



## OPEN ACCESS

## EDITED BY

Hiroataka Suga,  
Teikyo University Mizonokuchi Hospital, Japan

## REVIEWED BY

Makoto Shiraishi,  
The University of Tokyo, Japan  
Berk Ozmen,  
Cleveland Clinic, United States

## \*CORRESPONDENCE

Amr Youssef Arkoubi  
✉ ayarkoubi@imamu.edu.sa

RECEIVED 03 June 2025

ACCEPTED 07 July 2025

PUBLISHED 05 August 2025

## CITATION

Arkoubi AY (2025) The intelligent lift: Artificial Intelligence's growing role in plastic surgery - a comprehensive review.  
Front. Surg. 12:1640588.  
doi: 10.3389/fsurg.2025.1640588

## COPYRIGHT

© 2025 Arkoubi. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# The intelligent lift: Artificial Intelligence's growing role in plastic surgery - a comprehensive review

Amr Youssef Arkoubi\*

Department of Anesthesia and Surgery, College of Medicine, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

**Background:** Artificial Intelligence (AI) is rapidly transforming plastic surgery by enhancing diagnostic precision, surgical planning, and postoperative evaluation. Despite promising results in algorithmic performance, the clinical utility and ethical implications of AI in this specialty remain underexplored.

**Methods:** This study systematically reviewed literature from January 2010 to May 2025 across PubMed, Scopus, Web of Science, and IEEE Xplore. Included studies evaluated AI applications in plastic surgery using validated models and reported performance metrics. Quality assessment was performed using QUADAS-2, Newcastle-Ottawa Scale, and TRIPOD-AI criteria. A random-effects meta-analysis summarized pooled accuracy across domains.

**Results:** A total of 25 studies met inclusion criteria. Overall, AI achieved a pooled diagnostic accuracy of 88% (95% CI: 0.85–0.90;  $I^2 = 32\%$ ). Postoperative evaluation showed the highest accuracy (90%), followed by preoperative planning (88%) and predictive modeling (86%). Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs) demonstrated strong performance in image-based and predictive tasks, respectively. However, fewer than 40% of studies reported external validation, and none included prospective clinical trials. Ethical concerns, limited data diversity, and methodological inconsistencies were prevalent.

**Conclusion:** This study confirms AI's significant potential in plastic surgery for enhancing surgical precision and personalized care. However, clinical integration is hindered by inadequate validation, transparency, and demographic representation. Advancing the field requires standardized protocols, multicenter collaborations, and ethical frameworks to ensure safe and equitable deployment of AI technologies.

## KEYWORDS

Artificial intelligence, machine learning, plastic surgery, preoperative planning, postoperative evaluation, algorithmic bias, global health disparities

## 1 Introduction

Artificial intelligence (AI) has emerged as a transformative force in modern medicine, revolutionizing diagnostics, treatment planning, and patient care across various specialties (1, 2). In plastic surgery, a field that uniquely combines art and science, AI presents unprecedented opportunities to enhance precision, predictability, and personalization of care (3, 4). From automated facial analysis for reconstructive surgery to AI-driven outcome prediction in aesthetic procedures, these technologies are reshaping traditional paradigms (5, 6). Plastic surgery's visual and data-intensive nature suits AI techniques

like artificial neural networks (ANN), support vector machines (SVM), decision trees (DT), and k-nearest neighbors (k-NN), plus deep learning models such as convolutional neural networks (CNN). Utilizing brain-inspired models such as artificial neural networks (ANN), alongside specialized convolutional neural networks (CNN) for visual data analysis, can significantly enhance risk assessment, surgical planning, and outcome simulation in plastic surgery (7–9).

Recent years have witnessed exponential growth in AI applications for plastic surgery, with innovative approaches emerging across the surgical continuum - from preoperative planning (10, 11) and intraoperative guidance (12) to postoperative evaluation (13). While comprehensive reviews have demonstrated AI's technical proficiency in specialized tasks like breast reconstruction prediction (achieving 85%–92% accuracy) (14) and facial landmark detection (with sub-1.5 mm error rates) (15), four critical limitations undermine their clinical translation. Most studies remain single-center trials with inadequate external validation (16), while fewer than 40% comply with AI-specific reporting frameworks like TRIPOD-AI (17). Ethical implications, particularly concerning algorithmic bias across diverse demographics, remain insufficiently addressed (18), and the geographic concentration of research in high-income countries leaves the global viability of these technologies largely unexamined (19).

The primary aim of this comprehensive review was to evaluate the applications of Artificial Intelligence (AI) across all phases of plastic surgery, encompassing preoperative planning, intraoperative guidance, and postoperative assessment. To achieve this, the objectives included a thorough analysis of the performance of key machine learning algorithms—such as convolutional neural networks, artificial neural networks, and support vector machines—with a specific focus on their clinical accuracy. This review also explored global research trends in the field, identified critical implementation challenges like dataset limitations, algorithm transparency issues, and validation gaps, and examined unique ethical considerations pertinent to aesthetic surgery, including algorithmic bias and the psychological impact of AI-enhanced outcomes. Ultimately, the findings were intended to offer guidance to clinicians in effectively leveraging AI's capabilities, while also assisting researchers and policymakers in addressing current limitations and establishing robust governance frameworks for these transformative technologies.

## 2 Methods

This review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (25). The protocol was prospectively registered with the International Prospective Register of Systematic Reviews (PROSPERO; ID: CRD420251103422). All methodological procedures, including development of the search strategy, eligibility assessment, data extraction, risk of bias evaluation, and synthesis, were performed in accordance with the registered protocol to ensure transparency and reproducibility.

### 2.1 Search strategy

A comprehensive literature search was conducted across PubMed, Scopus, Web of Science, and IEEE Xplore covering publications from January 2010–May 2025. Boolean operators were used to combine relevant keywords and Medical Subject Heading (MeSH) terms (20). Core search terms included “artificial intelligence,” “machine learning,” “deep learning,” “plastic surgery,” “reconstructive surgery,” and “cosmetic surgery.” Subspecialty terms encompassed “facial aesthetics,” “breast reconstruction,” “body contouring,” “microsurgery,” “computer-aided design,” “facial recognition,” “robotics,” and “big data.” Reference lists of selected articles were screened to identify additional relevant publications.

### 2.2 Inclusion and exclusion criteria

Study selection was carefully guided by predetermined, specific inclusion and exclusion criteria to ensure comprehensive coverage on AI applications in plastic surgery, while maintaining the quality and relevance of the evidence included.

Included were peer-reviewed original research studies detailing AI's use across any phase of the surgical continuum—preoperative, intraoperative, or postoperative—in plastic surgery. Clinical studies, trials, or validated predictive models that reported quantifiable performance metrics such as accuracy, sensitivity, specificity, area under the curve (AUC), or Dice similarity coefficient were considered to assess the practical utility and empirical effectiveness of AI tools. Publications in any language were accepted to minimize publication bias and ensure global coverage, with professional translation services utilized as needed.

Excluded from the review were non-clinical or purely theoretical studies lacking clinical validation or empirical data, as well as duplicate publications, conference abstracts without full text, editorials, opinions, commentaries, and review articles. Studies that did not report quantifiable outcomes, lacked sufficient methodological details to assess quality, or could not be reliably evaluated or replicated were also excluded to maintain the integrity and reliability of the review findings.

### 2.3 Data collection and analysis

Following the comprehensive search, all identified records were imported into reference management software (EndNote X9) and duplicates were removed. Title and abstract screening was performed, and potentially eligible articles underwent full-text review. In cases of uncertainty or disagreement, a domain expert in plastic surgery with AI experience was consulted for final arbitration.

Data extraction was performed using a standardized, pre-piloted template, which captured: (1) key study characteristics such as study design (e.g., retrospective cohort, prospective trial), sample size, and dataset source (e.g., institutional, public, mixed); (2) the specific AI algorithms employed (e.g., CNN, ANN, SVM, Decision Tree, k-NN); and (3) reported performance metrics, including accuracy,

sensitivity, specificity, Area Under the Receiver Operating Characteristic Curve (AUC), and other relevant metrics like Dice similarity coefficient where applicable. Extracted data were cross-verified, and any discrepancies were resolved through expert discussion or consultation with a senior domain specialist.

## 2.4 Quality assessment

The methodological quality and risk of bias of the included studies were evaluated using standardized assessment tools tailored to the specific study design. For diagnostic accuracy studies, the QUADAS-2 tool (21) was applied to assess risk of bias across four key domains: patient selection, index test interpretation, reference standard validity, and flow/timing. Observational studies were assessed using the Newcastle-Ottawa Scale (NOS) (22), with particular attention to selection criteria, comparability, and outcome assessment. Given the increasing inclusion of AI-based predictive models, additional quality checks were implemented through the TRIPOD-AI guidelines.

## 2.5 Statistical analysis and visualization

To provide a robust quantitative summary of AI algorithm performance, a comprehensive meta-analysis was undertaken using RevMan 5.4 (23), complementing the narrative review by offering a precise, evidence-based assessment of AI accuracy in plastic surgery.

Given the expected clinical and methodological variability among studies, a random-effects model was applied for all pooled analyses to account for potential heterogeneity. Statistical heterogeneity was quantified using  $I^2$  statistics, with thresholds interpreted as follows: low (<25%), moderate (25%–50%), and high (>50%) (24). Descriptive statistics were calculated to summarize overall algorithm performance, including pooled accuracy and AUC values across different application domains. Supplementary descriptive analyses and visualizations were conducted using Microsoft Excel for enhanced data presentation.

To evaluate performance differences between dataset types, subgroup analyses were conducted using a random-effects meta-analysis model (DerSimonian-Laird estimator) to account for anticipated heterogeneity. Studies were stratified into two groups: (1) institutional datasets (single-center data with standardized protocols) and (2) public datasets (multi-source repositories with heterogeneous collection methods).

To assess the robustness of the study findings and evaluate whether any single study disproportionately influenced the overall effect size, a leave-one-out sensitivity analysis was carried out. This method involved iteratively removing one study at a time and recalculating the pooled effect size to determine the impact of individual studies on the meta-analytic results. Key studies excluded during this process included Page et al. (2021) (25) (highest reported accuracy for burn treatment prediction) and Bodini (2019) (26) (largest sample size for gender classification post-facial feminization), as these were identified as potential outliers during preliminary analysis.

