathway (schematic; not patient imaging data).**

### 4.1. Panoramic radiography

Panoramic radiography is widely available and suited to automated screening. Krois et al. reported a mean classification accuracy of approximately 0.81 for periodontal bone-loss detection on panoramic image segments [2]. A modified YOLOv3 detector was trained on 1,282 panoramic radiographs covering dentigerous cysts, periapical cysts, odontogenic keratocysts, ameloblastomas, and normal jaws [14], and a Faster R-CNN model with a Swin-Transformer backbone, developed on 6,404 radiographs from 5,621 patients, achieved an F1 score of 0.947 and an area under the curve (AUC) of 0.977 for mandibular fracture detection [15]. These studies were, however, single-center and retrospective. Positioning, magnification, distortion, and superimposition further limit quantitative interpretation, so AI-assisted condylar measurements on panoramic images should not be treated as equivalent to three-dimensional morphology.

### 4.2. Cephalometric imaging

Cephalometry is the best-benchmarked craniofacial AI task. The International Symposium on Biomedical Imaging (ISBI) 2015 challenge established a public dataset and evaluation framework for lateral cephalometric landmark detection [16]; a random-forest regression-voting approach that won the challenge was developed into a fully automatic system reported to perform within the inter-observer variability of two clinical experts [17], and an early CNN approach reported a 75.3% successful detection rate within 2 mm [1]. Automated localization can reduce repetitive work and enable standardized large-scale analysis. Localization error should nevertheless be interpreted in terms of derived angular and linear measurements and their clinical thresholds, and repeated testing on a single public benchmark carries a risk of overfitting to its acquisition characteristics.

### 4.3. Cone-beam computed tomography

CBCT supports three-dimensional segmentation, morphometry, volumetry, and


DOI : 10.62610/RJOR.2026.3.18.5

surface analysis. Verhelst et al. developed a layered deep-learning mandibular segmentation method [3]. Preda et al. reported a mean Dice similarity coefficient of 92.6% and automated maxillofacial segmentation in approximately 39 seconds versus more than two hours manually [4]. Cui et al. analyzed 4,938 CBCT examinations from 15 centers, reporting average Dice scores of 91.5% for teeth and 93.0% for alveolar bone [5]. For the mandibular canal, a fully convolutional network trained on 637 coarsely annotated volumes reached a mean curve distance of 0.56 mm [18], and a subsequent study tested the system on 150 scans from five scanners in European and Southeast Asian institutions, including devices absent from training [19]. Mandibular fracture detection on CBCT has also been benchmarked with publicly released code [20].

These results indicate strong segmentation feasibility and, increasingly, cross-device generalizability, but do not establish diagnostic or therapeutic benefit. CBCT grey values are not standardized and vary considerably between devices and protocols [21]; together with voxel size, field of view, and artifacts, this variability limits the reproducibility of texture and radiomic features unless intensities are harmonized and feature definitions follow standardized references [22].

### 4.4. Conventional computed tomography

CT is essential for selected craniofacial trauma, complex skeletal pathology, tumors, and reconstruction. A two-stage U-Net and ResNet pipeline trained on CT scans of 686 patients with mandibular fractures reported high accuracy and reliability for fracture detection and classification against a reference standard set by three experienced maxillofacial surgeons [23], and a similar segmentation-then-classification approach has been applied to zygomatic fractures [24]. Clinical validation should consider subtle and multiple fracture patterns, postoperative changes, and unexpected abnormalities. An algorithm validated for a narrow task should not be assumed to interpret the entire examination.

### 4.5. Magnetic resonance imaging

MRI provides complementary soft-tissue information, particularly on TMJ disc position and morphology, effusions, marrow abnormalities, and selected lesions. A CNN for anterior disc displacement was developed on sagittal MRI of 2,520 TMJs from 1,260 patients [25]; a proof-of-concept study segmented the articular disc on 217 images from 20 participants [26]; and a multistage system combining region-of-interest detection, segmentation of the disc, condyle, fossa, and articular eminence, and classification was trained on 618 TMJ cases from two hospitals with internal and external validation [27]. Imaging abnormalities do not independently establish pain, functional impairment, or treatment need, and integration of MRI with CBCT or CT requires registration, harmonized reference standards, and clinical testing.

### 4.6. Cross-modality evidence synthesis

Table 3 shows a consistent pattern. Most studies operate at the detection and quantification levels, report overlap or accuracy metrics on retrospective data and validate internally. Multicenter or cross-


DOI : 10.62610/RJOR.2026.3.18.5

device testing is now reported for segmentation tasks [5,19], but rarely for detection. Phenotyping and decision-support studies are few, concentrated in TMJ osteoarthritis, and based on cohorts of approximately 100 patients; only one of the studies identified here reported prospective longitudinal endpoints [28]. None reported a reader study measuring the effect of AI assistance on clinicians' decisions.

