Review | Open Access

Artificial Intelligence Surgery

Shi *et al. Art Int Surg.* 2026;6:725-46 **DOI:**10.20517/ais.2026.94

# YOLO-based computer vision in oral and maxillofacial surgical care: applications and clinical translation

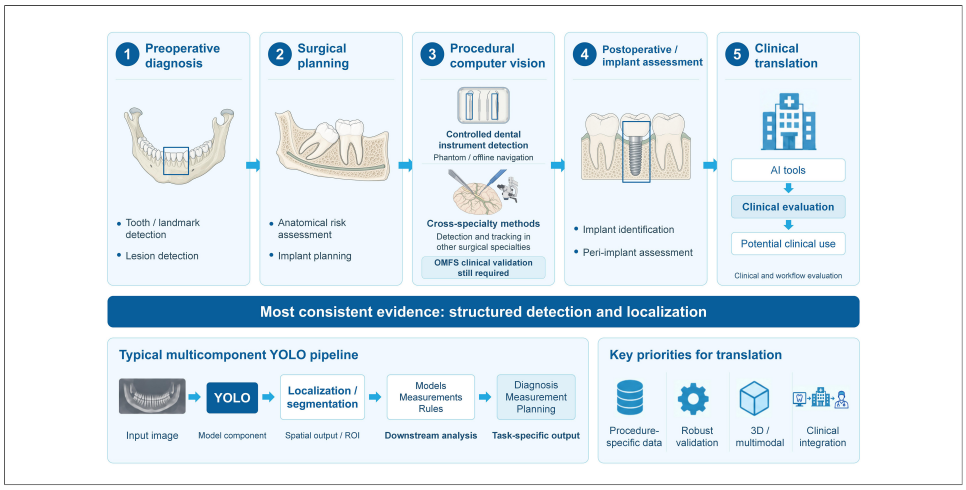
**Jianchen Shi, Jiarui Liu, Jiayu Li, Guiquan Zhu, Zhiyong Guo, Chunjie Li**

**Keywords:**
YOLO, oral and maxillofacial surgery, computer vision, object detection, dental imaging, surgical planning, clinical translation

**Citation:** Shi J, Liu J, Li J, Zhu G, Guo Z, Li C. YOLO-based computer vision in oral and maxillofacial surgical care: applications and clinical translation. *Art Int Surg.* 2026;6:725-46. https://dx.doi.org/10.20517/ais.2026.94

**Received:** 8 Sep 2026
**First Decision:** 10 Oct 2026
**Revised:** 4 Nov 2026
**Accepted:** 5 Nov 2026
**Published:** 19 Nov 2026

**Academic Editor:**
Andrew Gumbs
**Copy Editor:**
Tong Wang
**Production Editor:**
Tong Wang




## Abstract

You Only Look Once (YOLO)-based computer vision has been investigated across oral and maxillofacial imaging and related surgical applications. This narrative review examines its applications in preoperative diagnosis, treatment planning, procedural computer vision, postoperative assessment, and implant-related evaluation. Structured searches of PubMed, Scopus, and the Web of Science Core Collection, supplemented by Google Scholar and citation tracking, informed the purposive selection of 61 representative studies published or available online by 2 September 2026. Evidence is most consistent for structured detection and localization tasks, including tooth and landmark identification, lesion and fracture detection, implant recognition, and selected planning applications. In multicomponent systems, YOLO often provides localization, segmentation, or region-of-interest generation, whereas downstream models, measurements, or rule-based approaches generate diagnostic, planning, or prognostic outputs. Direct evidence for procedural applications in oral and maxillofacial surgery remains limited; studies from other surgical specialties provide methodological insights into instrument tracking, occlusion handling, small-target detection, and orientation-aware localization. Clinical translation will require independent

Check for updates

State Key Laboratory of Oral Diseases & National Center for Stomatology & National Clinical Research Center for Oral Diseases, Department of Head and Neck Oncology, West China Hospital of Stomatology, Sichuan University, Chengdu 610041, Sichuan, China.

**Correspondence to:** Prof. Chunjie Li, Dr. Zhiyong Guo, State Key Laboratory of Oral Diseases & National Center for Stomatology & National Clinical Research Center for Oral Diseases, Department of Head and Neck Oncology, West China Hospital of Stomatology, Sichuan University, Chengdu 610041, Sichuan, China. E-mail: lichunjie@scu.edu.cn; zhiyongguo@scu.edu.cn

validation, task-specific error analysis, procedure-specific datasets, and evaluation in real-world clinical workflows, along with further development of three-dimensional, multimodal, and image-guided applications.

## INTRODUCTION

Oral and maxillofacial surgical care relies on visual information throughout diagnosis, treatment planning, perioperative assessment, and postoperative follow-up. Radiographs, cone-beam computed tomography (CBCT), clinical photographs, and procedural visual data are used to localize anatomical structures, identify pathologic findings, assess surgical risk, and support clinical decision-making[1-3]. These tasks can be complicated by anatomical overlap, small or poorly defined targets, variable acquisition conditions, and the need for rapid interpretation.

You Only Look Once (YOLO) is a family of single-stage object detection models designed for rapid spatial localization[4]. In dentistry and oral and maxillofacial applications, YOLO-based models have been investigated for tooth and landmark localization, lesion and fracture detection, mucosal lesion assessment, implant-related analysis, and treatment planning[4-7], as well as instrument recognition[8-11]. Surgical computer vision studies in other specialties have further explored instrument tracking, occlusion handling, small-target detection, and orientation-aware localization, providing methodological references for procedural applications[3].

Recent reviews have examined YOLO applications in dentistry from different perspectives. A systematic review of 73 studies evaluated detection and segmentation across dental imaging tasks, highlighting inconsistent performance reporting and limited external validation[12]. A narrative review surveyed a broader range of dental applications, including diagnosis, restorative and prosthodontic care, implant planning, education, and salivary biomarker analysis[13]. Although both reviews cover diagnostic and implant-related research relevant to oral and maxillofacial care, their findings are synthesized mainly by imaging task or dental discipline. This review instead examines the clinical role of YOLO-based computer vision across the oral and maxillofacial surgical-care continuum, from preoperative diagnosis and planning to procedural vision and post-treatment assessment. We synthesize 61 representative studies, distinguishing direct surgical-care evidence from supporting dental research and selected cross-specialty methodological examples. Particular attention is given to the role of YOLO within multicomponent systems, the relationship between model outputs and downstream clinical tasks, task-specific failure modes, validation settings, and requirements for clinical translation. Figure 1 illustrates the clinical scope of the review and the three categories of evidence.

## LITERATURE SEARCH AND REVIEW APPROACH

This narrative review was informed by structured searches of PubMed, Scopus, and the Web of Science Core Collection for studies published or available online from January 2015 through 2 September 2026. Google Scholar was used for supplementary searches and citation tracking. Search terms combined "YOLO", "You Only Look Once", and individual YOLO versions with terms related to dentistry, oral and maxillofacial surgery, surgical imaging, computer vision, dental and craniofacial imaging, treatment planning, surgical instruments, procedural video, object tracking, intraoperative detection, and image-guided applications. The search syntax was adapted for each database. The complete database-specific search strategies, execution dates, search limits, and retrieval counts are provided in Supplementary Table 1.

Titles and abstracts were screened for relevance, with automated triage used to prioritize potentially eligible records for further review. These preliminary flags were not treated as final inclusion or exclusion decisions.

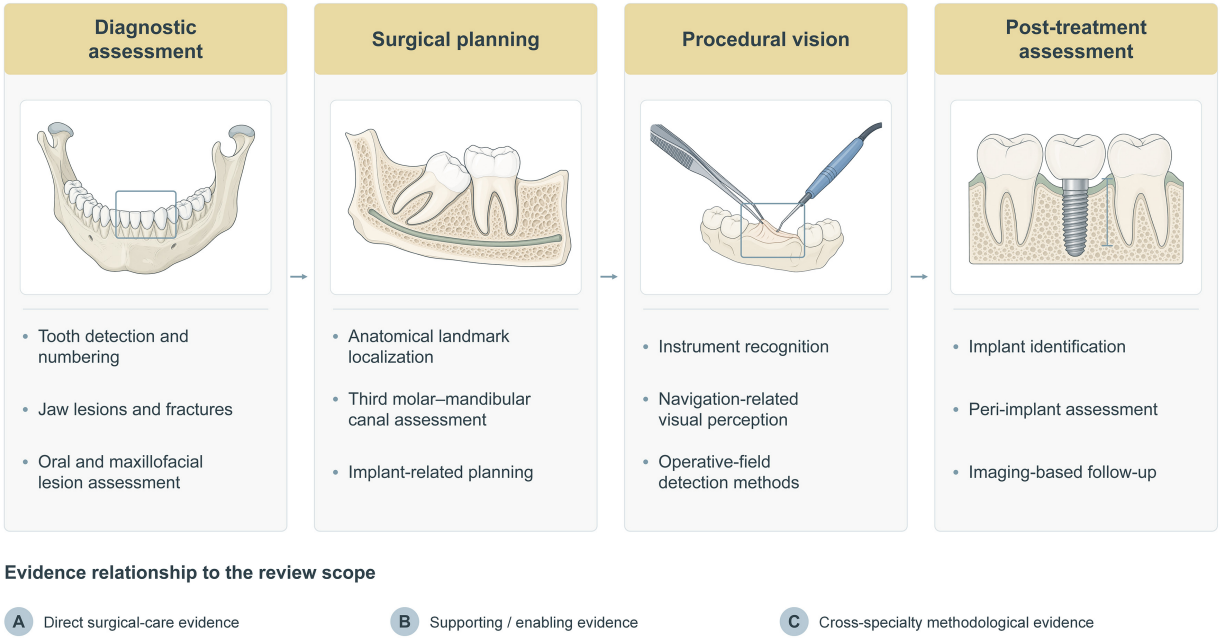



**Figure 1.** YOLO-related visual tasks along the oral and maxillofacial surgical-care continuum. The schematic organizes representative research tasks into diagnostic assessment, surgical planning, procedural vision, and post-treatment assessment. Categories A, B, and C represent direct surgical-care, supporting or enabling dental, and cross-specialty methodological evidence, respectively. YOLO: You Only Look Once.

The primary synthesis included studies applying YOLO or YOLO-based architectures to radiographic, tomographic, ultrasound, photographic, or procedural visual data relevant to oral and maxillofacial diagnosis, anatomical localization, treatment planning, perioperative assessment, implant-related evaluation, or procedural support.

Because direct evidence for continuous procedural computer vision in oral and maxillofacial surgery remains limited, selected studies from other surgical specialties were included when they addressed methodological challenges relevant to oral and maxillofacial procedures, such as instrument detection, temporal tracking, small-target recognition, occlusion handling, dataset construction, and orientation-aware localization. These studies were used as methodological references rather than as direct evidence of clinical performance in oral and maxillofacial surgery. For the narrative synthesis, studies were grouped into direct surgical-care evidence (A), supporting or enabling dental evidence (B), and cross-specialty methodological evidence (C).