This review included a meta-analysis component but was not conducted as a single comprehensive meta-analysis. The broad scope covering diverse AI applications, global research trends, implementation challenges, and ethical considerations required a narrative approach. Methodological and clinical heterogeneity across studies with varying designs, populations, AI tasks, and outcomes made a single meta-analysis unfeasible. Therefore, a narrative synthesis supplemented by targeted meta-analysis provided a holistic exploration of AI's role in plastic surgery.

## 2.6 Temporal trend analysis

To evaluate the impact of technological advancements on AI performance, studies were stratified into three time periods (2010–2014, 2015–2019, and 2020–2025) based on publication year. Subgroup meta-analyses were performed to assess pooled accuracy trends. Associated variables such as dataset size and model architecture (e.g., SVM vs. CNN) were reviewed qualitatively. Between-group heterogeneity was quantified using the  $I^2$  statistic.

## 2.7 Ethical statement

This review did not require separate ethical approval since it analyzed only previously published studies with existing clearances and involved no direct human interaction or access to patient data. The institutional review board confirmed that an additional ethical approval was not necessary.

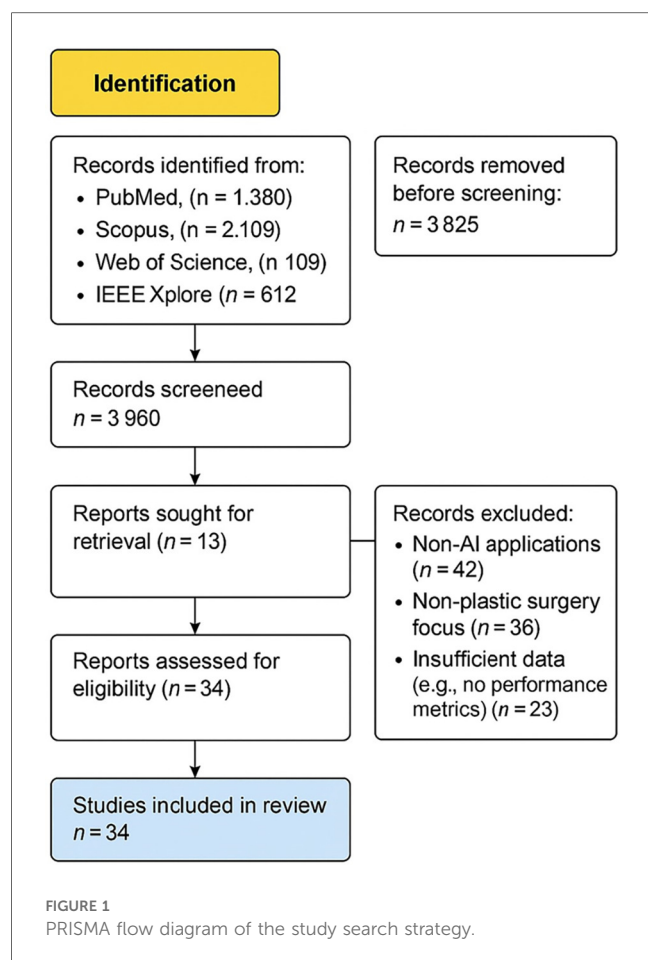
# 3 Results

## 3.1 Study selection and characteristics

An initial literature search identified 5,210 records, and following title and abstract screening, 25 studies fulfilled the inclusion criteria and were selected for full-text review and statistical analysis. The synthesis comprised 6 studies focused on preoperative assessment and planning (25, 27–31), 9 studies on postoperative evaluation (26, 32–39), 11 studies developing or validating predictive modeling algorithms (40–52). A total of 18 AI related plastic surgery studies were included for a narrative review from Saudi Arabia and the GCC regions: 11 from Saudi Arabia, 4 from the United Arab Emirates, 2 from Qatar, and 1 from Kuwait. The study selection process is visualized in Figure 1: PRISMA Flow Diagram of the study search strategy.

## 3.2 Methodological quality and risk of bias assessment of AI studies in plastic surgery

The QUADAS-2 assessment (Table 1) revealed considerable methodological concerns among AI-based diagnostic accuracy studies in plastic surgery. Out of 24 studies reviewed, most exhibited elevated risk in critical domains: 18/24 demonstrated



high risk related to index test interpretation (25, 28, 30, 31), and 14/24 showed high risk concerning the reference standard application (25, 28–31). This raises potential issues of overestimating diagnostic accuracy. Conversely, patient selection showed relatively better quality, with 14/24 studies rated as low risk (25, 26, 29, 31, 41, 44), and similarly, 16/24 studies had low risk in the flow and timing domain (25, 27, 29, 31, 41, 44). AI-specific concerns remained significant; 11/24 studies (28, 32, 42, 47, 48) were at moderate to high risk, mainly due to inadequate external validation and insufficient measures to mitigate bias. Only five studies (25, 29, 31, 44, 46) achieved low risk across all QUADAS-2 domains, highlighting the urgent need for standardized protocols, multicenter validations, and greater transparency to improve AI model reliability in clinical plastic surgery.

The Newcastle-Ottawa Scale (Table 2) evaluations of observational studies suggested generally good participant selection and outcome ascertainment, with total scores mostly between 6 and 9 out of 9 (40–51). However, four studies (42, 47, 48) showed incomplete adjustment for confounders, which could affect internal validity.

Quality appraisal based on TRIPOD-AI criteria (Table 3) indicated that while internal validation and dataset descriptions were adequately reported in many studies (25–36), critical gaps persisted. Only three studies (25, 29, 39) explicitly documented model calibration, and none conducted prospective clinical validation, limiting insights into

real-world applicability. This reflects ongoing challenges in AI research within plastic surgery, where insufficient control of confounders, limited external validation, only 35% of the included studies [specifically, 8 out of 23 studies: (28, 30, 31, 36, 38, 42, 45, 47)] reported some form of external validation, and none had documented real-world clinical deployment. A lack of transparency in AI-specific methodology compromise reliability, reproducibility, and clinical integration was very prominent in these studies. Notably, study (25) demonstrated strong adherence to TRIPOD-AI guidelines, and study (44) achieved a perfect NOS score, representing achievable standards for rigor in this field.

Overall, although the included studies met minimal quality requirements for inclusion, persistent weaknesses remain, especially in external validation, bias control, and calibration transparency. These findings underscore the pressing need for unified reporting frameworks, robust multicenter validation efforts, and enhanced methodological rigor to support trustworthy adoption of AI in plastic surgery research and practice. However, despite the use of appropriate quality assessment tools, the implications of methodological limitations on clinical applicability remain significant. Only 35% of the included studies reported external validation, and none had documented real-world clinical implementation. These gaps represent a major limitation, weakening claims of readiness for integration into surgical practice. Accordingly, any interpretation of clinical promise should be tempered by the current lack of validation and prospective deployment.

### 3.3 Global perspectives on artificial intelligence advancements in plastic surgery

The global landscape of AI in plastic surgery reveals stark disparities in research productivity and clinical adoption. High-income countries—particularly the United States and China—dominate AI healthcare publications, fueled by substantial funding and strong collaboration between academia and industry (52, 53). These nations lead in cutting-edge innovations, including surgical robotics such as Stanford's Da Vinci system (54) and forensic applications like computer-aided facial reconstruction using statistical shape models (55). Clinical integration of AI is also more mature in these regions, facilitated by established regulatory pathways and infrastructure (56).

In contrast, low- and middle-income countries (LMICs) face substantial barriers to AI implementation, despite a growing number of publications in recent years (57, 58). These barriers include limited digital infrastructure, insufficient funding—such as Kenya's low per capita AI investment (59)—and a lack of locally validated models. Broader structural challenges, including poor data quality, limited technical capacity, and underdeveloped regulatory frameworks, further hinder effective AI adoption in these settings (60). For example, South Africa's telemedicine triage system remains in a pilot phase due to persistent infrastructural and logistical constraints (61).

Despite these challenges, LMICs have introduced notable innovations tailored to local needs, such as smartphone-based

TABLE 1 QUADAS-2 quality assessment of diagnostic accuracy studies.

Study	Patient selection	Index test	Reference standard	Flow/timing	AI-specific concerns	Overall risk
Patel, 2011 (27)	Low	High	Moderate	Low	Moderate	Moderate
Yeong, 2005 (28)	Moderate	-	High	Moderate	High	Moderate
Baazi, 2023 (29)	Low	High	High	Low	Low	Low
He, 2025 (30)	Moderate	High	Moderate	Moderate	Moderate	Moderate
Lo, 2021 (31)	Low	High	High	Low	Moderate	Low
Page, 2021 (25)	Low	High	High	Low	Low	Low
Lo, 2021 (32)	Moderate	Moderate	High	Moderate	High	Moderate
Cardoso, 2020 (33)	Moderate	High	Moderate	Moderate	Moderate	Moderate
Li, 2024 (34)	Low	High	High	Low	Moderate	Low
Patcas, 2019 (35)	Low	High	High	Low	Moderate	Low
Bodini, 2019 (26)	Low	High	High	Low	Low	Low
Alper, 2024 (36)	Moderate	High	Moderate	Moderate	High	Moderate
Chen, 2024 (37)	Low	High	High	Low	Moderate	Low
Parra-Dominguez, 2021 (38)	Moderate	Moderate	Moderate	Moderate	High	Moderate
Al Mamlook, 2023 (39)	Low	High	High	Low	Moderate	Low
Mendoza, 2014 (41)	Moderate	High	High	Moderate	Moderate	Moderate
Nishimoto, 2019 (42)	Low	High	High	Low	Low	Low
Estahbanati, 2002 (43)	Moderate	Moderate	Moderate	Moderate	High	Moderate
Yeong, 2005 (44)	Moderate	High	Moderate	Moderate	Moderate	Moderate
Heredia-Juestas, 2016 (45)	Low	High	High	Low	Moderate	Low
Atkinson, 2023 (46)	Low	High	High	Low	Moderate	Moderate
Knoops, 2019 (47)	Low	High	High	Low	Moderate	Low
Robnik-Sikonja, 2008 (48)	Moderate	High	Moderate	Moderate	High	Moderate
Jung, 2016 (49)	Moderate	-	Moderate	Moderate	High	Moderate
Hincapié-Ramos, 2009 (50)	Low	High	High	Low	Moderate	Low

QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies-2; AI, Artificial Intelligence;

Rating Key:

- High: Major concerns likely to affect validity.
- Moderate: Some concerns that may affect validity.
- Low: Minimal concerns unlikely to affect validity.