## 5. From Automated Detection to Quantitative Morphological Characterization

Detection identifies whether and where a structure or abnormality is present. Quantification characterizes dimensions, volumes, surface shape, spatial relationships, and texture. Segmentation provides anatomical boundaries for many downstream measurements, but boundary uncertainty propagates into volume and shape estimates; studies should therefore report measurement reproducibility in addition to segmentation overlap. Computational shape analysis may reveal anatomical variation missed by isolated distances and angles. Radiomic features require intensity harmonization and standardized definitions[22], and should not be called clinical biomarkers without independent validation.


DOI : 10.62610/RJOR.2026.3.18.5

**Table 3. Representative evidence for AI in craniofacial imaging, ordered by level of the continuum.**

| Study | Modality and task | Data and validation | Key reported result | Level |
|---|---|---|---|---|
| Wang et al. 2016 [16] | Cephalometry; landmark detection benchmark | ISBI 2015 public dataset; 7 methods compared | Common benchmark and evaluation framework | Detection |
| Lindner et al. 2016 [17] | Cephalometry; 19 landmarks | Fully automatic system; earlier version won ISBI 2015 | Within inter-observer variability of two experts | Detection |
| Kwon et al. 2020 [14] | Panoramic; cysts and tumors | 1,282 radiographs, 4 lesion classes + normal; internal test | High sensitivity, specificity, and AUC (internal) | Detection |
| Vinayahalingam et al. 2022 [15] | Panoramic; mandibular fractures | 6,404 radiographs, 5,621 patients; single center | F1 0.947; AUC 0.977 | Detection |
| Wang X et al. 2022 [23] | CT; mandibular fractures | 686 patients; train/validation/test 222/56/408 scans | High accuracy vs surgeon-annotated reference | Detection |
| van Nistelrooij et al. 2024 [20] | CBCT; mandibular fractures | 171 scans with and 164 without fractures; public code | Precision 0.978; sensitivity 0.956 | Detection |
| Lee YH et al. 2022 [25] | MRI; TMJ anterior disc displacement | 2,520 TMJs, 1,260 patients | CNN-based automated detection | Detection |
| Yu et al. 2024 [27] | MRI; TMJ disc displacement | 618 cases, 2 hospitals; internal and external validation | Precision >92% for classification | Detection / quantification |
| Ito et al. 2022 [26] | MRI; TMJ disc segmentation | 217 images, 20 participants | Proof of concept | Quantification |
| Preda et al. 2022 [4] | CBCT; maxillofacial segmentation | 144 scans | Dice 92.6%; 39.1 s vs 132.7 min manual | Quantification |
| Cui et al. 2022 [5] | CBCT; tooth and alveolar bone | 4,938 scans, 4,215 patients, 15 centers | Dice 91.5% (teeth), 93.0% (bone) | Quantification |
| Jaskari et al. 2020 [18] | CBCT; mandibular canal | 637 training volumes; 15 voxel-level test volumes | Mean curve distance 0.56 mm | Quantification |
| Järnstedt et al. 2022 [19] | CBCT; mandibular canal | 982 training scans; 150 test scans, 5 scanners, 2 regions | Symmetric mean curve distance 0.39 mm vs consensus | Quantification |
| Schneider et al. 2023 [29] | Panoramic; tooth segmentation | 4,177 radiographs, 9 centers; federated learning | Federated outperformed local training in 8 of 9 centers | Quantification |
| de Dumast et al. 2018 [30] | CBCT; 3D condylar shape | 259 training condyles; 34 test condyles | 91% agreement with clinician consensus on 5 degenerative stages | Phenotyping |
| Bianchi et al. 2020 [31] | CBCT radiomics + clinical + biological markers; TMJ OA | 52 markers; cross-sectional, single cohort | AUC 0.870 for TMJ OA status | Decision support (diagnostic) |
| Al Turkestani et al. 2024 [28] | CBCT radiomics, shape + clinical + biological; TMJ OA progression | Prospective; 106 participants, 74 followed 2–3 years | AUC 0.72; F1 0.82 for progression | Decision support (prognostic) |

AUC, area under the receiver operating characteristic curve; CBCT, cone-beam computed tomography; CT, computed tomography; ISBI, International Symposium on Biomedical Imaging; MRI, magnetic resonance imaging; OA, osteoarthritis; TMJ, temporomandibular joint. Results are those reported by the original authors; metrics are not directly comparable across tasks.