The three core databases yielded 2,721 records (Scopus, $n$ = 1,052; PubMed, $n$ = 656; Web of Science, $n$ = 1,013). After removal of 971 duplicates, 1,750 unique records remained [Figure 2]. Of these, 61 representative original studies and methodological benchmarks were purposively selected for narrative synthesis based on clinical relevance, the role of YOLO, the validation setting, and methodological contribution (A, $n$ = 18; B, $n$ = 33; C, $n$ = 10).

The literature was synthesized according to clinical task, visual-data modality, the role of YOLO within the analytical pipeline, validation characteristics, reported performance, and relevance to clinical translation. We distinguished YOLO-specific results from downstream model performance and examined external validation, comparisons with clinicians, and clinical workflow evaluation, where reported.

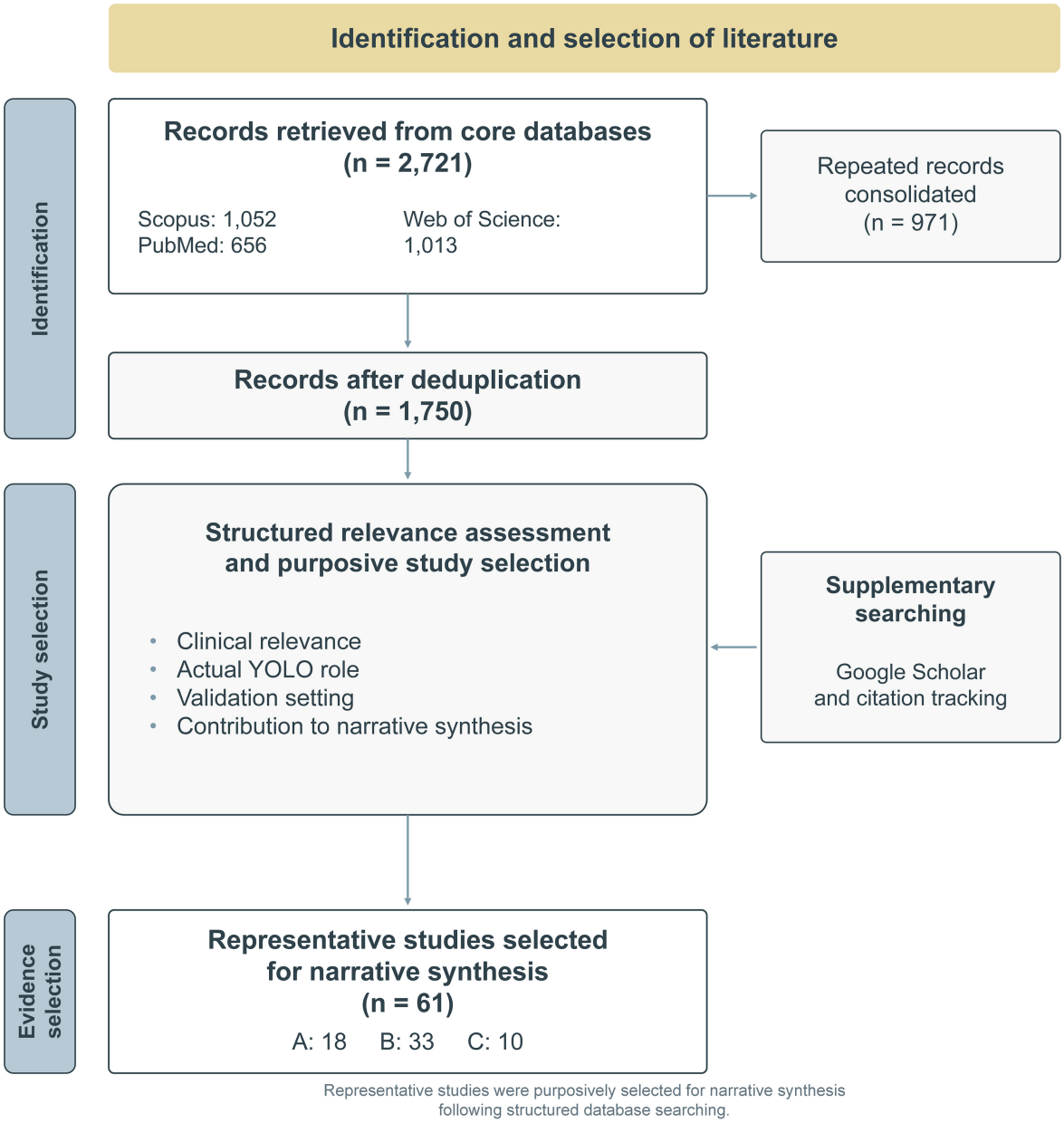



**Figure 2.** Literature identification and purposive selection of representative studies for narrative synthesis. Records were retrieved from the core databases (Scopus, PubMed, and Web of Science), with Google Scholar and citation tracking used as supplementary sources. A: Direct surgical-care evidence; B: supporting or enabling dental evidence; C: cross-specialty methodological evidence; YOLO: You Only Look Once.

## YOLO-BASED COMPUTER VISION AND ITS RELEVANCE TO ORAL AND MAXILLOFACIAL SURGICAL CARE

### Principles and evolution of YOLO-based object detection

The original YOLO framework introduced a unified, single-stage approach to object localization and classification[14]. Successive YOLO versions have modified feature extraction, multi-scale feature aggregation, and detection heads, while improving computational efficiency[15]. These developments have produced models with different trade-offs among accuracy, speed, and computational demands. A simplified overview of major developments across the YOLO family is provided in Figure 3.

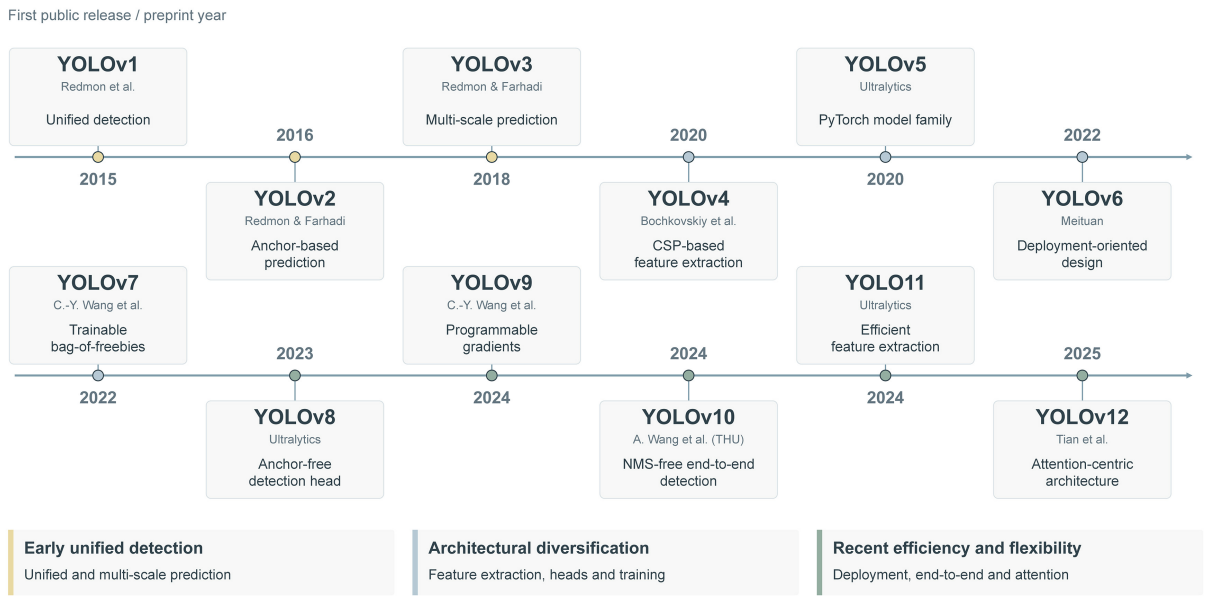



**Figure 3.** Major developments across the YOLO family. The simplified timeline outlines representative detection-model developments and parallel architectural directions. YOLO: You Only Look Once; CSP: cross-stage partial; NMS: non-maximum suppression; THU: Tsinghua University.

### Task characteristics relevant to surgical imaging and computer vision

Because YOLO produces object-level spatial outputs rather than only image-level labels, detected regions can be used as regions of interest for subsequent analysis. In multistage systems, YOLO has been used as a front-end localization component preceding downstream classification, measurement, or risk prediction[16-19]. The performance of the final output therefore reflects the complete analytical pipeline rather than the detector alone.

Localization outputs can also serve as inputs to quantitative assessment and planning pipelines. Existing applications include preliminary implant-pathway estimation, radiographic assessment of third-molar extraction difficulty, and CBCT-based anatomical measurements linked to treatment recommendations[20-22]. In these systems, spatial information generated by YOLO is combined with measurements, predefined criteria, image-processing steps, or other downstream components.

In procedural video, frame-level detections can be linked across consecutive frames using tracking algorithms, preserving target identity and enabling analysis of movement over time[23,24]. Orientation-aware detection can also encode the direction of an instrument or implanted object[25]. Efficient frame-level inference may facilitate procedural video analysis, as demonstrated by YOLO-based dental bur detection in controlled imaging experiments[8].

## PREOPERATIVE DETECTION AND DIAGNOSTIC SUPPORT

YOLO-based diagnostic studies in oral and maxillofacial imaging vary in target definition, acquisition standardization, and the extent to which localization contributes to a clinically interpretable output. Anatomically discrete targets, such as teeth and cephalometric landmarks, generally allow reproducible annotation and direct visual verification. In contrast, lesions, fractures, and mucosal abnormalities exhibit greater variation in morphology and appearance, increasing the importance of appropriate reference standards, class-specific performance, and validation design.