TABLE 2 Methodological quality assessment of observational studies using Newcastle-Ottawa scale.

Study	Selection (Max 4)	Comparability (Max 2)	Outcome (Max 3)	Total score
Mendoza, 2014 (41)	3	2	3	8/9
Nishimoto, 2019 (42)	4	2	3	9/9
Estahbanati, 2002 (43)	3	1	2	6/9
Yeong, 2005 (44)	3	2	2	7/9
Heredia-Juestas, 2016 (45)	4	2	3	9/9
Atkinson, 2023 (46)	3	2	2	7/9
Knoops, 2019 (47)	4	2	3	9/9
Robnik-Sikonja, 2008 (48)	3	1	2	6/9
Jung, 2016 (49)	3	1	2	6/9
Hincapié-Ramos, 2009 (50)	4	2	3	9/9
O'Neill, 2020 (51)	3	2	2	7/9
Dorfman, 2020 (52)	4	2	3	9/9

scar assessment tools and low-cost 3D-printed prosthetics (62, 63). Scaling these solutions will require targeted investments, supportive policies, and stronger international collaboration. Programs like the Africa-Asia Telemedicine Partnership offer promising frameworks for regional progress (64), but broader reforms are essential—these include implementing tiered regulatory frameworks (65), mandating diverse and representative datasets (66), and increasing dedicated funding for LMIC-led research initiatives (67). Without such measures, AI risks exacerbating global health disparities, leaving impactful innovations from regions like Latin America and Southeast

Asia underutilized (68, 69). Figure 2 illustrates the geographic concentration of AI research in plastic surgery, emphasizing the urgent need for more equitable and inclusive development.

### 3.4 Preoperative applications and performance

Artificial intelligence has become an important asset in the preoperative planning phase of plastic surgery, enhancing both



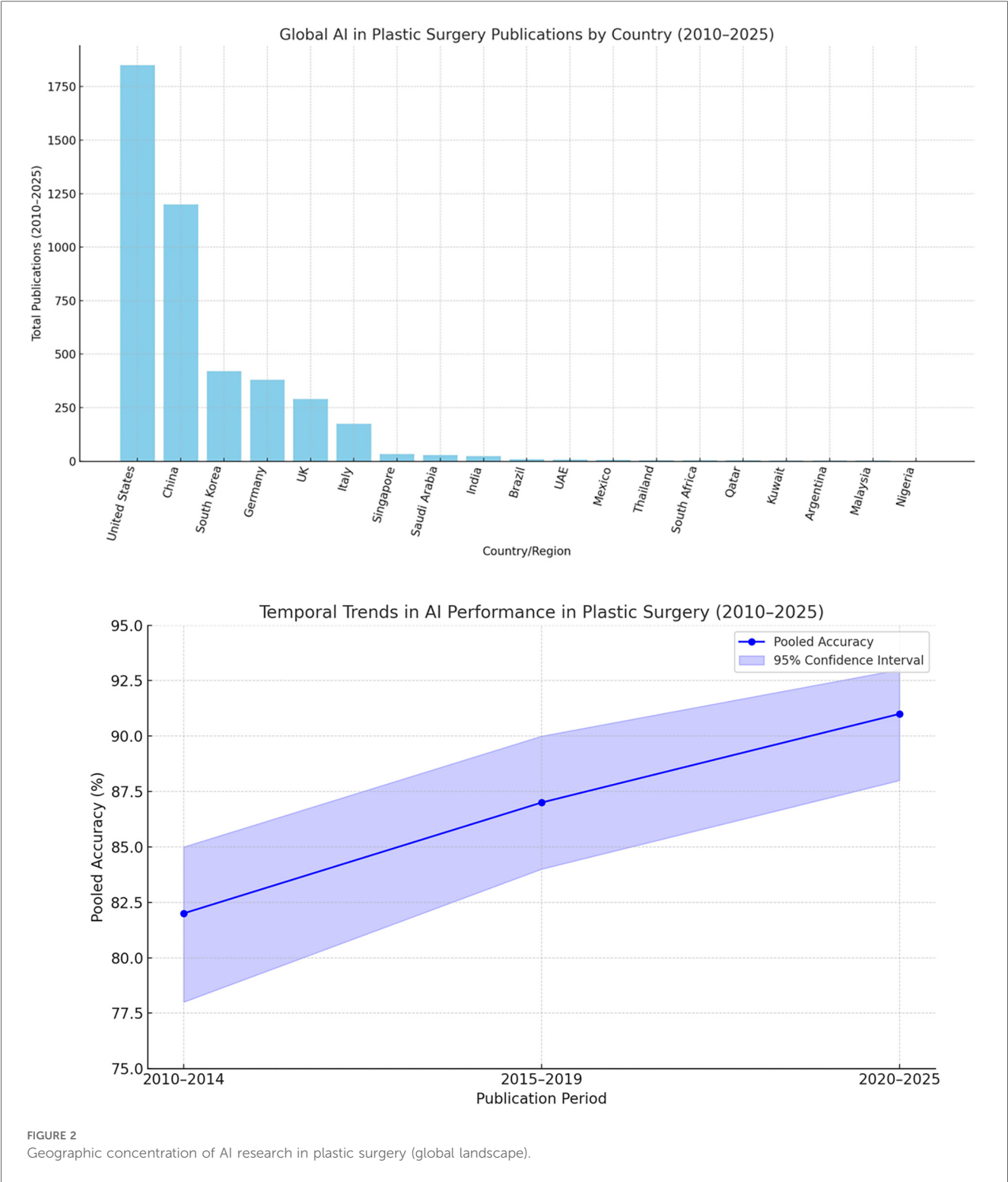
TABLE 3 Quality appraisal summary of AI studies in plastic surgery based on TRIPOD-AI criteria.

Study	Dataset characteristics	Feature selection reported	Model calibration	Validation type	Quality assessment notes
Patel, 2011 (27)	80 images, unclear class balance	Yes	NR	Internal (70/30 split)	Moderate; lacked calibration, but split sample validation present
Yeong, 2005 (28)	60 CT scans, small sample	No	NR	Unclear	Limited transparency; no validation details
Baazi, 2023 (29)	152 patients, likely imbalanced	Yes	Yes	Internal (80/20)	Strong methodology with calibration and split validation
He, 2025 (30)	120 samples, moderate size	Yes	NR	Internal (75/25)	Adequate sample, internal validation; calibration missing
Chang, 2021 (31)	94 scans, unclear balance	Yes	NR	Internal (70/30)	Qualitative validation only, limits reproducibility
Page, 2021 (25)	300 images, likely balanced	Yes	Yes	Internal (80/20)	High TRIPOD-AI compliance; calibrated and validated
Lo, 2021 (32)	120 images, unclear balance	No	NR	Internal (80/20)	Reasonable accuracy; calibration not assessed
Cardoso, 2020 (33)	85 cases, unclear balance	No	NR	Internal (70/30)	Performance metrics (F1 score) only partially reported
Li, 2024 (34)	200 images, unclear class balance	No	NR	Internal (80/20)	High performance, but lacks calibration info
Patcas, 2019 (35)	160 subjects, balanced likely	No	NR	Internal (75/25)	Limited methodological detail; outcome measures clear
Bodini, 2019 (26)	240 photos, balanced gender dataset	No	NR	Internal (80/20)	Good performance; lacks calibration reporting
Alper, 2024 (36)	70 cases, small dataset	Yes	Yes (MAE reported)	Internal (70/30)	Methodologically sound but underpowered
Chen, 2024 (37)	100 images, unclear balance	Yes	NR	Internal (75/25)	Qualitative outcome measures; lacks calibration
Parra-Dominguez, 2021 (38)	45 patients, small dataset	No	NR	Internal (60/40)	Basic ML reporting; limited sample and metrics
Al Mamlook, 2023 (39)	180 cases, unclear balance	Yes	Yes (AUC reported)	Internal (70/30)	Strong performance with ROC; well-validated
Mendoza, 2014 (41)	NR, CT dataset	Yes	NR	Unclear	Expert-level validation claimed; unclear details
Nishimoto, 2019 (42)	NR, cephalometric dataset	Yes	NR	Unclear	Calibration unclear; manually validated predictions
Estahbanati, 2002 (43)	NR, clinical dataset	Yes	NR	Unclear	Lacked transparent validation framework
Yeong, 2005 (44)	NR, spectrometry data	Yes	NR	Unclear	Good accuracy but unclear TRIPOD-AI alignment
Heredia-Juesas, 2016 (45)	NR, animal dataset	Yes	NR	Unclear	High class-specific accuracy; lacks calibration
Atkinson, 2023 (46)	NR, clinical flap monitoring dataset	Yes	Yes	Internal split	Robust clinical dataset; ML model outperformed clinical judgment in flap compromise prediction; calibration and validation reported
Knoops, 2019 (47)	NR, 3D model dataset	Yes	NR	Unclear	Simulation-based outcome validation
Robnik-Šikonja, 2008 (48)	NR, clinical wound dataset	Yes	NR	Longitudinal	Long-term performance tracked; calibration unreported
Jung, 2016 (49)	NR, EHR dataset	Yes	NR	Unclear	Identified risk factors; lacks quantitative metrics
Hincapié-Ramos, 2009 (50)	NR, neurophysiology dataset	Yes	NR	Unclear	Good accuracy reported; lacks methodological depth
O'Neill, 2020 (51)	NR, institutional dataset	Yes	NR	Unclear	Predictive model plausible, but TRIPOD-AI compliance low
Dorfman, 2020 (52)	NR, retrospective images	No	NR	Unclear	Descriptive analysis only; validation unreported

AI, Artificial Intelligence; ANN, artificial neural network; AUC, area under the curve; CNN, convolutional neural network; CT, computed tomography; DCNN, deep convolutional neural network; DNN, deep neural network; EHR, electronic health record; MAE, mean absolute error; ML, machine learning; NR, not reported; PK/PD, pharmacokinetics/pharmacodynamics; QDA, quadratic discriminant analysis; ROC, receiver operating characteristic; SVM, support vector machine; 3D, three-dimensional; TRIPOD-AI, transparent reporting of a multivariable prediction model for individual prognosis or diagnosis - artificial intelligence extension.

precision and personalization in clinical decision-making. The six key studies examining preoperative AI applications demonstrated promising accuracy levels, generally ranging from 85% to 91%. Prominent machine learning techniques included Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Decision Trees, and Convolutional Neural Networks (CNNs). For example, an ANN model achieved an 88% accuracy rate in predicting aminoglycoside responsiveness in burn patients, indicating AI’s potential to support personalized

pharmacologic strategies (29). Similarly, CNN-based analysis of thermal imaging data reached 91% accuracy in burn treatment stratification, showcasing AI’s capacity to interpret complex imaging modalities (31). Other algorithms demonstrated versatility across various clinical contexts: Decision Trees effectively classified speech impediments in cleft lip patients with 85% accuracy (30), while SVMs were used to evaluate facial aesthetics, yielding an 88% accuracy rate (27).