DOI : 10.62610/RJOR.2026.3.18.5

## 6. AI-Driven Craniofacial Morphological Phenotyping

Morphological phenotyping seeks reproducible patterns of anatomical variation rather than isolated measurements. AI can integrate many shape descriptors using clustering, representation learning, or statistical shape models. Mandibular and condylar phenotyping might incorporate condylar dimensions, curvature, volume, orientation, and bilateral asymmetry. Existing work illustrates both potential and limits: a neural-network classifier of three-dimensional condylar morphology staged degenerative change in close agreement with clinician consensus, but was trained against clinician-defined categories on a single cohort [30].

A defensible phenotyping pipeline should begin with segmentation quality assurance and repeated anatomical measurements, followed by harmonization of acquisition-related variation. Unsupervised clusters should be assessed using resampling stability, sensitivity analyses across analytical choices, and an independent validation cohort. Associations with age, skeletal pattern, symptoms, and independently defined outcomes should be prespecified where possible; otherwise phenotype labels risk becoming descriptive groupings without established biological meaning. Computational clusters are not automatically disease entities.

## 7. From Imaging Biomarkers to Clinical Decision Support

Clinical decision support uses imaging-derived information alongside clinical context to inform diagnosis, risk stratification, monitoring, or treatment planning. Predictive claims require longitudinal endpoints; cross-sectional classification cannot establish future progression. Multimodal models require robust registration and explicit handling of missing modalities, and should demonstrate incremental value over clinical variables alone.

Utility should be tested in reader-assistance or prospective workflow studies comparing clinicians with and without AI, with outcomes that include missed findings, calibration of risk estimates, changes in management, interpretation time, and potential downstream harms. Decision-curve analysis can quantify net benefit across clinically meaningful threshold probabilities [32]. Early clinical evaluations should be reported according to DECIDE-AI [13], and prediction models according to TRIPOD+AI [12].

## 8. Case Study: The Temporomandibular Joint Across the Continuum

The TMJ illustrates all four levels and the evidence each requires. At the detection level, CNNs identify anterior disc displacement on MRI [25,27]. At the quantification level, the articular disc can be segmented on MRI [26], while CBCT supports condylar surface models and trabecular texture features. At the phenotyping level, three-dimensional condylar shape has been used to stage degenerative change [30]. At the decision-support level, models combining CBCT radiomics with clinical and salivary or serum biomarkers discriminated TMJ osteoarthritis


DOI : 10.62610/RJOR.2026.3.18.5

status in cross-sectional cohorts [31,33], and a prospective study of 106 participants, 74 of whom were followed for two to three years, predicted disease progression with an AUC of 0.72 [28].

This sequence also exposes the gaps. Osseous CBCT features and MRI soft-tissue findings have largely been analyzed separately; structural change does not equate to pain or dysfunction; and the only longitudinal model reported moderate discrimination in a single prospective cohort, without external validation or a reader study. The TMJ is therefore a suitable test bed for genuinely multimodal, longitudinally validated decision support, provided that imaging findings are linked to standardized clinical diagnostic criteria.

## 9. Methodological Challenges and Clinical Translation

### 9.1. Reproducibility, metrics, and uncertainty

Metrics should match the clinical question [34]. For segmentation, Dice overlap should be reported together with boundary-sensitive metrics such as the 95th-percentile Hausdorff distance and downstream measurement error. For classification, discrimination, calibration, confidence intervals, and prevalence-dependent predictive values are needed. Dice overlap, landmark error, sensitivity, specificity, and AUC describe different tasks and cannot be compared directly. Patient-level data partitions, subgroup analyses, transparent missing-data handling, and explicit uncertainty communication guard against optimistic performance estimates. Explainability displays do not replace external validation. CLAIM 2024 and the dental AI checklist support transparent reporting of data sources, reference standards, and evaluation [10,11].

### 9.2. Bias, generalizability, and data governance

Models trained on narrow populations or a single device may fail silently elsewhere. In chest radiography, AI classifiers trained on large public datasets systematically underdiagnosed historically under-served patient groups [35]; comparable subgroup analyses by age, sex, ethnicity, and device are rarely reported in craniofacial AI. Data protection limits central pooling of dental images. Federated learning, in which models are trained across institutions without sharing images, outperformed isolated local training for panoramic tooth segmentation across nine centers, although centrally pooled training remained superior [29].