**Table 1. Representative YOLO-based studies of tooth detection, numbering, and dental structure recognition**

| Clinical task | Imaging modality, dataset, and validation | YOLO model/role | Main findings | Clinical interpretation and evidence limitations |
|---|---|---|---|---|
| Tooth detection, numbering, and impacted-tooth classification | Panoramic radiographs; 1,200 retrospectively selected images; internal model comparison[26] | YOLOv4 compared with Faster R-CNN | YOLOv4 achieved a precision of 99.90%, a recall of 99.18%, and an F1-score of 99.54% across 32 permanent-tooth and 4 impacted-tooth classes | High internal performance for anatomically structured targets; external validation and workflow-based evaluation were not reported |
| Tooth segmentation and numbering | Periapical radiographs; 6,446 images randomly divided into training, validation, and test sets; 644 test images[28] | YOLOv5 with polygonal segmentation | Of 2,578 test tooth labels, 2,434 were correctly predicted (94.4%); precision 0.9548, recall 0.9882, and F1-score 0.9712 | Technical feasibility was supported by a relatively large internal dataset; motion-degraded images were excluded, and validation used a random internal split |
| Permanent and deciduous tooth numbering with tooth-condition classification | Panoramic radiographs; 1,085 images from 10 facilities, including 994 permanent- and 91 mixed-dentition images[29] | YOLOv7 with rule-based post-processing | F1-score was 98.45% for tooth numbering and 95.47% across five tooth-condition classes | Data were acquired from 10 facilities and covered a broad age range; mixed dentition was underrepresented, and no independent site-level evaluation was performed |
| Pediatric tooth detection under domain shift | Pediatric panoramic radiographs; internal dataset of 200 images and external open-source dataset of 192 images; reciprocal cross-domain testing[30] | YOLOv8 and YOLOv10 | Cross-domain mAP ranged from 0.901 to 0.910; bite blocks and rare supernumerary teeth produced domain-specific errors | Provides direct evidence of external generalizability, with persistent domain-specific errors involving bite blocks and rare supernumerary teeth |

YOLO: You Only Look Once; R-CNN: region-based convolutional neural network; mAP: mean average precision.

## Tooth and dental structure detection

Tooth detection and numbering are among the most frequently investigated YOLO applications in dental imaging. Teeth are repeated, spatially ordered structures with relatively distinct boundaries, allowing model outputs to be checked directly against the radiograph. Studies using panoramic and periapical radiographs have demonstrated the feasibility of automated tooth localization, numbering, segmentation, and impacted-tooth recognition[26-28]. These characteristics make tooth-level analysis well suited to automated charting and region-of-interest generation.

Recent studies have examined more heterogeneous populations and validation conditions. A model trained on panoramic radiographs acquired from multiple facilities extended tooth numbering to both permanent and deciduous teeth and incorporated tooth-condition labels[29]. However, because images from all participating facilities contributed to model development, the study did not provide independent site-level evaluation[29]. In a pediatric study, reciprocal testing between internal and external datasets provided direct evidence of external generalizability[30]. Nevertheless, errors remained associated with device-related structures, such as bite blocks, and rare findings, including supernumerary teeth. Overall, evidence is most consistent for routine tooth localization and numbering, whereas comprehensive tooth-condition assessment remains less well validated across populations and imaging environments [Table 1].

## Oral and maxillofacial pathology and trauma

Detection of pathologic and traumatic findings is more challenging because clinically relevant targets vary in size, density, border definition, anatomical site, and prevalence. Studies using panoramic radiographs have evaluated cystic lesions, odontogenic lesions, idiopathic osteosclerosis, radiopaque and radiolucent abnormalities, and mandibular fractures[31-35]. Technical feasibility has been demonstrated for predefined abnormalities, although performance varies substantially across disease and fracture categories. Anatomically complex regions and less frequent classes remain more challenging, and aggregate metrics may obscure clinically important patterns of false-negative detection.

Cross-sectional imaging has extended these applications to additional anatomical targets. Maxillary sinus structures and associated abnormalities have been segmented on selected coronal images derived from CBCT examinations[36], while zygomatic fractures have been detected on axial and coronal computed tomography (CT) images[37]. However, analysis of selected two-dimensional slices does not constitute complete volumetric assessment, and slice-based validation requires clear patient-level separation across data partitions.

YOLO also serves as a localization component in multicomponent diagnostic pipelines. In the evaluation of radiopaque and radiolucent lesions, YOLO was used for detection and segmentation, while separate convolutional networks performed lesion classification[35]. In multimodal screening for degenerative temporomandibular joint disease, YOLO localized the condylar region, while disease classification incorporated additional image-analysis components and demographic information[16]. Similarly, a retrospective magnetic resonance imaging (MRI) study of early-stage oral tongue squamous cell carcinoma used YOLOv8 to detect primary tumors before combining image-derived features with the clinical T stage to predict lymph node metastasis[38].

Collectively, these studies illustrate the dual role of YOLO in standalone lesion localization and multistage diagnostic assessment [Table 2].

### Oral mucosal lesion detection and referral

Photographic assessment of oral mucosal lesions is less standardized than radiographic imaging and depends more heavily on illumination, camera angle, image distance, lesion exposure, and operator technique. Visual overlap among benign, potentially malignant, and malignant conditions further complicates classification. YOLO-based systems have been investigated for the detection of oral potentially malignant disorders, referral-oriented triage, broader mucosal lesion classification, and differentiation of clinically similar oral white lesions[39-42].

Earlier studies primarily established technical feasibility using retrospective datasets and internal validation. A more recent study included a held-out image test set and comparisons with both general dentists and oral surgery specialists[43]. In this comparison, aggregate model performance was comparable to that of the participating specialists and exceeded that of the participating general dentists. Performance varied substantially across lesion categories, with particularly poor detection of leukoplakia. Overall metrics therefore did not uniformly reflect performance for lesions requiring timely referral or biopsy.

Current evidence therefore aligns most directly with lesion documentation and referral triage, for which relevant evaluation outcomes include acquisition consistency and category-specific referral errors [Table 3].

### Craniofacial and cephalometric image analysis

Cephalometric landmark localization is a relatively structured YOLO application because the landmarks are spatially defined and localization errors can be quantified directly. Earlier work demonstrated that multi-scale YOLO models could identify landmarks on lateral and frontal cephalograms within predefined distance thresholds[44]. However, the reliability of derived angular measurements, skeletal classifications, and treatment-planning outputs requires separate evaluation.

A more recent study linked YOLO-based detection of four cephalometric landmarks to skeletal classification based on the A-point-Nasion-B-point (ANB) angle and compared the resulting classifications with assessments by orthodontists[45]. This design connected localization performance to a downstream diagnostic output. The study evaluated 12 YOLO model configurations using 11 independent test cephalograms.

**Table 2. Representative YOLO-based studies of pathologic and traumatic findings in oral and maxillofacial imaging**

| Clinical task | Imaging modality, dataset, and validation | YOLO model/role | Main findings | Clinical interpretation and evidence limitations |
|---|---|---|---|---|
| Nasopalatine duct cyst and periapical cyst detection | Panoramic radiographs; 1,209 lesion-containing images divided into training, validation, and test sets; 60 additional negative images for clinician comparison[31] | YOLOv5 compared with EfficientDet-D3, Faster R-CNN, RetinaNet, and SSD | YOLOv5 achieved an mAP50 of 89.5%, compared with 93.8% for EfficientDet-D3 | Demonstrates internal feasibility of cyst localization; the clinician comparison was performed with EfficientDet-D3 rather than YOLOv5 |
| Multiclass odontogenic lesion detection | Panoramic radiographs; retrospective internal dataset[34] | Improved YOLOv3 | Overall classification accuracy was 95.6%; sensitivity was 98.4% for odontogenic keratocysts | High internal performance for predefined odontogenic lesion classes; external evaluation was not reported |
| Primary oral tongue tumor localization within a lymph node metastasis risk-prediction pipeline | Two-center retrospective MRI cohort; 348 patients (163 training, 65 validation, 84 internal test, 36 external test)[38] | YOLOv8 for primary-tumor localization; downstream imaging/clinical fusion model for risk prediction | The external-test AUC was 0.909 for the complete downstream predictive pipeline, not for the YOLO detector | Provides externally tested proof of concept for multistage risk prediction; the small external cohort and retrospective design do not establish improvement in surgical decisions or outcomes |
| Mandibular fracture localization and identification | 1,710 panoramic radiographs (855 fracture-positive, 855 fracture-negative; 1,423 fracture sites); fracture-positive images split 70:10:20 by accession number[32] | YOLOv5 for bounding-box localization; DenseNet-169 and ResNet-50 for separate image-level classification | YOLOv5: precision 86.12%, recall 92.23%, F1-score 89.07%, PR-AUC 0.90; a 100-image clinician comparison assessed classification | Retrospective, single-center evaluation without external validation. The clinician comparison concerned classification models, not YOLOv5 localization |
| Radiopaque and radiolucent lesion detection and segmentation | Panoramic radiographs; 1,000 images; retrospective internal evaluation[35] | YOLOv8m for detection and segmentation; separate CNNs for lesion classification | mAP was 71.5% for detection and 72.1% for segmentation | Detection and segmentation performance was moderate; lesion classification was performed by separate CNNs rather than YOLO |
| Maxillary sinus and sinus-pathology segmentation | CBCT-derived coronal images; 307 patients and 1,377 selected images; single-center internal evaluation[36] | Modified YOLOv5x instance-segmentation model | F1-scores were 0.992 for the maxillary sinus, 0.889 for mucosal thickening, 0.924 for mucous retention cysts, and 0.970 for sinusitis | High internal segmentation performance was reported on selected two-dimensional CBCT-derived images rather than complete volumetric examinations |
| Zygomatic fracture detection at seven anatomical sites | Spiral CT; 283 scans yielding 13,988 axial and 14,107 coronal slices; reported 6:2:2 split[37] | YOLOv8 | Recall exceeded 90% across fracture sites; AP ranged from 0.692 to 0.827 | Patient-level independence of the reported slice-based data split was not clearly established |
| Screening for degenerative TMJ disease | Panoramic radiographs, lateral cephalograms, age, and sex; 500 patients and 1,000 TMJs; CBCT diagnosis as the reference standard[16] | YOLOv8 for condylar localization; EfficientNetV2 and multimodal fusion for disease classification | YOLO localization mAP50 was 0.995; the complete multimodal model achieved an AUC of 0.898 | YOLO performed condylar localization within a multimodal pipeline; disease classification was generated downstream. Evaluation was single-center and retrospective |

YOLO: You Only Look Once; R-CNN: region-based convolutional neural network; SSD: Single Shot MultiBox Detector; mAP50: mean average precision at an intersection-over-union threshold of 0.50; MRI: magnetic resonance imaging; PR-AUC: area under the precision-recall curve; AUC: area under the curve; AP: average precision; CBCT: cone-beam computed tomography; CNN: convolutional neural network; CT: computed tomography; mAP: mean average precision; TMJ: temporomandibular joint.

Structured craniofacial image analysis has also been explored in prenatal ultrasound, where YOLO was used to identify standard imaging planes relevant to fetal lip and palate assessment[5]. The task involved image-plane recognition and acquisition support rather than direct diagnosis of cleft lip or palate. Across these applications, YOLO serves primarily for reproducible localization or image-plane recognition preceding downstream measurement, classification, or expert interpretation [Table 4].

## SURGICAL PLANNING AND PERIOPERATIVE DECISION SUPPORT

YOLO-based planning studies in oral and maxillofacial surgery mainly address anatomical risk assessment, quantitative measurement, and structured decision support. In most systems, YOLO provides localization or segmentation, whereas measurements, clinical indices, or additional algorithms generate the final planning output.