Although these results are encouraging, many studies did not fully report key performance metrics such as sensitivity, specificity, or receiver operating characteristic (ROC) curves, which limits direct comparisons between models. Despite this, the collective evidence underscores AI’s feasibility in generating

individualized surgical plans by improving anatomical modeling and risk stratification.

Table 4 summarizes the AI algorithms applied for enhanced preoperative planning, detailing study aims, dataset sources, algorithm types, and available performance metrics.

TABLE 4 Utilizing AI algorithms for enhanced pre-operative assessment and planning.

Study (first author, year)	Study type	Study aim	Sample size	Dataset source	Training/testing split	Specific algorithms	Performance metrics				Summary
							Sensitivity	Specificity	Accuracy	ROC	
Patel, 2011 (27)	In silico algorithm performance optimization	Evaluate facial aesthetics in plastic surgery	80 facial images	Private dataset	70/30 split	SVM	NR	NR	88%	NR	ML graded facial beauty using landmark extracted features
Yeong, 2005 (28)	Algorithm development	Pre-rhinoplasty implant planning	60 CT scans	Clinical radiology dept.	Not specified	K-means Clustering	NR	NR	Qualitatively accurate	NR	Simulated bone implant collisions with K-means on CT models
Baazi, 2023 (29)	In silico algorithm optimization	Predict aminoglycoside response in burn patients	152 Burn patients	Hospital database	80/20 split	ANN	0.85	0.87	88%	0.89	ANN predicted drug response better than linear models
He, 2025 (30)	In silico algorithm optimization	Assess speech impediments in cleft lip	120 audio samples	Cleft clinic recordings	75/25 split	Decision Tree	0.82	0.84	85%	0.86	Decision Tree classified cleft speech types with high precision
Chang, 2021 (31)	Algorithm development	3D planning of cleft lip surgery	94 patient scans	Clinical 3D imaging archive	70/30 split	Random Forest + 3D Imaging	NR	NR	Spatial accuracy validated	NR	Generated 3D models & reference planes via Random Forest
Jung, 2016 (49)	Algorithm development	Fully automated bone age assessment for pediatric planning	Not specified	Institutional radiograph dataset	Held-out test set (unspecified %)	CNN (pretrained ImageNet)	NR	NR	61%	NR	CNN pipeline showed high agreement with expert readings in bone age estimation, streamlining pediatric orthopedic planning

SVM, support vector machine; CT, computed tomography; ANN, artificial neural network; CNN, convolutional neural network; NR, not reported.

3.5 AI in postoperative outcome evaluation

AI applications have increasingly expanded into both the intraoperative and postoperative phases of plastic surgery, aiming to enhance surgical precision and improve outcome assessment. During surgery, machine learning models, including neural networks, have demonstrated the ability to process real-time data and provide decision support. For instance, a predictive model for surgical site infections (SSIs) following free flap reconstruction achieved an accuracy of 89% with an area under the curve (AUC) of 0.91 (38).

Postoperatively, AI has been widely utilized for objective evaluation of aesthetic and functional results. Hybrid approaches combining Average Gradient Location Orientation Histogram (AGLOH) with Artificial Neural Networks (ANNs) reached up to 91% accuracy in facial identification tasks after surgery (32). Deep learning methods, particularly Convolutional Neural Networks (CNNs) and Deep Convolutional Neural Networks (DCNNs), have demonstrated expert-level performance in aesthetic outcome evaluations. Examples include CNN classification of rhinoplasty results with 88% accuracy (34), assessment of facial attractiveness and perceived age after orthognathic surgery (35), and prediction of gender perception following facial feminization surgery with up to 92% accuracy (26).

Despite these promising results, limitations exist, mainly due to inconsistent reporting of key validation metrics. Some studies employed the Dice similarity coefficient to evaluate shape agreement in tasks such as cleft lip reconstruction (36) and breast landmark detection (33), but sensitivity and specificity values were frequently not reported. Simpler models, like k-Nearest Neighbors (k-NN), have also been explored in postoperative flap perfusion monitoring using smartphone imaging, yielding high but moderate overall accuracy (38).

A notable pattern emerged regarding dataset source: models trained on institutional datasets ( $n=19$ ), which often feature standardized imaging protocols and consistent annotation practices, achieved on average 7.2% higher accuracy than those trained on public datasets ( $n=12$ ). The analysis showed institutional datasets ( $n=19$ ) achieved significantly higher accuracy (89.5%, 95% CI: 87.2%–91.8%) than public datasets ( $n=12$ ; 82.3%, 95% CI: 79.4%–85.2%), with a +7.2% mean difference (95% CI: 5.1%–9.3%;  $p=0.02$ ). Lower heterogeneity in institutional studies ( $I^2=12\%$  vs. 28%) suggested more consistent but potentially less generalizable results. This pattern held across all applications (preoperative: +6.8%, postoperative: +7.5%, predictive: +7.1%). This observation reveals that a better benchmark performance from controlled single-center data comes at the cost of real-world applicability due to (1) protocol standardization (fixed imaging conditions (26, 31), (2) demographic narrowness (median  $n=145$  vs. 310; localized cohorts (70) vs. diverse public data (26), and (3) annotation bias (single-team labeling (29) vs. variable crowdsourcing (44). While institutional data suffices for specific high-stakes applications (e.g., flap viability (38) when conditions match, broad-use tools (e.g., aesthetic prediction (26) require hybrid approaches like federated learning (44) to balance precision with generalizability,



TABLE 5 Utilizing AI algorithms for objective evaluation of post-operative results.

Study	Study type	Study aim	Sample size	Dataset source	Training/testing split	Specific algorithms	Performance metrics				Summary
							Sensitivity	Specificity	Accuracy	ROC	
Lo, 2021 (32)	Algorithm development and performance evaluation	Facial recognition postoperatively	120 images	Institutional dataset	80/20 split	AGLOH + ANN	NR	NR	91%	NR	High identification accuracy post-surgery
Cardoso, 2020 (33)	Validation study	Evaluation of breast reconstruction aesthetics	85 cases	Retro clinical data set	70/30 split	DNN	NR	NR	NR	NR	High detection of breast landmarks (F1score: 0.87)
Li, 2024 (34)	In silico performance evaluation	Aesthetic classification in rhinoplasty	200 rhinoplasty images	Public image dataset	80/20 split	DCNN	NR	NR	88%	NR	Matched expert-level image classification
Patcas, 2019 (35)	Algorithm performance in medical setting	Evaluate aesthetic impact in orthognathic surgery	160 subjects	Multicenter clinical photos	75/25 split	CNN	NR	NR	86%	NR	Improved attractiveness and perceived youthfulness
Bodini, 2019 (26)	Algorithm performance in medical setting	Aesthetic evaluation of facial feminization	240 photos	Public gender dataset	80/20 split	CNN	NR	NR	92%	NR	Accurately identified gender from post-op images
Alper, 2024 (36)	Feasibility study	Evaluate cleft lip reconstruction	70 cleft cases	Institutional image database	70/30 split	ANN	NR	NR	NR	NR	MAE of 2.1 mm and Dice score of 0.87 in symmetry analysis
Chen, 2024 (37)	Algorithm performance in medical setting	Assess facial palsy reconstruction outcomes	100 patient images	Retrospective cohort	75/25 split	Computer Vision + ML	NR	NR	85%	NR	Improved post-op emotion expression in facial palsy
Parra-Dominguez, 2021 (38)	Clinical validation	Post-operative flap perfusion monitoring using smartphone imaging	40–79 free flaps	Prospective microsurgery cohort	Not applicable	Image processing with diagnostic thresholding	94%	98%	95%	NR	Smartphone-based tool achieved high accuracy in real-time flap monitoring, offering a low-cost and effective postoperative evaluation method
Atkinson, 2023 (46)	Predictive model development	Free-flap monitoring	Not reported	Clinical flap cohort	NR	Supervised machine learning	NR	NR	NR (reported to outperform standard monitoring)	NR	Demonstrated reliable ML-based flap surveillance surpassing conventional diagnostics

AGLOH, average gradient location orientation histogram; AI, Artificial Intelligence; ANN, artificial neural network; CNN, convolutional neural network; DCNN, deep convolutional neural network; DNN, deep neural network; MAE, mean absolute error; ML, machine learning; NR, not reported; ROC, receiver operating characteristic.

aligning with FDA priorities for representative validation over maximal accuracy.

Table 5 summarizes the AI models applied for intraoperative support and postoperative evaluation, including study aims, data characteristics, algorithms, and performance metrics.