### 9.3. Regulatory considerations

In the European Union, software intended to provide information used for diagnostic or therapeutic decisions is regulated as a medical device under Regulation (EU) 2017/745 [36]. AI systems that are, or are safety components of, such devices and require third-party conformity assessment are additionally classified as high-risk under the Artificial Intelligence Act [37], which introduces obligations on data governance, documentation, human oversight, and post-market monitoring. These requirements align with the validation steps described here: a system intended for decision support will need evidence corresponding to the higher levels of Table 1, not only technical accuracy.


DOI : 10.62610/RJOR.2026.3.18.5

## 10. Future Research Priorities

Future research should emphasize reproducible three-dimensional morphometry, externally validated anatomical phenotypes, multimodal image integration, longitudinal analysis, and patient-centered outcomes. Public benchmarks with multicenter, multi-device test sets, analogous to the cephalometric challenges [16], are needed for CBCT and MRI tasks. Foundation models and generative AI may expand image reconstruction, enhancement, synthetic data generation, and representation learning, but anatomical fidelity and synthetic artifacts remain concerns [38]. Technical novelty should be balanced against diagnostic safety and patient benefit.

## 11. Discussion

The evolution of craniofacial AI is characterized by increasingly sophisticated extraction of anatomical information. Landmark identification and segmentation are technically mature, with multicenter evidence for some tasks, whereas phenotyping and multimodal decision support remain at an early stage. The proposed continuum separates computational output from clinical meaning and, by mapping each level to the efficacy hierarchy [8], makes explicit that quantitative features do not become clinically useful biomarkers without reproducibility and outcome-based validation. Imaging modalities are complementary rather than interchangeable, and the TMJ illustrates the need to combine osseous, soft-tissue, and clinical information without equating structural change with symptoms.

This review has limitations. As a narrative review, it does not provide a formal risk-of-bias assessment or quantitative synthesis, and study selection may reflect the authors' judgment. The studies in Table 2 are representative rather than exhaustive. The continuum is a conceptual proposal whose usefulness should itself be tested, for example by assessing inter-rater agreement when classifying published studies by level.

## 12. Conclusions

AI extends craniofacial radiology through detection, segmentation, quantitative analysis, and emerging integrated modeling. The Craniofacial AI Diagnostic Continuum distinguishes detection, quantification, morphological phenotyping, and clinical decision support, and links each to the evidence required for clinical claims. Evidence of technical feasibility is strongest for selected detection and segmentation tasks, whereas advanced phenotyping and multimodal decision support still require external, prospective, and outcome-based validation. The value of AI depends on reliable, interpretable, and actionable information rather than computational complexity alone.

## References

1. Arık SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. J Med Imaging. 2017;4(1):014501. doi:10.1117/1.JMI.4.1.014501.
2. Krois J, Ekert T, Meinhold L, et al. Deep learning for the radiographic detection of periodontal bone loss. Sci Rep. 2019; 9:8495. doi:10.1038/s41598-019-44839-3.


DOI : 10.62610/RJOR.2026.3.18.5

3. Verhelst PJ, Smolders A, Beznik T, et al. Layered deep learning for automatic mandibular segmentation in cone-beam computed tomography. J Dent. 2021;114:103786. doi: 10.1016/j.jdent.2021.103786.
4. Preda F, Morgan N, Van Gerven A, et al. Deep convolutional neural network-based automated segmentation of the maxillofacial complex from cone-beam computed tomography: a validation study. J Dent. 2022; 124:104238. doi: 10.1016/j.jdent.2022.104238.
5. Cui Z, Fang Y, Mei L, et al. A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. Nat Commun. 2022; 13:2096. doi:10.1038/s41467-022-29637-2.
6. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: a systematic review. Dentomaxillofac Radiol. 2020;49(1):20190107. doi:10.1259/dmfr.20190107.
7. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. Korean J Radiol. 2019;20(3):405–410.
8. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making. 1991;11(2):88–94. doi:10.1177/0272989X9101100203.
9. Baethge C, Goldbeck-Wood S, Mertens S. SANRA— a scale for the quality assessment of narrative review articles. Res Integr Peer Rev. 2019; 4:5. doi:10.1186/s41073-019-0064-8.
10. Tejani AS, Klontzas ME, Gatti AA, et al. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 update. Radiol Artif Intell. 2024;6(4):e240300. doi:10.1148/ryai.240300.
11. Schwendicke F, Singh T, Lee JH, et al. Artificial intelligence in dental research: checklist for authors, reviewers, readers. J Dent. 2021; 107:103610. doi: 10.1016/j.jdent.2021.103610.
12. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378.
13. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924–933. doi:10.1038/s41591-022-01772-9.
14. Kwon O, Yong TH, Kang SR, et al. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. Dentomaxillofac Radiol. 2020;49(8):20200185. doi:10.1259/dmfr.20200185.
15. Vinayahalingam S, van Nistelrooij N, van Ginneken B, et al. Detection of mandibular fractures on panoramic radiographs using deep learning. Sci Rep. 2022; 12:19596. doi:10.1038/s41598-022-23445-w.
16. Wang CW, Huang CT, Lee JH, et al. A benchmark for comparison of dental radiography analysis algorithms. Med Image Anal. 2016; 31:63–76. doi: 10.1016/j.media.2016.02.004.
17. Lindner C, Wang CW, Huang CT, Li CH, Chang SW, Cootes TF. Fully automatic system for accurate localisation and analysis of cephalometric landmarks in lateral cephalograms. Sci Rep. 2016; 6:33581. doi:10.1038/srep33581.