**Table 3. Representative YOLO-based studies of oral mucosal lesion assessment using intraoral photographs**

| Clinical task | Imaging modality, dataset, and validation | YOLO model/role | Main findings | Clinical interpretation and evidence limitations |
|---|---|---|---|---|
| Oral potentially malignant disorder detection and classification | Intraoral photographs; retrospective internal dataset[40] | YOLOv4-based lesion localization and classification | The model demonstrated the feasibility of localizing and classifying OPMDs from clinical photographs | Early proof-of-concept evidence based on retrospective internal validation; external testing, clinician comparison, and prospective screening evaluation were not reported |
| Referral-oriented oral mucosal lesion triage | White-light intraoral photographs; 6,903 original images expanded to more than 50,000 through augmentation[39] | YOLOv7 variants with coordinate attention | Lesions were categorized into benign, potentially malignant, and malignant referral groups; coordinate attention improved model performance | Referral-oriented categories provide a clinically interpretable output; validation relied on augmented data from the available dataset without independent patient-level or external testing |
| Detection and 11-class classification of oral mucosal lesions | Intraoral photographs; 4,079 images, including a held-out test set of 282 images; comparison with six general dentists and three oral surgery specialists[43] | YOLOv8 for image classification and object detection | Overall performance was comparable to that of specialists and higher than that of general dentists; leukoplakia detection was substantially weaker, with AP50 of 0.20 | Includes held-out test images and a clinician comparison; patient-level independence and external validation require separate confirmation. Performance varied substantially by class, particularly for leukoplakia |

YOLO: You Only Look Once; OPMD: oral potentially malignant disorder; AP50: average precision at an intersection-over-union threshold of 0.50.

**Table 4. Representative YOLO-based studies of cephalometric and structured craniofacial image analysis**

| Clinical task | Imaging modality, dataset, and validation | YOLO model/role | Main findings | Clinical interpretation and evidence limitations |
|---|---|---|---|---|
| Cephalometric landmark localization | Lateral and frontal cephalograms; retrospective internal evaluation[44] | Multi-scale YOLOv3 | Successful detection rates were 80.84%, 93.75%, and 98.14% within 2-, 3-, and 4-mm thresholds, respectively | Demonstrates automated landmark localization within predefined distance thresholds; downstream angular measurements and treatment-related outputs were not evaluated |
| Landmark detection and ANB-based skeletal classification | Lateral cephalograms; 120 images from one center, with 11 held-out test cephalograms; expert reference from four orthodontists[45] | YOLO-based localization of Sella, Nasion, A-point, and B-point followed by ANB classification | Best-model mean radial error: 3.10 mm; downstream ANB-based skeletal classification was evaluated against expert consensus | Links landmark localization to downstream skeletal classification. Evaluation was retrospective and limited to 11 independent test cephalograms from a single center; prospective clinical validation was not reported |
| Identification of fetal cleft lip and palate standard views | Prenatal ultrasound images; evaluation restricted to predefined standard imaging planes[5] | YOLOv5-based standard-plane detection | AUC values were 0.971, 0.988, and 0.977 for the three standard planes | The task involved standard-plane recognition and acquisition support rather than direct cleft diagnosis; nonstandard or routine acquisition conditions were not evaluated |

YOLO: You Only Look Once; ANB: A-point-Nasion-B-point angle; AUC: area under the curve; AI: artificial intelligence.

## Anatomical risk assessment for surgical planning

The relationship between mandibular third molars and the inferior alveolar canal is a major focus of YOLO-based preoperative assessment. Studies using panoramic radiographs have estimated root-canal contact or proximity, using CBCT assessment as the reference standard[46-48]. These approaches may help identify cases that require further three-dimensional assessment.

A YOLO11-based system also assessed third-molar extraction difficulty using predefined radiographic criteria[21]. The model combined radiographic features into a structured difficulty category, extending YOLO from anatomical localization to preoperative risk assessment. However, the resulting category reflects radiographic criteria rather than observed operative difficulty or complication risk.

### Image-based planning and quantitative decision support

YOLO has also been used as a front-end component in quantitative planning pipelines. One implant-planning framework identified edentulous regions and the orientation of adjacent teeth on panoramic radiographs, followed by additional image-processing steps to generate a preliminary implant pathway[20]. The predicted pathway showed close angular agreement with clinician-defined planning.

A comparative implant-planning study evaluated YOLOv7 and Faster Region-based Convolutional Neural Network (Faster R-CNN) using implant-site images displayed in four implant-planning software packages[49]. The relative detection and localization results varied across software environments, and YOLOv7 was not consistently superior. Because the software comparisons involved different renderings of a limited set of test images, they do not constitute independent multicenter validation.

Another system used YOLOv8-seg on CBCT cross-sections to segment mandibular alveolar bone and the inferior alveolar canal. Automated measurements were then linked to suggested treatment options[22].

In a separate CBCT study involving 192 patients, YOLOv8-seg was compared with other segmentation architectures for mandibular bone and inferior alveolar canal delineation on two-dimensional cross-sectional images[50]. Although the reported segmentation agreement was high, the authors identified false-positive canal regions in anterior mandibular sections where no canal was present. This example illustrates why anatomically implausible predictions require specific assessment beyond average Dice scores.

## PROCEDURAL COMPUTER VISION AND INTRAOPERATIVE APPLICATIONS

Evidence for procedural computer vision spans several distinct settings: dental instrument management, navigation-related visual perception evaluated under laboratory or offline conditions, and operative-field detection studied in other surgical specialties. Continuous tracking, occlusion handling, and orientation-aware localization have been explored in other surgical specialties, providing methodological insights for oral and maxillofacial applications. Translation to oral and maxillofacial procedures requires procedure-specific datasets, task definitions, and spatiotemporal evaluation.

### Dental instrument recognition and navigation-related prototypes

Published dental YOLO studies have addressed instrument management, recognition, and detection under controlled conditions. A YOLOv8-based system detected and classified 14 restorative dental instruments arranged on standardized trays, supporting automated instrument identification in dental supply workflows[9]. Related work showed that annotation of characteristic instrument regions could improve recognition efficiency compared with whole-instrument labeling[10]. A public dataset comprising 22 categories of dental surgical instruments and 847 images has also been released for YOLO-based model development[11]. Collectively, these studies encompass dental instrument recognition under controlled imaging conditions and the development of annotated datasets for model training and evaluation.

Dental bur detection is another controlled object-recognition task. An optimized YOLO-based model incorporating asymmetric convolution and adaptive feature fusion achieved a mean average precision at an intersection-over-union threshold of 0.50 (mAP50) of 99.3% and a detection speed of 128 frames per second (FPS) in experiments using tray-based dental bur images[8].

Navigation-related studies have begun to evaluate perception components more closely aligned with procedural use. A system combining YOLO11 with an RGB-D (color and depth) camera identified anatomical landmarks and supported three-dimensional registration for robot-assisted implant surgery on a dental training phantom[51]. A separate YOLO-based instance-segmentation study examined marker-free tooth localization as a front end for dynamic implant navigation[52]. Its processing speed was evaluated using previously collected image frames rather than during live robotic navigation.

**Table 5. Representative cross-domain surgical computer vision approaches relevant to oral and maxillofacial procedural applications**

| Methodological capability | Surgical context | Representative approach | Main technical contribution | Methodological relevance to oral and maxillofacial procedural computer vision |
|---|---|---|---|---|
| Real operative-field instrument detection | Open neck procedures; recorded surgical video from 22 procedures | YOLOv3 electrocautery detection[53] | Held-out image testing across different procedures yielded an F1 ≈ 0.906; recorded-video analysis was also explored | Genuine head-and-neck operative imagery; single-instrument detection with false positives from visually similar objects, not demonstrated tracking or intraoperative guidance |
| Intraoperative anatomical landmark localization | Neurosurgical microscopic operative video; 78 videos from 76 patients | AENEAS, YOLOv7x[54] | Recognized neurovascular landmarks in microscope imagery; mean AP50 of approximately 0.39 with substantial between-structure variation | Shows the feasibility of target-specific anatomical localization in a true operative field; five-fold internal cross-validation does not constitute independent external or OMFS-specific validation |
| Dataset construction and benchmarking | Robotic surgery | SID-RAS dataset; YOLOv10 and YOLOv11[55] | Established a high-resolution dataset of 6,000 images covering nine robotic instrument categories, with YOLO-based benchmark evaluations | Provides a reference for instrument taxonomy, annotation protocols, and benchmark construction |
| Discrimination of visually similar instruments | Laparoscopic surgery | DBH-YOLO[57] | Used dual-branch detection heads to improve discrimination between instruments with similar appearance | Relevant to dental and maxillofacial instruments with similar shapes and overlapping visual features |
| Detection of small or bifurcated instrument tips | Laparoscopic surgery | Modified YOLOv8 with random object masking and Mixup Plus[58] | Reduced detection errors involving small and bifurcated instrument tips | Relevant to burs, needles, forceps tips, and other small operative targets |
| Fine-grained multi-tool tracking | Laparoscopic surgery | SurgiTrack[24] | Enabled simultaneous identification and temporal tracking of multiple instrument categories in surgical video | Provides a framework for preserving instrument identity and analyzing tool use over time |
| Tracking during rapid motion and partial occlusion | Laparoscopic surgery | YOLOv9n with ByteTrack and BoT-SORT[23] | Improved continuity of instrument trajectories during rapid movement and partial occlusion | Relevant to maintaining instrument trajectories in dynamic operative scenes |
| Occlusion-resilient instrument segmentation | Robotic minimally invasive surgery | YOLO with GAN-based occlusion modeling[59] | Improved instance segmentation of partially occluded surgical instruments | Relevant to instrument-tissue overlap and intermittent instrument visibility |
| Orientation-aware localization | Spinal fusion surgery | YOLOv8 with oriented bounding boxes[25] | Combined object localization with estimation of screw orientation | Provides a reference for representing instrument, implant, or trajectory direction in addition to location |

These cross-domain studies are included as methodological references and do not constitute direct evidence of clinical performance in oral and maxillofacial procedures. YOLO: You Only Look Once; GAN: generative adversarial network; SID-RAS: surgical instrument dataset for robot-assisted surgery; OMFS: oral and maxillofacial surgery.

## Operative-field evidence and methods from other surgical specialties

YOLO-based detection has also been investigated using recorded surgical videos from other specialties. A YOLOv3 model was used to detect electrocautery instruments in recordings from 22 open neck procedures, achieving an image-level F1-score of approximately 0.906 on a procedure-separated test set. Detection performance was further evaluated across 228 minutes of recorded surgical video[53]. In neurosurgical microscope footage, a YOLOv7x-based system localized anatomical structures, reporting an overall AP50 of approximately 0.39, with class-specific values of 0.73 for the optic nerve and 0.67 for the internal carotid artery[54].