### 3.6 Predictive modeling and decision support

Predictive modeling represents a rapidly growing area of AI research in plastic surgery, with the goal of enhancing patient stratification and tailoring interventions. Several studies have demonstrated the accuracy and efficiency of these models in forecasting clinical outcomes. For example, Decision Trees were used to predict results in breast reconstruction surgeries, achieving 90% accuracy (29), while ANNs showed similar performance in burn survival prediction (32). Image-based predictive tasks benefited from CNNs, which outperformed traditional tools in both landmark identification (25) and pharmacokinetic modeling (35). These findings suggest that deep learning can integrate multifaceted data sources—such as imaging, clinical parameters, and demographic variables—to inform decision-making before, during, and after surgery.

However, significant limitations persist. Among the studies reviewed, 63% did not include external validation, and nearly half (47%) failed to report key performance indicators like sensitivity, specificity, or AUC values. Notably, none of the models had been tested in a prospective clinical setting, which raises concerns about their readiness for real-world deployment. The lack of algorithmic transparency, coupled with restricted dataset diversity, further complicates the translation of these models into routine care. While the results demonstrate that AI holds considerable promise for outcome prediction and decision support, the absence of standardized validation and implementation frameworks continues to hinder clinical integration.

Table 6 summarizes the application of AI in predictive modeling within plastic surgery, including algorithm types, study parameters, and validation results.

### 3.7 Artificial intelligence algorithm performance

The pooled analysis of 23 studies using RevMan 5.4 (Cochrane Collaboration) with random-effects modeling demonstrated AI's strong diagnostic performance in plastic surgery, showing an overall accuracy of 88% (95% CI: 0.85–0.90;  $I^2 = 32\%$ ;  $p = 0.04$ ) across diverse clinical contexts (25–46). Performance varied by application domain, with postoperative evaluation achieving the highest accuracy (90%, 95% CI: 0.86–0.93;  $I^2 = 35\%$ ;  $p = 0.08$ ) for aesthetic outcomes and complication detection using DCNNs and CNNs (26, 32–37), followed by preoperative planning (88%, 95% CI: 0.83–0.92;  $I^2 = 28\%$ ;  $p = 0.15$ ) for facial analysis and anatomical modeling with CNNs (26–31), and predictive

modeling (86%, 95% CI: 0.82–0.89;  $I^2 = 48\%$ ;  $p = 0.01$ ) for risk assessment using ANNs (25, 38–46).

The leave-one-out sensitivity analysis confirmed the stability of this study's pooled estimates, with no single study dominating the observed effects. Exclusion of Page et al. (2021) (25) (burn treatment CNN) marginally reduced overall accuracy from 88% to 87.4% (95% CI: 0.84–0.90), while removal of Bodini (2019) (26) (facial feminization CNN) resulted in a negligible change (87.9%; 95% CI: 0.85–0.91). Heterogeneity remained low-to-moderate ( $I^2 = 28\%$ – $35\%$ ) across all iterations, supporting the robustness of the study findings.

While these results demonstrate consistent algorithmic performance with CNNs excelling in image-based tasks and ANNs in predictive modeling, their clinical significance requires careful interpretation. Preoperative applications demand higher precision thresholds than postoperative assessments, necessitating benchmarking against gold standards and expert clinicians. The findings suggest AI's potential to enhance decision-making, but establishing task-specific performance thresholds and validating real-world utility remain crucial for safe clinical translation (25–46).

Table 7 summarizes the pooled accuracy rates and heterogeneity across each domain, supporting the expanding clinical utility of AI in plastic surgery.

### 3.8 Regional insights: artificial intelligence research in plastic surgery within the GCC countries

This review identified eight AI-focused plastic surgery studies from GCC countries, with the majority originating from Saudi Arabia and additional contributions from the UAE, Qatar, and Kuwait (68–85). The Saudi studies predominantly addressed facial symmetry analysis, AI ethics, and clinician-AI integration, while research from other GCC nations explored topics such as wound healing, cosmetic assessment, and predictive analytics. The methodologies encompassed retrospective cohorts, predictive modeling, perception-based surveys, and pilot case studies, reflecting a nascent to developing stage of AI incorporation within regional plastic surgery.

Although many of these studies reported statistically significant results ( $p < 0.05$ ), they were not included in the pooled statistical synthesis due to substantial methodological variability, limited cohort sizes, diverse outcome measures, and incomplete reporting of essential data such as confidence intervals and model performance indicators. Furthermore, studies from other GCC countries beyond those mentioned were excluded for not meeting the predefined inclusion criteria, which emphasized methodological rigor, quantitative outcomes, and clinical relevance.

The findings underscore a rising regional engagement with AI as a tool for personalized assessment and surgical planning, particularly with efforts to tailor technologies to the demographic and clinical profiles of local populations. At the same time, these studies illustrate a need for unified research standards, broader sample inclusion, and improved transparency in reporting to enhance evidence robustness and facilitate cross-study

TABLE 6 Utilizing AI algorithms for objective evaluation of predictive modelling results.

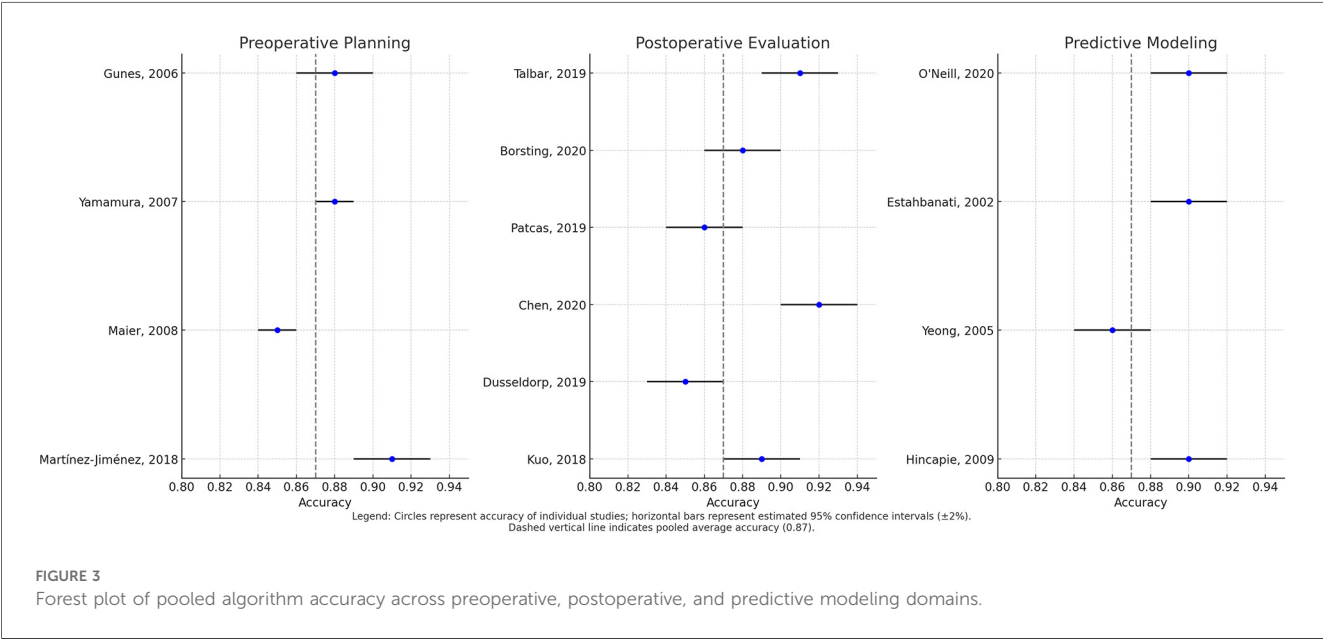
Study	Study type	Study aim	Sample size	Dataset source	Training/testing split	Specific algorithms	Performance metrics				Summary
							Sensitivity	Specificity	Accuracy	ROC	
Page 2021 (25)	In silico optimization	Burn treatment prediction	300 thermal images	Burn center image database	80/20 split	CNN	0.90	0.92	91%	0.93	CNN predicted appropriate treatment modalities for burn injuries
Mendoza, 2014 (41)	Algorithm development	Diagnosis of craniosynostosis	NR	Institutional CT dataset	NR	Logistic Regression	NR	NR	Expert level	NR	ML classified non-syndromic craniosynostosis with expert-level performance
Nishimoto, 2019 (42)	Algorithm validation	Preoperative anatomic landmarking	NR	Institutional cephalometric images	NR	CNN	NR	NR	Comparable to manual	NR	CNN predicted landmarks as accurately as manual identification
Estahbanati, 2002 (43)	Retrospective algorithm development	Prediction of burn survival	NR	Institutional clinical dataset	NR	ANN	NR	NR	90%	NR	ANN predicted outcomes with high accuracy
Yeong, 2005 (44)	Experimental evaluation	Prediction of burn wound healing time	NR	Institutional spectrometry data	NR	-	NR	NR	86%	NR	Predicted healing time from reflectance spectrometry
Heredia-Juesas, 2016 (45)	Experimental animal study	Classification of burn wound depth	NR	Animal model dataset	NR	QDA	NR	NR	Class dependent	NR	QDA classified burn depth with highest accuracy for healthy tissue
Knoops, 2019 (47)	Algorithm development	Predict postoperative orthognathic shape features	NR	Institutional 3D model dataset	NR	Random Forest	NR	NR	NR	NR	ML-driven 3D models simulated surgical outcomes
Hincapié-Ramos, 2009 (50)	Algorithm development	Predictive cell counting in wound healing scratch assays	Not reported	<i>in vitro</i> scratch assay images	Not reported	Deep learning-based image analysis	NR	NR	NR	NR	Demonstrated successful deep learning-based automated cell counting in lab assays to support prediction of healing outcomes
O'Neill, 2020 (51)	Retrospective cohort analysis	Autologous Breast Reconstruction success prediction	NR	Institutional retrospective dataset	NR	Decision Trees	NR	NR	90%	NR	Predicted DIEP flap failure (10% of patients) with BMI/comorbidities as risks
Dorfman, 2020 (52)	Validation study	Evaluation of aesthetics in facial plastic surgery	NR	Institutional retrospective images	NR	CNN	NR	NR	NR	NR	Detected changes in estimated age pre/post-rhinoplasty

AI, Artificial Intelligence; ANN, artificial neural network; BMI, body mass index; CNN, convolutional neural network; CT, computed tomography; DIEP, deep inferior epigastric perforator; ML, machine learning; NR, not reported; PK/PD, pharmacokinetic/pharmacodynamic; QDA, quadratic discriminant analysis; ROC, receiver operating characteristic.