DOI : 10.62610/RJOR.2026.3.18.5

18. Jaskari J, Sahlsten J, Järnstedt J, et al. Deep learning method for mandibular canal segmentation in dental cone beam computed tomography volumes. Sci Rep. 2020;10:5842. doi:10.1038/s41598-020-62321-3.
19. Järnstedt J, Sahlsten J, Jaskari J, et al. Comparison of deep learning segmentation and multigrader-annotated mandibular canals of multicenter CBCT scans. Sci Rep. 2022; 12:18598. doi:10.1038/s41598-022-20605-w.
20. van Nistelrooij N, Schitter S, van Lierop P, et al. Detecting mandible fractures in CBCT scans using a 3-stage neural network. J Dent Res. 2024. doi:10.1177/00220345241256618.
21. Pauwels R, Nackaerts O, Bellaiche N, et al. Variability of dental cone beam CT grey values for density estimations. Br J Radiol. 2013;86(1021):20120135. doi:10.1259/bjr.20120135.
22. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020;295(2):328–338. doi:10.1148/radiol.2020191145.
23. Wang X, Xu Z, Tong Y, et al. Detection and classification of mandibular fracture on CT scan using deep convolutional neural network. Clin Oral Investig. 2022; 26:4593–4601. doi:10.1007/s00784-022-04427-8.
24. Tong Y, Jie B, Wang X, et al. Is convolutional neural network accurate for automatic detection of zygomatic fractures on computed tomography? J Oral Maxillofac Surg. 2023. doi: 10.1016/j.joms.2023.04.013.
25. 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. doi:10.1038/s41598-022-15231-5.
26. Ito S, Mine Y, Yoshimi Y, et al. Automated segmentation of articular disc of the temporomandibular joint on magnetic resonance images using deep learning. Sci Rep. 2022;12. doi:10.1038/s41598-021-04354-w.
27. 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. doi:10.3389/fphys.2024.1445258.
28. Al Turkestani N, Li T, Bianchi J, et al. A comprehensive patient-specific prediction model for temporomandibular joint osteoarthritis progression. Proc Natl Acad Sci U S A. 2024;121(8):e2306132121. doi:10.1073/pnas.2306132121.
29. Schneider L, Rischke R, Krois J, et al. Federated vs local vs central deep learning of tooth segmentation on panoramic radiographs. J Dent. 2023;135:104556. doi:10.1016/j.jdent.2023.104556.
30. de Dumast P, Mirabel C, Cevidanes L, et al. A web-based system for neural network based classification in temporomandibular joint osteoarthritis. Comput Med Imaging Graph. 2018; 67:45–54. doi: 10.1016/j.compmedimag.2018.04.009.
31. Bianchi J, de Oliveira Ruellas AC, Gonçalves JR, et al. Osteoarthritis of the temporomandibular joint can be diagnosed earlier using biomarkers and machine learning. Sci Rep. 2020;10(1):8012. PMID:32415284.


DOI : 10.62610/RJOR.2026.3.18.5

32. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565–574. doi:10.1177/0272989X06295361.
33. 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. doi:10.3389/fdmed.2022.1007011.
34. Maier-Hein L, Reinke A, Godau P, et al. Metrics reloaded: recommendations for image analysis validation. Nat Methods. 2024;21(2):195–212. doi:10.1038/s41592-023-02151-z.
35. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med. 2021;27(12):2176–2182. doi:10.1038/s41591-021-01595-0.
36. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices. Official Journal of the European Union. 2017; L117:1–175.
37. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. 2024; L 2024/1689.
38. Hung KF, Wang F, Lu MY, Shi X, Bornstein MM, Ai QYH. Current evidence of generative artificial intelligence specifically developed for dental and maxillofacial radiology: a systematic review. Dentomaxillofac Radiol. 2026;55(6):538–567. doi: 10.1093/dmfr/twag024.


DOI : 10.62610/RJOR.2026.3.18.5