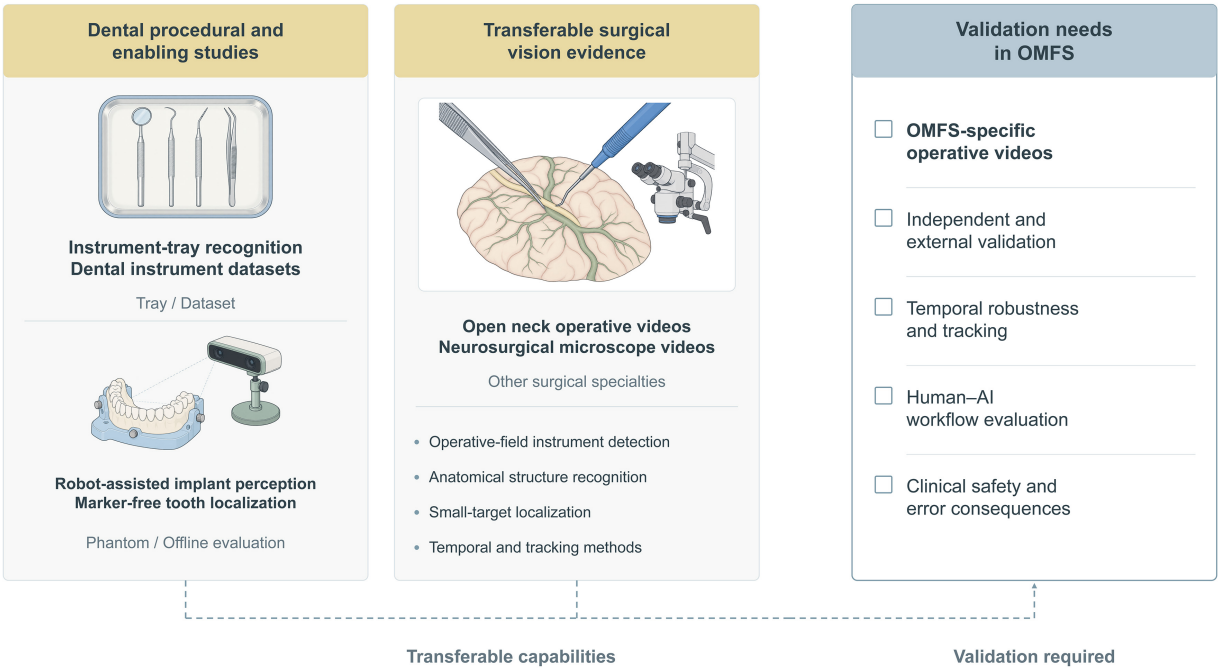



**Figure 4.** Procedural computer vision: existing evidence and validation needs in OMFS. The schematic summarizes controlled dental instrument and navigation studies alongside operative-video research from other surgical specialties and outlines considerations for evaluation in oral and maxillofacial procedures. Dashed connections indicate methodological relevance across these settings. OMFS: Oral and maxillofacial surgery; AI: artificial intelligence.

Other cross-specialty work has addressed dataset development and robust instrument recognition. The SID-RAS dataset established instrument categories and benchmark data for robotic surgical instrument detection[55]. Architectural optimization has also been explored for surgical instrument detection, including feature-fusion and attention-based modifications in UK-YOLOv10[56]. Other studies have addressed visually similar instruments[57], small or bifurcated instrument tips[58], and partial occlusion[59]. These approaches target failure modes that arise when only a small portion of an instrument is visible or when different instruments share similar visual features. Representative approaches are summarized in Table 5.

Temporal and spatial representations further extend procedural computer vision beyond single-frame detection. SurgiTrack supports fine-grained multi-class and multi-tool tracking across surgical videos[24], whereas YOLO combined with ByteTrack and BoT-SORT has been used to maintain instrument trajectories during rapid motion and partial occlusion[23]. Orientation-aware detection using oriented bounding boxes has also been applied to estimate the orientation of implanted hardware in spinal surgery[25].

## Requirements for translation to oral and maxillofacial surgery

Translation to oral and maxillofacial procedures requires datasets and task definitions that reflect the operative environment. Instrument categories, anatomical targets, procedural phases, and annotation rules should be defined for specific procedures. A multicenter endoscopic-video study of minimally invasive submandibular gland resection examined surgical workflow recognition and workflow-based skill assessment[60]. Continuous operative video is particularly important for tracking tasks because frame-level image datasets do not capture temporal changes in instrument visibility, motion, or interaction with anatomy[10,11,24,55]. Figure 4 summarizes the reported procedural-vision settings and associated validation needs.

The visual conditions of oral and maxillofacial procedures also place specific demands on detection models. Small instrument tips may occupy only a limited portion of the image, while tissue overlap, specular reflections, blood or saliva, motion blur, and changing viewpoints can alter target visibility[8,58,59]. Soft-tissue deformation and movement further alter the spatial relationships between instruments and anatomical structures. Evaluation should therefore include performance under these procedure-specific conditions.

Procedural guidance requires richer spatial and temporal information than object detection alone provides. Instrument identity must be maintained across frames, and some tasks require estimation of trajectory or orientation[23-25]. Image-guided applications that combine preoperative imaging with intraoperative visual information additionally require spatial correspondence between image-derived anatomy and the operative field. Evaluation should distinguish among offline frame-level performance, retrospective video analysis, laboratory or phantom testing, and prospective integration into clinical workflows. In related work, robotic platforms for head-and-neck surgery have been evaluated in porcine models and a prospective clinical series[61,62]. These requirements distinguish procedural computer vision from frame-level instrument recognition.

## POSTOPERATIVE AND IMPLANT-RELATED ASSESSMENT

YOLO-based implant applications range from the identification of existing implant systems to the assessment of peri-implant conditions and implant-related outcomes. These tasks are relevant to follow-up, maintenance, and retrospective implant evaluation.

### Dental implant identification

YOLO-based implant identification has been investigated using panoramic and periapical radiographs for implant localization, segmentation, and classification by manufacturer or implant type[17,63-65]. These studies generally report favorable performance when candidate categories are limited to implant systems represented during model development. Most evaluations use closed-set classification, in which the model must select among a predefined set of known brands. Evidence remains limited for rare or discontinued implants, unknown-class rejection, and discrimination between visually similar systems that are poorly represented in the training data. Retrospective identification of implants with incomplete records represents a potential clinical application. In one two-stage system, YOLOv7 localized the implant region, while EfficientNet classified the manufacturer and implant type[17] [Table 6]. A separate study compared YOLOv12x with dental professionals in identifying five implant brands on a 100-image test subset, reporting higher accuracy for the model[64].

### Peri-implant and longitudinal assessment

YOLO-based methods have been applied to peri-implantitis grading, crown-implant ratio measurement, peri-implant bone-defect classification, marginal bone-loss assessment, and implant failure-risk prediction[6,18,19,66,67]. Current studies mainly assess a single radiograph or selected image section and therefore address cross-sectional peri-implant status rather than longitudinal changes over time. Performance also varies across disease categories. In marginal bone-loss grading, for example, F1-scores were lower for moderate and severe bone loss than for healthy and mild sites[6]. A study using panoramic radiographs also compared a YOLOv8 ensemble with two periodontal surgeons; their aggregate results differed from some severe-defect subgroup findings[67].

Some systems extend object localization into quantitative or prognostic pipelines. In crown-implant ratio assessment, YOLOv4 localized the prosthesis before convolutional neural network (CNN)-based key-point detection and measurement[18]. In implant failure-risk prediction, YOLOv8n localized the implant, while ResNet-34 generated the risk classification[19]. The reported prognostic performance therefore reflects the complete two-stage system rather than YOLO alone. Serial change detection, progression modeling, and independent cross-device validation remain understudied [Table 7].

**Table 6. Representative YOLO-based studies of dental implant detection, identification, and classification**

| Clinical task | Imaging modality, dataset, and validation | YOLO model/role | Main findings | Clinical interpretation and evidence limitations |
|---|---|---|---|---|
| Implant brand identification | Periapical and panoramic radiographs; 2,290 labeled images[63] | Comparison of YOLO variants, including YOLOv10m | Precision 0.931, recall 0.929, and mAP50 0.912 | High performance within a predefined set of implant brands; rare, discontinued, and previously unseen brands and unknown-class rejection were not evaluated |
| Implant detection and segmentation | Dental radiographs; 2,573 images containing six implant brands[65] | YOLOv8m-seg for implant detection and instance segmentation | mAP50 0.972; average detection time 1.3 s | High internal detection and segmentation performance; evaluation was limited to six implant brands, and external cross-device performance was not assessed |
| Manufacturer and implant-type identification | Panoramic radiographs; 1,574 images containing 3,675 implants[17] | YOLOv7 for implant localization followed by EfficientNet classification | F1-score 0.989 for the two-stage pipeline | YOLO performed implant localization, while manufacturer and implant-type classification was performed downstream by EfficientNet; evaluation was restricted to predefined implant categories |
| Implant brand classification with clinician comparison | Periapical radiographs; 5,851 images representing five brands; internal 80:10:10 split and clinician comparison using 100 held-out images[64] | Comparison of YOLOv7-YOLOv12 models; YOLOv12x selected for final evaluation | mAP50 0.989, mAP50-95 0.900, and accuracy 0.950; performance exceeded that of the participating dental professionals | Includes clinician comparison within a held-out internal test subset; evaluation remained limited to five known brands, without external multisite testing or unknown-class evaluation |

YOLO: You Only Look Once; mAP50: mean average precision at an intersection-over-union threshold of 0.50; mAP50-95: mean average precision at intersection-over-union thresholds 0.50-0.95.

## FROM MODEL PERFORMANCE TO SURGICAL TRANSLATION

Surgical translation requires more than optimization of model performance on a defined dataset. It also depends on clinically relevant validation, integration into real workflows, and evaluation of emerging three-dimensional, multimodal, and image-guided applications. Across the reviewed tasks, YOLO most often provides an object-level localization or segmentation output; the clinical meaning of that output depends on the downstream decision, the error type, and the evaluation setting [Figure 5].

### Task-oriented model selection and optimization

Comparative studies have not identified a consistently superior YOLO version across oral and maxillofacial applications[64,68-71]. Performance differences depend on dataset composition, annotation protocols, model size, training settings, and evaluation methods. A comparison across implant-planning software environments also found that detection and localization performance varied across image displays, and YOLOv7 was not consistently superior[49]. Model selection should therefore reflect the intended task, target characteristics, error consequences, and computational setting. Sensitivity and localization accuracy may be prioritized for small or low-contrast targets, whereas procedural or chairside applications additionally impose constraints on inference speed and computational demands.