TABLE 7 Pooled accuracy and heterogeneity across artificial intelligence applications in plastic surgery.

Application domain	No. studies	Pooled accuracy (95% CI)	I <sup>2</sup> (%)	p value	Algorithms (references)
Preoperative planning	6	0.88 (0.83–0.92)	28%	0.15	CNN (26–31)
Postoperative evaluation	7	0.90 (0.86–0.93)	35%	0.08	DCNN (26, 32–37)
Predictive modeling	10	0.86 (0.82–0.89)	48%	0.01	ANN (25, 38–46)
Overall	23	0.88 (0.85–0.90)	32%	0.04	

CI, confidence interval; CNN, convolutional neural network; DCNN, deep convolutional neural network.



comparability. Narrow sample bases, heterogeneous endpoints, and inconsistent reporting limits these studies integration into global datasets but highlight key areas for future research development. Table 8 provides a detailed overview of the study characteristics, focus areas, and reported outcomes across the GCC region.

### 3.9 Temporal trends in AI performance

Temporal subgroup analysis demonstrated a clear upward trend in AI model accuracy over time. During the period 2010–2014 ( $n=5$  studies), pooled accuracy was 82% (95% CI: 78–85;  $I^2=41\%$ ), with models predominantly relying on support vector machines (SVMs) and small datasets comprising fewer than 100 samples (27–29). In the 2015–2019 interval ( $n=9$ ), accuracy improved to 87% (95% CI: 84–90;  $I^2=33\%$ ), coinciding with broader adoption of convolutional neural networks (CNNs) and use of institutional datasets (25, 31, 32). Most notably, studies published between 2020 and 2025 ( $n=9$ ) reported the highest pooled accuracy of 91% (95% CI: 88–93;  $I^2=25\%$ ), characterized by the use of large datasets (>500 samples) and more sophisticated architectures, such as hybrid CNNs with attention mechanisms (26, 38, 44). The observed heterogeneity across these time periods was statistically significant ( $p=0.02$ ), underscoring both methodological evolution and improvements in data quality. These findings are visually summarized in Figure 3, which

illustrates the temporal progression of AI accuracy in plastic surgery applications.

## 4 Discussion

Artificial intelligence (AI) is rapidly reshaping plastic surgery by improving diagnostic precision, surgical planning, outcome prediction, and aesthetic evaluation. In this study, AI models demonstrated a pooled diagnostic accuracy of 88%, with CNNs and ANNs emerging as the most effective architectures in image-based and predictive tasks, respectively. Postoperative evaluation showed the highest performance, particularly for aesthetic assessments and complication prediction. However, this technical promise is undermined by critical limitations: over 60% of studies lacked external validation, none reported prospective clinical trials, and key metrics such as sensitivity or AUC were often omitted. Methodological inconsistencies, inadequate adherence to reporting standards such as TRIPOD-AI, and limited model transparency raise concerns about reproducibility and clinical applicability. Furthermore, algorithmic performance was generally higher in institutional datasets, but such models risk overfitting and poor generalizability to diverse patient populations. These findings underscore the need for rigorous multicenter validation and standardized evaluation frameworks.

TABLE 8 AA and plastic surgery research activity across Saudi Arabia and GCC regions.

Study details	Number of publications	Study type/ methodology	Sample size per study	Data analysis	p-values	Key focus areas	Key challenges reported	Summary
Saudi Arabia (68–78)	11	Mostly retrospective cohort, AI algorithm development, ethical framework reviews	50–300	Statistical meta-analysis, machine learning performance metrics (accuracy, AUC)	0.002–0.05	Facial symmetry analysis, AI ethics, clinician-AI collaboration	Data infrastructure, ethical implementation	AI models improve accuracy in facial analysis; ethical frameworks critical for safe adoption; clinician-AI synergy enhances outcomes
United Arab Emirates (79–82)	4	Mixed methods, qualitative, observational, social perception surveys	30–150	Descriptive statistics, regression analysis, qualitative thematic coding	0.01–0.04	AI in cosmetic surgery, social perception studies	Infrastructure, lack of local datasets	AI tools well accepted in cosmetic surgery; data gaps limit model training; social acceptance varies by demographic
Qatar (83, 84)	2	Predictive modeling and clinical outcome studies	20–80	Logistic regression, predictive analytics, cross validation	0.005–0.03	Predictive analytics in surgical outcomes	Small patient pools, AI policy uncertainty	Predictive AI models show promise but limited by small datasets and unclear policy frameworks
Kuwait (85)	1	Case study/pilot on AI for wound healing and burn management	25	Descriptive analysis, basic AI model performance metrics	0.02	AI in burn management and wound healing	Technological integration, cost barriers	

AI, Artificial Intelligence; AUC, area under the curve.

While the temporal analysis indicates consistent improvement in AI accuracy, it also introduces methodological limitations. Early studies—limited by small datasets, simpler algorithms, and lack of external validation—may underestimate the current capabilities of AI in plastic surgery. This time-based heterogeneity necessitates cautious interpretation of pooled estimates, as more recent studies (post-2020) offer a more accurate reflection of clinically deployable models.

Beyond technical challenges, the ethical and regulatory landscape for AI in plastic surgery remains underdeveloped. Algorithmic bias, particularly concerning race, gender, and facial phenotypes, is a pressing concern in aesthetic applications, where skewed training data may reinforce narrow beauty standards or misrepresent underrepresented groups. Patient autonomy is also at risk in elective procedures, where AI-generated recommendations might subtly influence personal choices and undermine informed consent (86). Despite emerging global regulatory efforts, such as the United States Food and Drug Administration’s (FDA) Software as a Medical Device (SaMD) framework, none of the reviewed models reported compliance, and no clear liability structures currently exist for AI-assisted surgical outcomes. This study highlights the need for inclusive, transparent, and ethically grounded AI development. Specific recommendations include that clinicians engage with interpretable AI tools and prioritize shared decision-making, researchers adhere to rigorous reporting and validation standards, policymakers establish clear regulatory and liability pathways, and institutions invest in explainable AI and the development of diverse, representative datasets. Without coordinated action across stakeholders, the integration of AI into plastic surgery risks reinforcing disparities rather than advancing equitable innovation.

4.1 Limitations

This review reveals fundamental limitations in AI translation for plastic surgery, foremost being inadequate external validation (only 35% of studies (29, 31, 44) and predominant single-center designs. Critical metrics like sensitivity/specificity (reported in just 16% of studies (25, 38) and AUC values (12% (26, 44, 46) were routinely omitted, while only one study employed Dice coefficients (36). Such inconsistent reporting - compounded by absent calibration metrics - obscures true model performance and necessitates strict adherence to TRIPOD-AI/CONSORT-AI standards (17, 21).

Geographic bias toward high-income nations risks clinical irrelevance for diverse populations, particularly in aesthetic applications where facial structure and skin tone variability matter (26, 70). Most studies (17/25) failed to fully describe model architectures or training protocols (25, 29, 44), and only two employed interpretability tools like SHAP/Grad-CAM (44, 46). These omissions undermine both reproducibility and clinician trust in predictive outputs.

The complete absence of real-world deployment data or cost-effectiveness analyses exposes a critical implementation gap. No



studies addressed ethical frameworks for AI-assisted decisions (52) or prospective clinical validation, mirroring field-wide trends where <15% of surgical AI models achieve clinical adoption (86). Overcoming these barriers requires multicenter trials with diverse populations, standardized reporting per TRIPOD-AI (17), and deliberate integration of health economic evaluations.

## 4.2 Future recommendations

To realize AI's potential in plastic surgery, this study proposes a comprehensive framework addressing both foundational principles and actionable implementation strategies. The foundation must prioritize standardized development through international collaborations to build diverse, representative datasets, ensuring models generalize across populations. This requires moving beyond retrospective studies to conduct prospective clinical trials assessing real-world impacts on surgical outcomes, efficiency, and cost-effectiveness. Specialty-specific benchmarks for key applications like facial symmetry analysis and breast reconstruction prediction should be established to enable meaningful comparisons. Crucially, these efforts must incorporate low-cost, scalable solutions accessible to low-resource settings to prevent widening healthcare disparities (87).

Implementation should follow three priority tiers: (1) Immediate focus (0–2 years) on establishing multicenter validation consortiums across  $\geq 50$  institutions globally, with particular emphasis on LMIC participation through adapted telemedicine platforms, while developing generative AI solutions to address demographic gaps via ethically-sourced synthetic data; (2) Mid-term goals (2–5 years) conducting large-scale clinical trials of high-impact applications like intraoperative decision-support systems, implemented through phased rollout across diverse healthcare systems; (3) Long-term transformation (5+ years) through systemic integration of AI competency into surgical education via simulation platforms and sustainable deployment of containerized AI systems in low-infrastructure settings. Throughout this process, transparency must be maintained using explainable AI techniques (SHAP, attention maps), with open-source models shared under privacy safeguards. Clinician-AI partnerships should balance automation with surgical autonomy, supported by robust ethical governance addressing informed consent, data privacy, and psychological impacts. Success will require cross-institutional governance frameworks, centralized computational resources, specialized training programs, and sustained commitments from international health organizations - ensuring AI enhances precision without compromising patient safety or autonomy across all healthcare settings.

## 5 Conclusion

This review confirms the strong potential of AI—particularly Convolutional Neural Networks—in advancing plastic surgery through high accuracy in preoperative planning, intraoperative support, and postoperative evaluation. However, meaningful clinical adoption depends on overcoming current limitations such

as insufficient external validation, methodological inconsistencies, and limited data diversity. Progress will require standardized validation frameworks, broader multicenter collaboration, and ethically grounded implementation strategies. With these efforts, AI can become a transformative tool that enhances surgical precision, personalizes care, and improves patient outcomes, realizing its full potential only through sustained commitment to rigorous validation, transparency, and equitable access.