Optimization strategies have addressed specific challenges, including small targets, overlapping anatomy, limited or imbalanced data, annotation burden, and computational constraints. Reported approaches include multi-scale feature fusion and attention mechanisms, pruning and lightweight architectures, targeted data augmentation, model ensembling, and task-specific annotation[8,10,39,72-74]. Representative strategies and their trade-offs are summarized in Table 8.

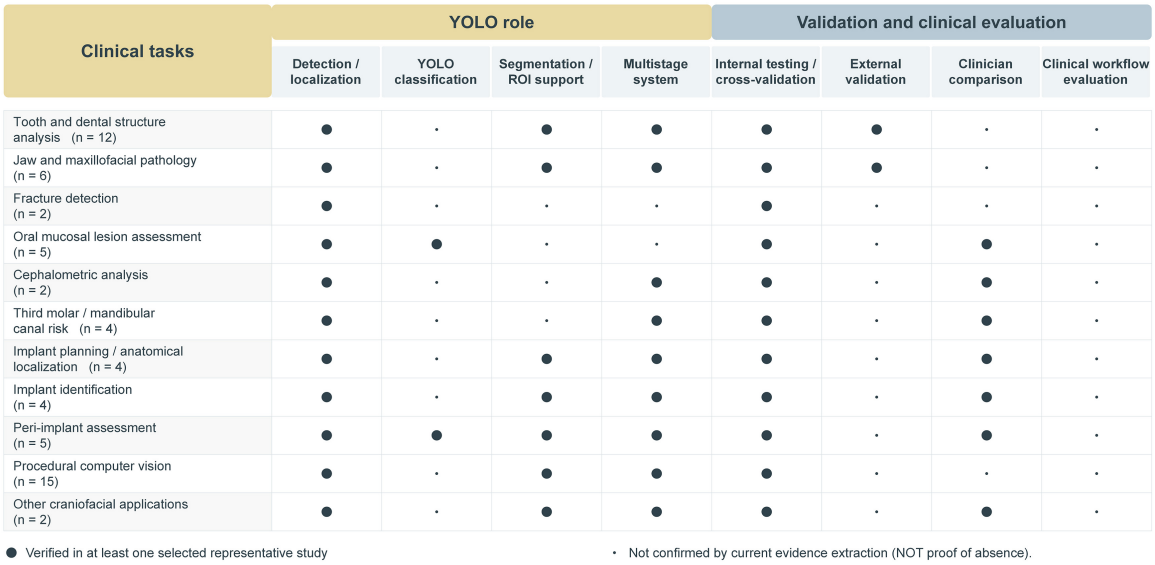



**Figure 5.** YOLO roles and validation characteristics among 61 selected representative studies. Filled circles (●) indicate characteristics documented in at least one selected study within each clinical task category, whereas dots (·) indicate characteristics not confirmed in the extracted evidence. Numbers in parentheses indicate the number of studies in each task category. YOLO: You Only Look Once; ROI: region of interest.

**Table 7. Representative YOLO-based studies of peri-implant assessment and implant-related outcome prediction**

| Clinical task | Imaging modality, dataset, and validation | YOLO model/role | Main findings | Clinical interpretation and evidence limitations |
|---|---|---|---|---|
| Peri-implantitis severity grading | Periapical radiographs; 800 images; retrospective internal evaluation[66] | YOLOv7 for detection and severity classification | Precision 1.000, recall 0.944, and F1-score 0.971 | High internal performance for predefined severity categories; evaluation was cross-sectional and did not assess longitudinal disease progression or external generalizability |
| Crown-implant ratio measurement | Dental radiographs; 1,106 images[18] | YOLOv4 for prosthesis localization followed by two CNNs for key-point detection and measurement | Automated and manual crown-implant ratio measurements did not differ significantly | YOLO performed prosthesis localization within a multicomponent measurement pipeline; the final quantitative output depended on downstream key-point detection and measurement |
| Peri-implant bone-defect classification | Panoramic radiographs; 1,075 images from 426 patients, covering 2,250 implant sites; comparison with two periodontal surgeons[67] | Ensemble of YOLOv8 models | Accuracy 0.853 and F1-score 0.854; overall accuracy of the two periodontal surgeons averaged 0.756 | Model accuracy exceeded the surgeons' mean overall result, but the surgeons performed better for selected severe-defect classes. All analyzed implants had peri-implantitis; healthy-site screening performance and external generalizability were not established |
| Marginal bone-loss detection and grading | 699 two-dimensional CBCT sections derived from 186 implants in 94 patients; sections randomly allocated to training, test, and validation sets (80:10:10)[6] | YOLOv8 for implant and marginal bone-loss detection and grading | Overall accuracy and F1-score were 0.900; mAP50 was 0.889. F1-score declined from 0.950 for healthy sites and 0.900 for mild loss to 0.620 and 0.700 for moderate and severe loss, respectively | F1-score was lower for moderate and severe bone loss. Assessment used selected two-dimensional sections rather than complete volumes; patient-/implant-level independence of the image split was not documented |
| Implant failure-risk prediction | Panoramic radiographs; 326 implants from 196 patients, including 58 implants lost within five years and 268 retained implants; stratified five-fold cross-validation[19] | YOLOv8n for implant localization followed by ResNet-34 for failure-risk classification | YOLO localization achieved mAP50 0.875; the classifier achieved an AUROC of 0.823 and an F1-score of 0.884 and was compared with two experienced clinicians | YOLO performed implant localization, while failure-risk prediction was generated downstream by ResNet-34. Evaluation was based on a single-center, imbalanced retrospective cohort without independent external testing |

YOLO: You Only Look Once; CNN: convolutional neural network; CBCT: cone-beam computed tomography; mAP50: mean average precision at an intersection-over-union threshold of 0.50; AUROC: area under the receiver operating characteristic curve.

**Table 8. Task-oriented optimization strategies for YOLO-based oral and maxillofacial imaging**

| Task, data, or deployment challenge | Optimization approach | Representative study or application | Interpretation and trade-offs |
|---|---|---|---|
| Dense or overlapping anatomical structures and targets with variable morphology | Deformable multi-scale feature fusion | DMAF-Net for dental structure detection on panoramic radiographs[73] | Designed to improve feature extraction and localization across scales in complex radiographs. Additional feature-fusion modules increase architectural complexity and should be evaluated with matched baselines and ablation analyses |
| Small or low-contrast targets under computational constraints | Model pruning and backbone modification | Lightweight YOLOv8 for tooth-crack detection in optical images[72] | Pruning reduced the number of parameters by 16.8% and GFLOPs by 24.3%, with limited loss of internal performance. Much of the dataset consisted of experimentally generated cracks, and external clinical performance was not evaluated |
| Limited or imbalanced training data | Targeted augmentation and attention-based feature enhancement | Coordinate-attention YOLOv7 models for oral mucosal lesion triage[39] | Augmentation increased variation within the available dataset, while attention emphasized lesion-related features. Independent patient-level or external data were not added, and class-specific performance remained important for rare or difficult lesions |
| Heterogeneous and non-standardized photographic acquisition | Model ensembling and transfer learning | Hybrid YOLO ensemble for caries detection in non-standardized dental photographs[74] | Ensembling combined predictions across models in heterogeneous photographic data but increased model complexity. Performance across different cameras, operators, illumination conditions, and clinical settings requires independent evaluation |
| Visually similar instruments and high annotation burden | Characteristic-region annotation | Dental instrument recognition using discriminative-region rather than whole-instrument labeling[10] | Region-focused annotation reduced labeling effort and emphasized class-specific features. Non-standard annotation definitions may limit comparability across datasets and omit contextual information from the complete instrument |
| Small-instrument detection with real-time requirements | Asymmetric convolution and adaptive feature fusion | Dental bur detection using an optimized YOLO-based architecture[8] | The architecture combined rapid detection with enhanced feature representation and reported rapid detector-level inference in a controlled, tray-based imaging setting. Robustness under intraoral occlusion, motion blur, blood or saliva contamination, variable magnification, and different hardware was not evaluated |

YOLO: You Only Look Once; DMAF-Net: deformable multi-scale adaptive fusion network; GFLOPs: giga floating-point operations.

Optimization should be evaluated against the failure mode it is intended to address. Additional feature-extraction modules may improve localization while increasing computational cost; pruning can reduce model size but may compromise performance on difficult targets; and data augmentation expands variation within the available dataset without adding new acquisition environments. Task-specific errors and deployment requirements should be considered when evaluating the value of model optimization. Detection accuracy, F1-score, average precision (AP), mean average precision (mAP), area under the curve (AUC), and Dice coefficient quantify different outcomes and are not directly comparable across tasks or study-specific test sets.

## Validation and clinically meaningful performance

Technical metrics such as precision, recall, mAP, and inference speed describe performance on a defined test dataset but do not establish clinical utility. Validation should be aligned with the intended use and supported by an appropriate reference standard. Data from the same patient or examination should not be shared across development and test sets, and independent external testing should be distinguished from internal evaluation[75,76]. The Checklist for Artificial Intelligence in Medical Imaging (CLAIM) and the artificial intelligence (AI) extension to the Standards for Reporting Diagnostic Accuracy (STARD-AI) provide reporting frameworks for documenting data sources, acquisition conditions, reference standards, test-set characteristics, and statistical uncertainty[75,76].

External evaluation should reflect the populations, institutions, imaging systems, and acquisition conditions relevant to the intended setting. Performance should also be examined by clinically important class, target, or subgroup rather than only through pooled metrics[75,76]. This is particularly important when uncommon findings or severe disease categories show lower sensitivity despite favorable aggregate results. The anatomically implausible false-positive canal regions in CBCT segmentation illustrate a task-specific error that an overall Dice score alone may not reveal[50]. In multicomponent systems, performance of the final diagnostic or planning output should be attributed to the complete pipeline, whereas the specific contribution of YOLO should be reported separately. For example, a mandibular fracture study compared clinicians with image-level CNN classifiers, whereas YOLOv5 was evaluated separately for fracture localization[32].

Clinical relevance also depends on how the system is intended to be used. For assistive applications, comparison of clinician performance with and without AI support is more informative than comparison between an unaided model and unaided clinicians. Outcomes may include diagnostic accuracy, interpretation time, referral decisions, planning decisions, and consistency among clinicians[76].

### Workflow integration and real-time deployment

After offline validation, including independent external testing, evaluation should extend to live clinical workflows. The Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by Artificial Intelligence (DECIDE-AI) guideline recommends reporting the clinical setting, intended users, workflow integration, user training, system inputs and outputs, and the role of AI-supported information in decision-making[77]. Early clinical evaluation should also document significant errors, safety concerns, user responses, and relevant human factors[77].

Real-time deployment requires more than rapid model inference. Dental bur recognition has demonstrated rapid frame-level detector inference under controlled conditions[8], whereas cross-domain studies have evaluated instrument detection and tracking in recorded surgical videos[23,24,53]. In a deployed system, performance also depends on image acquisition, preprocessing, tracking, output generation, and available hardware. Together, these components determine end-to-end latency and processing continuity during clinical use.

Workflow evaluation should assess whether the system can operate without disrupting the clinical procedure and how users respond when model outputs are uncertain or incorrect. Relevant outcomes include processing stability, interruption frequency, time added to the workflow, user interaction with model outputs, and recovery from detection or tracking failures. For procedural applications, these factors should be evaluated with the complete system under the conditions in which it is intended to operate.

### Research priorities for three-dimensional, multimodal, and image-guided applications

Three-dimensional analysis remains underdeveloped in current YOLO-based oral and maxillofacial applications. Existing CBCT studies commonly apply detection or segmentation to selected two-dimensional slices or cross-sections rather than complete volumetric data[6,22,36,50]. Future studies should evaluate volumetric localization and segmentation across complete CBCT examinations, with patient-level data separation and performance assessment across scanners, voxel sizes, and acquisition protocols. For surgical planning, evaluation should also include three-dimensional localization error and measurement reproducibility for anatomically relevant structures.

Multimodal systems offer another avenue beyond isolated image analysis. In screening for degenerative temporomandibular joint disease, YOLO was used to localize the condyle before image features from panoramic and cephalometric radiographs were combined with demographic information for downstream

classification[16]. Similarly, an MRI study of oral tongue cancer used YOLOv8 for primary-tumor localization before downstream imaging and clinical features were combined for lymph node metastasis risk prediction[38]. Future studies should define the contribution of each modality, evaluate whether multimodal input improves performance over appropriate unimodal baselines, and examine robustness when one source is unavailable or discordant.

Image-guided applications require integration of preoperative anatomical information with dynamic procedural data. Current oral and maxillofacial evidence includes preliminary image-based pathway estimation[20] and navigation-related visual perception tested using dental phantoms or prerecorded images[51,52]. Cross-domain studies provide methods for temporal tracking and orientation-aware localization[23-25], but image guidance additionally requires reliable spatial registration between image-derived anatomy, the operative field, and tracked instruments. Evaluation should therefore include registration accuracy, temporal stability, target-localization error, end-to-end latency, and failure behavior under realistic procedural conditions. Results from phantom or pseudo-stream tests cannot establish actual implant-placement precision or clinical benefit in patients.

## CONCLUSION

Published studies have examined YOLO-based computer vision across multiple stages of oral and maxillofacial surgical care, with the most consistent evidence currently supporting structured detection and localization tasks. In more complex applications, YOLO often serves as a localization, segmentation, or region-of-interest component within multistage systems rather than generating the final diagnostic, planning, or prognostic output alone. Procedural research to date includes dental navigation-perception prototypes and operative-field studies from other surgical specialties, but these studies do not establish routine intraoperative performance in oral and maxillofacial surgery. Further clinical translation will depend on independent validation, task-specific error analysis, procedure-specific datasets, and evaluation within real-world clinical workflows, together with continued development of three-dimensional, multimodal, and image-guided approaches.

## DECLARATIONS

### Authors' contributions

Made substantial contributions to the conception and design of the review: Shi J, Guo Z, Li C
Performed literature searching, study screening, and data curation: Shi J, Liu J, Li J, Zhu G
Drafted the manuscript: Shi J
Contributed to visualization: Li J
Critically revised the manuscript and provided supervision and project administration: Guo Z, Li C
Acquired funding: Zhu G, Guo Z, Li C
All authors read and approved the final manuscript.

### Availability of data and materials

Not applicable.

### AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool ChatGPT (version GPT-5.6 Sol, released 2026-07-09, OpenAI) was used solely for language editing and formatting adjustments. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

### Financial support and sponsorship

This work was supported by the Natural Science Foundation of Sichuan Province (2025ZNSFSC0645), the Health Commission of Sichuan Province Medical Science and Technology Program (24CXTD07), the Sichuan Provincial Youth Innovation Project for Medical Research (Q20250018), and the Research and Development Program of West China Hospital of Stomatology (RD-03-202506).

## Conflicts of interest

All authors declared that there are no conflicts of interest.

## Ethical approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Copyright

© The Author(s) 2026.

## Supplementary Materials

Supplementary Materials

## REFERENCES

1. Uranbey Ö, Gülbağ E, Ekinci B, et al. When CBCT looks borderline and standard radiology is inconclusive: should we plate or should we wait? *Diagnostics.* 2025;15:3140. DOI PubMed PMC
2. Ren R, Luo H, Su C, Yao Y, Liao W. Machine learning in dental, oral and craniofacial imaging: a review of recent progress. *PeerJ.* 2021;9:e11451. DOI PubMed PMC
3. Zhan B, Zhao W, Fang Y, et al. Tracking everything in robotic-assisted surgery. In: 2025 IEEE International Conference on Robotics and Automation (ICRA); 2025 May 19-23; Atlanta, GA, USA. IEEE; 2025. pp. 1-7. DOI
4. Vilcapoma P, Parra Meléndez D, Fernández A, et al. Comparison of faster R-CNN, YOLO, and SSD for third molar angle detection in dental panoramic X-rays. *Sensors.* 2024;24:6053. DOI PubMed PMC
5. Jiang W, Pang H, Deng S, Chen C. Artificial intelligence-based prenatal ultrasound for diagnosing fetal lip and palate during the second trimester. *Quant Imaging Med Surg.* 2025;15:7497-509. DOI PubMed PMC
6. Madani Z, Fakhar HB. Detection and classification of peri-implant marginal bone loss in cone-beam computed tomography using a deep learning approach. *Clin Exp Dent Res.* 2026;12:e70308. DOI PubMed PMC
7. Jiao X, Gao S, Huang F, Dou W, Zhou Y, Zhang C. An improved algorithm for full-mouth lesion detection based on YOLOv8. *Graph Models.* 2025;142:101302. DOI
8. Hou H, Yang A, Li X, Zhu K, Zhao Y, Wu Z. Dental bur detection system based on asymmetric double convolution and adaptive feature fusion. *Sci Rep.* 2024;14:31874. DOI PubMed PMC
9. Poomrittigul S, Mittong S, Thanathornwong B, Suebnukarn S. Deep learning-based object detection of restorative dental instruments with potential implications for workflow automation and infection control in dental supply units. *Sci Rep.* 2026;16:1168. DOI PubMed PMC
10. Oka S, Nozaki K, Hayashi M. An efficient annotation method for image recognition of dental instruments. *Sci Rep.* 2023;13:169. DOI PubMed PMC
11. Girija R, Gupta A, Anand R, Dhampetla AR, Kumar T, Kathait Y. *Annotated dental instrument dataset for YOLO object detection.* Mendeley Data. 2026.[DOI: 10.17632/mh6zpc7y9g.1].
12. Hartman H, Ruslandi RFA, Nurdin D, Akbar S, Setiawan AS. Exploring YOLO applications in dentistry through a systematic review of detection and segmentation models. *Discov Artif Intell.* 2026;6:292. DOI
13. Khurshid Z, Osathanon T, Alharbi AS, Egusa H. Applications for the YOLO deep learning framework in dentistry: a narrative review. *J Prosthodont Res.* 2026; Epub ahead of print. DOI PubMed
14. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27-30; Las Vegas, NV, USA. IEEE; 2016. pp. 779-88. DOI
15. Tian Y, Xu W, Yang B, et al. Development and evolution of YOLO in object detection: a survey. *Neurocomputing.* 2026;669:132436. DOI
16. Xiong X, Yang M, Zheng Y, Zhao L, Li K, Wang J. Multimodal deep learning-based screening of degenerative temporomandibular joint disease using 2D radiography: a cost-effective and low-radiation approach. *Oral Dis.* 2026;32:2007-15. DOI PubMed
17. Ariji Y, Kusano K, Fukuda M, et al. Two-step deep learning models for detection and identification of the manufacturers and types of dental implants on panoramic radiographs. *Odontology.* 2025;113:788-98. DOI PubMed

18. Zhang JP, Wang ZH, Zhang J, Qiu J. Convolutional neural network-based measurement of crown-implant ratio for implant-supported prostheses. *J Prosthet Dent*. 2025;134:189.e1-9. DOI PubMed
19. Zhao Y, Zhao X, Wang L, et al. Predicting dental implant failure risk from panoramic radiographs using deep learning. *Biomed Signal Process Control.* 2026;111:108498. DOI
20. Wu PY, Chen SL, Mao YC, et al. Automated implant placement pathway from dental panoramic radiographs using deep learning for preliminary clinical assistance. *Diagnostics.* 2025;15:2598. DOI PubMed PMC
21. Akdoğan S, Öziç MÜ, Tassoker M. Development of an AI-Supported clinical tool for assessing mandibular third molar tooth extraction difficulty using panoramic radiographs and YOLO11 sub-models. *Diagnostics.* 2025;15:462. DOI PubMed PMC
22. Rashid MO, Gaghor S. Clinical application of a deep learning system for automatic mandibular alveolar bone quantity assessment and suggested treatment options using CBCT cross-sections. *Medicine (Baltimore).* 2025;104:e43257. DOI PubMed PMC
23. Ujjainkar PA, Raut SA. Enhanced deep learning framework for real-time instrument detection and tracking in laparoscopic surgery using advanced augmentation and tracking techniques. *Surg Endosc.* 2025;39:4973-84. DOI PubMed
24. Nwoye CI, Padoy N. SurgiTrack: fine-grained multi-class multi-tool tracking in surgical videos. *Med Image Anal.* 2025;101:103438. DOI PubMed
25. Kim KH, Koo H, Lee B. Deep learning-based localization and orientation estimation of pedicle screws in spinal fusion surgery. *Korean J Neurotrauma.* 2024;20:90-100. DOI PubMed PMC
26. Yilmaz S, Tasyurek M, Amuk M, Celik M, Canger EM. Developing deep learning methods for classification of teeth in dental panoramic radiography. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2024;138:118-27. DOI PubMed
27. Peker RB, Kurtoglu CO. Evaluation of the performance of a YOLOv10-based deep learning model for tooth detection and numbering on panoramic radiographs of patients in the mixed dentition period. *Diagnostics.* 2025;15:405. DOI PubMed PMC
28. Ayyıldız H, Orhan M, Bilgir E, Çelik Ö, Bayrakdar İŞ. Tooth numbering with polygonal segmentation on periapical radiographs: an artificial intelligence study. *Clin Oral Investig.* 2024;28:610. DOI PubMed
29. Sakai K, Muramatsu C, Seino Y, et al. Detection and dual-label classification of tooth number and condition in dental panoramic radiographs including deciduous teeth. *Radiol Phys Technol.* 2025;18:821-32. DOI PubMed
30. Nishimura T, Mine Y, Iwamoto Y, et al. Domain shift analysis of deep learning models for tooth detection in pediatric panoramic radiographs. *Dent Mater J.* 2026;45:354-64. DOI PubMed
31. Lee HS, Yang S, Han JY, et al. Automatic detection and classification of nasopalatine duct cyst and periapical cyst on panoramic radiographs using deep convolutional neural networks. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2024;138:184-95. DOI PubMed
32. Warin K, Limprasert W, Suebnukarn S, Inglam S, Jantana P, Vicharueang S. Assessment of deep convolutional neural network models for mandibular fracture detection in panoramic radiographs. *Int J Oral Maxillofac Surg.* 2022;51:1488-94. DOI PubMed
33. Tassoker M, Öziç MÜ, Yuce F. Performance evaluation of a deep learning model for automatic detection and localization of idiopathic osteosclerosis on dental panoramic radiographs. *Sci Rep.* 2024;14:4437. DOI PubMed PMC
34. 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;49:20200185. DOI PubMed PMC
35. Tosun ZT, Kumbasar N, Sumbullu MA, Miloglu O. Evaluation of the effectiveness of artificial intelligence models in radiopaque and radiolucent lesions of the maxillofacial region on panoramic radiographs. *Oral Radiol.* 2026;42:36-50. DOI PubMed PMC
36. Altun O, Özen DÇ, Duman ŞB, et al. Automatic maxillary sinus segmentation and pathology classification on cone-beam computed tomographic images using deep learning. *BMC Oral Health.* 2024;24:1208. DOI PubMed PMC
37. Yari A, Fasih P, Kamali Hakim L, Asadi A. Automated detection of zygomatic fractures on spiral computed tomography using a deep learning model. *Int J Oral Maxillofac Surg.* 2026;55:68-76. DOI PubMed
38. Ren J, Zhou W, Zhao H, et al. Prediction of lymph node metastasis and recurrence risk in early-stage oral tongue squamous cell carcinoma with fully automated MRI deep learning. *Cancer Imaging.* 2026;26:19. DOI PubMed PMC
39. Hsu Y, Chou CY, Huang YC, et al. Oral mucosal lesions triage via YOLOv7 models. *J Formos Med Assoc.* 2025;124:621-7. DOI PubMed
40. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P. Performance of deep convolutional neural network for classification and detection of oral potentially malignant disorders in photographic images. *Int J Oral Maxillofac Surg.* 2022;51:699-704. DOI PubMed
41. Yu J, Huang R, Wang X, et al. A comparative study of single-stage and dual-stage classification models for OPMDs. *BMC Oral Health.* 2025;26:169. DOI PubMed PMC
42. Khovidhunkit SP, Phosri K, Thanathornwong B, Rungraungrayabkul D, Poomrittigul S, Treebupachatsakul T. Performance of deep learning models for the classification and object detection of different oral white lesions using photographic images. *Sci Rep.* 2025;15:30834. DOI PubMed PMC
43. Schwärzler J, Tolstaya E, Tichy A, et al. Machine learning versus clinicians for detection and classification of oral mucosal lesions. *J Dent.* 2025;161:105992. DOI PubMed