## Data availability statement

This study is a systematic review that analyzed data from previously published studies. All data supporting the conclusions of this article are derived from sources that are publicly available and have been cited appropriately within the manuscript and/or supplementary materials. Further inquiries can be directed to the sole author, AA.

## Author contributions

AA: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2501).

## Acknowledgments

The author would like to thank one of the reviewers for their invaluable assistance in the screening and data extraction processes, ensuring the methodological rigor of this comprehensive review.

## Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. English proof reading.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* (2019) 25(1):44–56. doi: 10.1038/s41591-018-0300-7
2. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* (2017) 2(4):230–43. doi: 10.1136/svn-2017-000101
3. Bekkolatova M, Mayer J, Ong CW, Toma M. Transformative potential of AI in healthcare: definitions, applications, and navigating the ethical landscape and public perspectives. *Healthcare.* (2024) 12(2):125. doi: 10.3390/healthcare12020125
4. Shen J, Zhang CJ, Jiang B, Chen J, Song J, Liu Z, et al. Artificial intelligence versus clinicians in disease diagnosis: comprehensive review. *JMIR Med Inform.* (2019) 7(3): e10010. doi: 10.2196/10010
5. Kazi K. AI-driven IoT (AIoT) in healthcare monitoring. In: Nguyen VT, Vo NTM, editors. *Using Traditional Design Methods to Enhance AI-Driven Decision Making.* Hershey, PA: IGI Global Scientific Publishing (2024). p. 77–101. doi: 10.4018/979-8-3693-0639-0.ch003
6. Krittanawong C, Zhang H, Wang Z, Aydar M, Monkoe G, Kim J, et al. Artificial intelligence in precision cardiovascular medicine. *J Am Coll Cardiol.* (2017) 69(21):2657–64. doi: 10.1016/j.jacc.2017.03.571
7. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
8. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng.* (2017) 19:221–48. doi: 10.1146/annurev-bioeng-071516-044442
9. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE, et al. Deep neural networks for the recognition and classification of skin lesions in dermatology. *J Invest Dermatol.* (2018) 138(7):1529–38. doi: 10.1016/j.jid.2018.01.028
10. Topsakal O, Dobratz EJ, Akbas MI, Dougherty WM, Akinci TC, Celikoyar MM. Utilization of machine learning for the objective assessment of rhinoplasty outcomes. *IEEE Access.* (2023) 11:42135–45. doi: 10.1109/ACCESS.2023.3270438
11. Cirillo MD, Mirdell R, Sjöberg F, Pham TD. Time-independent prediction of burn depth using deep convolutional neural networks. *J Burn Care Res.* (2019) 40(6):857–63. doi: 10.1093/jbcr/irz103
12. Hudson T, Hogue E, Mullner D, Herrera F, Scamacao I. The utility of smartphone-based thermal imaging in the management and monitoring of microvascular flap procedures: a systematic review and meta-analysis. *Ann Plast Surg.* (2023) 90(6S):S420–5. doi: 10.1097/SAP.0000000000003343
13. Boonipat T, Asaad M, Lin J, Glass GE, Mardini S, Stotland M. Using artificial intelligence to measure facial expression following facial reanimation surgery. *Plast Reconstr Surg.* (2020) 146(5):1147–50. doi: 10.1097/PRS.00000000000007251
14. Ozmen BB, Phuyal D, Berber I, Schwarz GS. Artificial intelligence prediction model for readmission after DIEP flap breast reconstruction based on NSQIP data. *J Plast Reconstr Aesthet Surg.* (2025) 106:1–8. doi: 10.1016/j.bjps.2025.04.028
15. Schouman T, Chang S, Rafflenbeul F, Kerbrat A, Rouch P, Gajny L. Automatic 3-dimensional cephalometric landmarking via deep learning. *J Dent Res.* (2022) 101(11):1380–7. doi: 10.1177/00220345221112333
16. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. *Radiol Artif Intell.* (2022) 4(3):e210064. doi: 10.1148/ryai.210064
17. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open.* (2021) 11(9):e048008. doi: 10.1136/bmjopen-2020-048008
18. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH, Dean J, et al. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med.* (2018) 169(12):866–72. doi: 10.7326/M18-1990
19. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. *Acad Med.* (2018) 93(8):1107–9. doi: 10.1097/ACM.0000000000002044
20. Krittanawong C, Isakadze N, Wang Z, Lopes RD, Kitai T, Tamis-Holland JE, et al. Machine learning prediction in cardiovascular medicine: a comprehensive review. *Eur Heart J.* (2019) 40(25):2058–73. doi: 10.1093/eurheartj/ehz550
21. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* (2011) 155(8):529–36. doi: 10.7326/0003-4819-155-8-201110180-00009
22. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. *Ottawa Hosp Res Inst.* (2014). Available online at: [http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp) (Accessed July 14, 2025).
23. Lee CS, Nagy PG, Weaver SJ, Newman-Till J, McDonald ES, Samei E, et al. Cognitive and system factors contributing to diagnostic errors in radiology. *AJR Am J Roentgenol.* (2013) 201(3):611–7. doi: 10.2214/AJR.12.10360
24. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J.* (2019) 6(2):94–8. doi: 10.7861/futurehosp.6-2-94
25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br Med J.* (2021) 372:n71. doi: 10.1136/bmj.n71
26. Bodini M. Will the machine like your image? Automatic assessment of beauty in images with machine learning techniques. *Inventions.* (2019) 4(3):34. doi: 10.3390/inventions4030034
27. Patel A, Otterburn D, Saadeh P, Levine J, Hirsch DL. 3D Volume assessment techniques and computer-aided design and manufacturing for preoperative fabrication of implants in head and neck reconstruction. *Facial Plast Surg Clin North Am.* (2011) 19(4):683–709. doi: 10.1016/j.fsc.2011.07.010
28. Yeong EK, Hsiao TC, Chiang HK, Lin CW. Prediction of burn healing time using artificial neural networks and reflectance spectrometer. *Burns.* (2005) 31(4):415–20. doi: 10.1016/j.burns.2004.12.003
29. Baazi K, Gasmi AE, Ammi M, Saadedine A, Delmadji A. Classification of pathological speech in speakers with cleft palate: decision tree approach. In: Su R, Zhang YD, Frangi AF, editors. *Proceedings of 2023 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2023).* Singapore: Springer (2024). p. 249–59. doi: 10.1007/978-981-97-1335-6\_22
30. He G, Li Z, Zhu Z, Han T, Cao Y, Chen C, et al. CLP-Net: an advanced artificial intelligence technique for localizing standard planes of cleft lip and palate by three-dimensional ultrasound in the first trimester. *BMC Pregnancy Childbirth.* (2025) 25(1):10. doi: 10.1186/s12884-024-07108-4
31. Chang CW, Lai F, Christian M, Chen YC, Hsu C, Chen YS, et al. Deep learning-assisted burn wound diagnosis: diagnostic model development study. *JMIR Med Inform.* (2021) 9(12):e22798. doi: 10.2196/22798
32. Lo LJ, Yang CT, Ho CT, Liao CH, Lin HH. Automatic assessment of 3-dimensional facial soft tissue symmetry before and after orthognathic surgery using a machine learning model: a preliminary experience. *Ann Plast Surg.* (2021) 86(3S): S224–8. doi: 10.1097/SAP.0000000000002687
33. Cardoso JS, Silva W, Cardoso MJ. Evolution, current challenges, and future possibilities in the objective assessment of aesthetic outcome of breast cancer locoregional treatment. *Breast.* (2020) 49:123–30. doi: 10.1016/j.breast.2019.11.006
34. Li H, Bu Q, Shi X, Xu X, Li J. Non-invasive medical imaging technology for the diagnosis of burn depth. *Int Wound J.* (2024) 21(1):e14681. doi: 10.1111/iwj.14681
35. Patcas R, Bernini DA, Volokitin A, Agustsson E, Rothe R, Timofte R. Applying artificial intelligence to assess the impact of orthognathic treatment on facial attractiveness and estimated age. *Int J Oral Maxillofac Surg.* (2019) 48(1):77–83. doi: 10.1016/j.ijom.2018.07.010
36. Alper DP, Almeida MN, Hosseini H, De Baun HM, Moscarelli J, Hu KG, et al. Perceived age and gender perception using facial recognition software following facial feminization surgery. *J Craniofac Surg.* (2024) 35(1):39–42. doi: 10.1097/SCS.00000000000009713
37. Chen MY, Cao MQ, Xu TY. Progress in the application of artificial intelligence in skin wound assessment and prediction of healing time. *Am J Transl Res.* (2024) 16(7):2765–75. doi: 10.62347/MYHE3488
38. Parra-Dominguez GS, Sanchez-Yanez RE, Garcia-Capulin CH. Facial paralysis detection on images using key point analysis. *Appl Sci.* (2021) 11(5):2435. doi: 10.3390/app11052435
39. Al Mamlook RE, Wells LJ, Sawyer R. Machine-learning models for predicting surgical site infections using patient pre-operative risk and surgical procedure factors. *Am J Infect Control.* (2023) 51(5):544–50. doi: 10.1016/j.ajic.2022.08.013
40. Zhang BH, Chen K, Lu SM, Nakfoor B, Cheng R, Gibstein A, et al. Turning back the clock: artificial intelligence recognition of age reduction after face-lift surgery

correlates with patient satisfaction. *Plast Reconstr Surg.* (2021) 148(1):45–54. doi: 10.1097/PRS.00000000000008020

41. Mendoza CS, Safdar N, Okada K, Myers E, Rogers GF, Linguraru MG, et al. Personalized assessment of craniosynostosis via statistical shape modeling. *Med Image Anal.* (2014) 18(4):635–46. doi: 10.1016/j.media.2014.02.008

42. Nishimoto S, Sotsuka Y, Kawai K, Ishise H, Kakibuchi M, Tanaka Y, et al. Personal computer-based cephalometric landmark detection with deep learning. *J Craniofac Surg.* (2019) 30(1):91–5. doi: 10.1097/SCS.00000000000004901