44. Zhao C, Yuan Z, Luo S, et al. Automatic recognition of cephalometric landmarks via multi-scale sampling strategy. *Heliyon.* 2023;9:e17459. DOI PubMed PMC
45. Kotula J, Konarzewski M, Polkowski J, et al. Automated YOLO-based cephalometric landmark detection for ANB-based skeletal classification: a retrospective single-centre study. *J Clin Med.* 2026;15:5149. DOI PubMed PMC
46. Zhu T, Chen D, Wu F, Zhu F, Zhu H. Artificial intelligence model to detect real contact relationship between mandibular third molars and inferior alveolar nerve based on panoramic radiographs. *Diagnostics.* 2021;11:1664. DOI PubMed PMC
47. Jing Q, Dai X, Wang Z, et al. Fully automated deep learning model for detecting proximity of mandibular third molar root to inferior alveolar canal using panoramic radiographs. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2024;137:671-8. DOI PubMed
48. Takebe K, Imai T, Kubota S, Nishimoto A, Amekawa S, Uzawa N. Deep learning model for the automated evaluation of contact between the lower third molar and inferior alveolar nerve on panoramic radiography. *J Dent Sci.* 2023;18:991-6. DOI PubMed PMC
49. Roongruangsilp P, Narkbuakaew W, Khongkhunthian P. Performance of two different artificial intelligence models in dental implant planning among four different implant planning software: a comparative study. *BMC Oral Health.* 2025;25:984. DOI PubMed PMC
50. Rashid M, Gaghor S, Talabani R. Clinically applicable deep learning model for segmentation of the mandibular bone and inferior alveolar canal in CBCT cross-sectional images. *BMC Oral Health.* 2026;26:511. DOI PubMed PMC
51. Tucan P, Hedesiu M, Bublucan V, et al. AI-based anatomical landmark recognition for robot-assisted dynamic dental implant surgery. In: 2025 IEEE 21st International Conference on Intelligent Computer Communication and Processing (ICCP); 2025 Oct 16-18; Cluj-Napoca, Romania. IEEE; 2025. pp. 1-8. DOI
52. Bulbucan V, Pisla D, Tucan P, et al. ; TEAM Project Group. Real-time markerless tooth detection towards dynamic robot-assisted dental implant navigation. *Dent J.* 2026;14:345. DOI PubMed PMC
53. Deng T, Gulati S, Rodriguez W, Dawant BM, Langerman A. Automated detection of electrocautery instrument in videos of open neck procedures using YOLOv3. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2021 Nov 1-5; Mexico. IEEE; 2021. pp. 2071-4. DOI PubMed
54. Olei S, Sarwin G, Staartjes VE, et al. The AENEAS Project: intraoperative anatomical guidance through real-time landmark detection using machine vision. *Mayo Clin Proc Digit Health.* 2026;4:100308. DOI PubMed PMC
55. Chellamani GK, N A, P BK, Reddy PSK, Babu RTS. A robust dataset for surgical instrument detection to aid autonomous robotic surgery. *Data Brief.* 2025;61:111798. DOI PubMed PMC
56. Zhang L, Guo G, Wang W. UK-YOLOv10: deep learning-based detection of surgical instruments. *Int J Med Robot.* 2025;21:e70083. DOI PubMed
57. Pan X, Bi M, Wang H, Ma C, He X. DBH-YOLO: a surgical instrument detection method based on feature separation in laparoscopic surgery. *Int J Comput Assist Radiol Surg.* 2024;19:2215-25. DOI PubMed
58. Liu Y, Hayashi Y, Oda M, Mori K. Enhancing YOLO for laparoscopic tool detection: novel data augmentation and structural modifications addressing mis-detection of bifurcated targets. *Int J Comput Assist Radiol Surg.* 2025;20:1899-910. DOI PubMed
59. Hamdi H, Ye C, Ahmad S, Cao J, Ma T. Occlusion-resilient instance segmentation of surgical instrument parts using YOLO and generative adversarial networks for minimal invasive robotic surgery. In: 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2025 Jul 14-18; Copenhagen, Denmark. IEEE; 2025. pp. 1-4. DOI PubMed
60. Ma Z, Hua Y, Que L, et al. Intelligent surgical workflow recognition-based skill assessment for minimally invasive submandibular gland resection. *NPJ Digit Med.* 2026;9:824. DOI PubMed PMC
61. Ma Z, Guo Z, Ding Z, et al. Evaluation of a newly developed oral and maxillofacial surgical robotic platform (KD-SR-01) in head and neck surgery: a preclinical trial in porcine models. *Int J Oral Sci.* 2024;16:51. DOI PubMed PMC
62. Ma Z, Guo Z, Cao C, et al. Evaluation of a new developed robotic system for head and neck surgery: a prospective study. *Gland Surg.* 2025;14:1250-62. DOI PubMed PMC
63. Tekis H, Zirek T, Tassoker M. Comparative analysis of YOLO variants for dental implant brand identification. *Sci Rep.* 2025;15:33304. DOI PubMed PMC
64. Ahmed WM, Azhari AA, Almufti A, et al. Development and evaluation of an AI model for dental implant type detection: a comparison of diagnostic accuracy between a deep learning model and dental professionals. *J Prosthodont.* 2026;35:81-93. DOI PubMed PMC
65. Hassan NA, Kamel AE, Omran AE, et al. Automated identification of dental implants: a new, fast and accurate artificial intelligence system. *Eur J Prosthodont Restor Dent.* 2024;32:162-7. DOI PubMed
66. Lee WF, Day MY, Fang CY, et al. Establishing a novel deep learning model for detecting peri-implantitis. *J Dent Sci.* 2024;19:1165-73. DOI PubMed PMC
67. Lee JH, Kim YT, Schwendicke F. Performance of artificial intelligence-based diagnosis and classification of peri-implantitis compared with periodontal surgeon assessment: a pilot study of panoramic radiograph analysis. *J Periodontal Implant Sci.* 2025;55:436-46. DOI PubMed PMC

68. Tekis H, Zirek T, Tassoker M. AI-powered detection of dental anatomy: a YOLO-based approach. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2026;141:558-65. DOI PubMed
69. Şevik U, Mutlu O. Detection of dental anomalies in digital panoramic images using YOLO: a next generation approach based on single stage detection models. *Diagnostics.* 2025;15:1961. DOI PubMed PMC
70. Mutlu O, Aslan E, Mert A. Deep learning-based automated diagnostic charting on panoramic radiography: comparison of YOLOv11 and YOLOv12. *Odontology.* 2026;114:2658-68. DOI PubMed PMC
71. Edik M, Çelebi F, Çukurluoğlu A. Deep-learning-based detection of open-apex teeth on panoramic radiographs using YOLO models. *Oral Radiol.* 2026;42:520-30. DOI PubMed
72. Xie Z, Hu X, Guo L, et al. A lightweight detection algorithm for tooth cracks in optical images. *Comput Biol Med.* 2024;182:109153. DOI PubMed
73. Li W, Wang Y, Liu Y. DMAF-Net: deformable multi-scale adaptive fusion network for dental structure detection with panoramic radiographs. *Dentomaxillofac Radiol.* 2024;53:296-307. DOI PubMed PMC
74. Tareq A, Faisal MI, Islam MS, et al. Visual diagnostics of dental caries through deep learning of non-standardised photographs using a hybrid YOLO ensemble and transfer learning model. *Int J Environ Res Public Health.* 2023;20:5351. DOI PubMed PMC
75. Tejani AS, Klontzas ME, Gatti AA, et al. ; CLAIM 2024 Update Panel. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. *Radiol Artif Intell.* 2024;6:e240300. DOI PubMed PMC
76. Sounderajah V, Guni A, Liu X, et al. ; STARD-AI Consensus Group. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat Med.* 2025;31:3283-9. DOI PubMed
77. Vasey B, Nagendran M, Campbell B, et al. ; DECIDE-AI Expert Group. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med.* 2022;28:924-33. DOI PubMed


**Disclaimer/Publisher's Note:** All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.

© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.