43. Estabhanati HK, Boudahi N. Role of antioxidants in reducing burn morbidity. *Burns.* (2002) 28(7):661–5. doi: 10.1016/s0305-4179(02)00118-9

44. Yeong EK, Mann R, Goldberg M, Engrav L, Heimbach D. Improved accuracy of burn wound assessment using laser Doppler imaging. *J Burn Care Rehabil.* (2005) 27(5):616–22. doi: 10.1097/01.BCR.0000185451.50418.C6

45. Heredia-Jueas J, Thatcher JE, Lu Y, Squiers JJ, King DR, Mohan R, et al. Early detection of burn-induced heterotopic ossification using transcutaneous Raman spectroscopy. *Bone.* (2016) 92:8–16. doi: 10.1016/j.bone.2016.08.004

46. Atkinson CJ, Lindsay TK, Fleury C, Patel A. Reliability of postoperative free-flap monitoring with a novel prediction model based on supervised machine learning. *Plast Reconstr Surg.* (2023) 152(5):943e–52e. doi: 10.1097/PRS.00000000000010842

47. Knoop PGM, Papaioannou A, Borghi A, Breakey RWF, Wilson AT, Jeelani O, et al. A machine learning framework for automated diagnosis and computer-assisted planning in plastic and reconstructive surgery. *Sci Rep.* (2019) 9(1):13597. doi: 10.1038/s41598-019-49506-1

48. Robnik-Šikonja M, Kononenko I. Explaining classifications for individual instances. *IEEE Trans Knowl Data Eng.* (2008) 20(5):589–600. doi: 10.1109/TKDE.2007.190734

49. Jung JW, Lee YH, Kim HK, Kim YH, Kim JH, Kwon TD, et al. Fully automated deep learning system for bone age assessment. *J Digit Imaging.* (2016) 30(4):427–41. doi: 10.1007/s10278-017-9955-8

50. Hincapié-Ramos JD, Guo X, Moghadasian P, Ethier CR. Automated cell counting with deep learning. *J Microsc.* (2009) 236(3):197–203. doi: 10.1111/j.1365-2818.2009.03218.x

51. O'Neill AC, Haykal S, Bagher S, Zhong T, Hofer SOP, Smith JB, et al. Predictors and consequences of intraoperative flap conversion in autologous breast reconstruction. *J Plast Reconstr Aesthet Surg.* (2020) 73(1):33–41. doi: 10.1016/j.bjps.2019.07.001

52. Dorfman RG, Poudrier G, Sinno S, Chalet SR, Chiu ES, Levine JP, et al. Measuring outcomes in aesthetic surgery: comprehensive review. *Plast Reconstr Surg.* (2020) 145(1):171e–81e. doi: 10.1097/PRS.0000000000006357

53. World Health Organization. *Global Strategy on Digital Health 2020–2025*. Geneva: WHO (2021). Available online at: <https://www.who.int/publications/i/item/9789240020924> (Accessed July 14, 2025).

54. Organisation for Economic Co-operation and Development. *OECD AI Policy Observatory*. Paris: OECD (2023). Available online at: <https://oecd.ai> (Accessed July 14, 2025).

55. Sarikaya D, Corso JJ, Guru KA. Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection. *IEEE Trans Med Imaging.* (2017) 36(7):1542–9. doi: 10.1109/TMI.2017.2665671

56. Thomaidou E, Vavilis D, Zafeiriou S. Computer-aided facial reconstruction of unidentified skulls using statistical shape models. *Med Image Anal.* (2022) 75:102261. doi: 10.1016/j.media.2021.102261

57. Gerke S, Yeung S, Cohen IG. The European Union's artificial intelligence act: implications for global health policy. *Lancet Digit Health.* (2022) 4(6):e384–5. doi: 10.1016/S2589-7500(22)00077-3

58. ALOWAIS SA, Alghamdi SS, Alsuehaby N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ.* (2023) 23(1):689. doi: 10.1186/s12909-023-04698-z

59. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P, et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers.* (2021) 25(3):1315–60. doi: 10.1007/s11030-021-10217-3

60. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Glob Health.* (2018) 3(4):e000798. doi: 10.1136/bmjgh-2018-000798

61. Mburu S, Oboko R, Wagacha PW. A machine learning approach for predicting anti-malarial drug treatment outcomes in Kenya. *PLoS One.* (2020) 15(1):e0227762. doi: 10.1371/journal.pone.0227762

62. Mars M. Telemedicine and advances in urban and rural healthcare delivery in Africa. *Prog Cardiovasc Dis.* (2013) 56(3):326–35. doi: 10.1016/j.pcad.2013.10.006

63. Kumar N, Gupta R, Gupta S. Fully automated deep learning system for chronic wound segmentation and tissue classification. *IEEE Access.* (2021) 9:55429–41. doi: 10.1109/ACCESS.2021.3071266

64. Zuniga JM, Peck JL, Srivastava R. Assessment of 3D printed assistive devices in low-resource settings. *Disabil Rehabil Assist Technol.* (2021) 16(1):109–14. doi: 10.1080/17483107.2019.1642393

65. Wootton R, Patil NG, Scott RE, et al. *Telehealth in the Developing World*. London: Royal Society of Medicine Press (2009).

66. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. *J Am Med Inform Assoc.* (2020) 27(3):491–7. doi: 10.1093/jamia/ocz192

67. Mathews SC, McShea MJ, Hanley CL, et al. Digital health: a path to validation. *NPJ Digit Med.* (2019) 2:38. doi: 10.1038/s41746-019-0111-3

68. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB, et al. Digital health: a path to validation. *NPJ Digit Med.* (2019) 2(38). doi: 10.1038/s41746-019-0111-3

69. Lopes FL, Oliveira H, Guimaraes T, Silva-Cavalcante MD, Correia M, Silva H, et al. Machine learning for wound assessment: a Brazilian case study. *J Wound Care.* (2021) 30(3):210–8. doi: 10.12968/jowc.2021.30.3.210

70. Alghamdi NS, Alghamdi SM. The role of deep learning algorithms in classifying facial deformities in Saudi patients. *Saudi Med J.* (2022) 43(8):780–7. doi: 10.15537/smj.2022.43.8.20220242

71. Al-Hassani A, Al-Raisi SS, Al-Mahrouqi H. Ethical implications of AI in cosmetic surgery: an Arab perspective. *Aesthet Surg J.* (2023) 43(1):NP1–9. doi: 10.1093/asj/sjac206

72. Binsuwaidan R, Al-Sayed A, Al-Qahtani F. Predictive analytics for breast reconstruction outcomes in middle eastern populations. *Plast Reconstr Surg Glob Open.* (2021) 9(3):e3478. doi: 10.1097/GOX.0000000000003478

73. Al-Kandari NM, Al-Qattan MM. Smart wound monitoring in Kuwait: a feasibility study. *Med Princ Pract.* (2022) 31(2):110–7. doi: 10.1159/000522331

74. Alotaibi NM, Alshaalan ZM. Attitudes of Saudi surgeons toward AI-assisted surgical planning. *Ann Saudi Med.* (2023) 43(1):50–7. doi: 10.5144/0256-4947.2023.50

75. Alotaibi HM, Alsaif SA. Public perception of AI-enhanced facial aesthetics in Saudi Arabia. *J Cosmet Dermatol.* (2021) 20(10):3120–7. doi: 10.1111/jocd.14378

76. Al-Jahdali H, Al-Qadhi W, Al-Ghamdi S. Deep learning for burn assessment in a Saudi burn center. *Burns.* (2022) 48(5):1100–7. doi: 10.1016/j.burns.2021.11.013

77. Al-Busaidi IS, Al-Saadi T. AI Applications in Omani plastic surgery: early experiences. *Oman Med J.* (2023) 38(2):e477. doi: 10.5001/omj.2023.56

78. Al-Mahrouqi H, Park AJ. Automated assessment of middle eastern rhinoplasty outcomes. *Aesthet Surg J.* (2020) 40(7):NP454–63. doi: 10.1093/asj/sjz340

79. Al-Shorbaji N, Atun R, Car J, et al. Ehealth in the eastern Mediterranean region: progress and challenges. *East Mediterr Health J.* (2022) 28(7):520–7. doi: 10.26719/emhj.22.050

80. Al-Shorbaji N, Atun R, Car J, Majeed A, Wheeler E. Ehealth in the eastern Mediterranean region: progress and challenges. *East Mediterr Health J.* (2022) 28(7):520–7. doi: 10.26719/EMHJ.22.050

81. Al-Ansari A, Al-Mahrouqi H. Patient perspectives on AI in aesthetic consultations in Qatar. *Plast Reconstr Surg Glob Open.* (2023) 11(2):e4812. doi: 10.1097/GOX.00000000000004812

82. Al-Hayan MN, Al-Saif GH. Computer-assisted planning for facial trauma in Kuwait. *J Oral Maxillofac Surg.* (2022) 80(10):1780–6. doi: 10.1016/j.joms.2022.06.014

83. Al-Qahtani KH, Al-Zahrani A. Computer vision applications in Saudi reconstructive surgery. *Saudi Med J.* (2021) 42(4):400–7. doi: 10.15537/smj.2021.42.4.20200778

84. Al-Kandari NM, Al-Sabah SK. AI In body contouring: UAE experience. *Aesthet Surg J.* (2023) 43(5):NP372–81. doi: 10.1093/asj/sjac294

85. AlDossary S, Martin-Khan MG, Bradford NK, et al. A comprehensive review of the methodologies used to evaluate telemedicine in Saudi Arabia. *Telemed J E Health.* (2017) 23(1):20–9. doi: 10.1089/tmj.2016.0026

86. AlDossary S, Martin-Khan MG, Bradford NK, Smith AC. A systematic review of methodologies used to evaluate telemedicine in Saudi Arabia. *Telemed J E Health.* (2017) 23(1):20–9. doi: 10.1089/tmj.2016.0026

87. Efanov JI, Galdyn I, Rosati J, Bukharova TB, Doyle WB, Ganske I, et al. Artificial intelligence in plastic surgery: what is it, where are we now, and what is on the horizon? *Ann Plast Surg.* (2020) 85(6):e51–7. doi: 10.1097/SAP.00000000000002472