

Chapter

AI-Enhanced Bariatric and Endobariatric Surgery: Predictive Analytics, Complication Surveillance and Intelligent Patient Care

Mohammed Maree

Abstract

Artificial intelligence (AI) and machine learning (ML) are increasingly influencing bariatric surgery by enabling personalized care and better risk management. This review summarizes recent evidence on AI/ML in surgical and endoscopic weight-loss procedures, based on approximately 300 studies from 2018 to 2025. Findings cover preoperative risk prediction, postoperative monitoring, intraoperative AI, endobariatric applications, long-term outcome prediction, and ethical considerations. Preoperative AI models, such as neural networks and deep learning, modestly improved complication prediction with AUROC gains of 0.02–0.12 over traditional methods. Predictive models for sleeve gastropasty achieved high accuracy for 12-month weight loss using early postoperative data. Long-term models reliably forecast five-year BMI, and deep learning outperformed scores in predicting diabetes remission. However, limitations include class imbalance, rare-adverse-event detection, limited external validation, calibration gaps, and scant evidence for some endobariatric procedures. Intraoperative computer vision is still in its early stages of development. Successful AI implementation requires multicenter validation, explainability, adherence to standards, and versatile strategies for diverse healthcare settings.

Keywords: artificial intelligence, machine learning, deep learning, bariatric surgery, metabolic surgery, endobariatric surgery, endoscopic sleeve gastropasty, predictive modeling, computer vision, risk stratification

1. Introduction

1.1 The global obesity epidemic and weight-loss interventions

Obesity affects more than 650 million adults worldwide and is a major driver of cardiovascular disease, type 2 diabetes, several malignancies, impaired quality of life, and premature mortality [1]. Treatment options span a spectrum from lifestyle modification and pharmacotherapy to minimally invasive endoscopic interventions and definitive metabolic and bariatric surgery.

Surgical procedures such as Roux-en-Y gastric bypass (RYGB) and sleeve gastrectomy (SG) consistently outperform medical and lifestyle management in long-term weight reduction and metabolic outcomes, with growing evidence suggesting superior durability compared with pharmacotherapy alone [2]. Over the last decade, endoscopic bariatric therapies such as endoscopic sleeve gastroplasty (ESG), intragastric balloons, transoral outlet reduction (TORe), and primary obesity surgery endoluminal (POSE) have emerged as less invasive options to traditional surgery and offer additional tools in the management of obesity when lifestyle and pharmacologic approaches alone are insufficient [3].

Despite these advances, bariatric and endobariatric procedures carry significant risks, including anastomotic or staple line leaks, venous thromboembolism (VTE), bleeding, perforation, and device-related issues such as balloon migration [3, 4]. Long-term trajectories are heterogeneous: a significant subset of patients experiences suboptimal weight loss, weight regain, or incomplete metabolic response [5]. This variability underscores the need for tools that personalize procedure selection, stratify risk, and guide long-term management.

1.2 Artificial Intelligence in surgical care

Artificial intelligence (AI), particularly machine learning (ML) and deep learning, enables the discovery of complex, nonlinear patterns in high-dimensional clinical data [6]. In surgery, AI applications now extend across the perioperative continuum: risk prediction, patient selection, intraoperative guidance, skill assessment, postoperative surveillance, and long-term outcomes modeling [7].

Standard supervised learning algorithms used in bariatric research include logistic regression (LR), support vector machines (SVM), random forests (RF), gradient boosting methods such as XGBoost, and various neural network architectures (multilayer perceptrons, convolutional neural networks (CNNs), and other deep learning pipelines) [8]. Compared to traditional scores, these models can integrate large numbers of variables, including demographics, comorbidities, laboratory data, imaging, and digital phenotypes, and often yield improved discrimination [6, 8]. However, improvements in AUROC do not automatically translate into better patient outcomes or safer care [9, 10].

1.3 Current landscape and review objectives

The application of AI and ML in bariatric and endobariatric interventions has expanded rapidly since 2018, as shown in **Figure 1**. This acceleration has been driven, in part, by the increasing availability of large, high-quality clinical registries, including the Metabolic and Bariatric Surgery Accreditation and

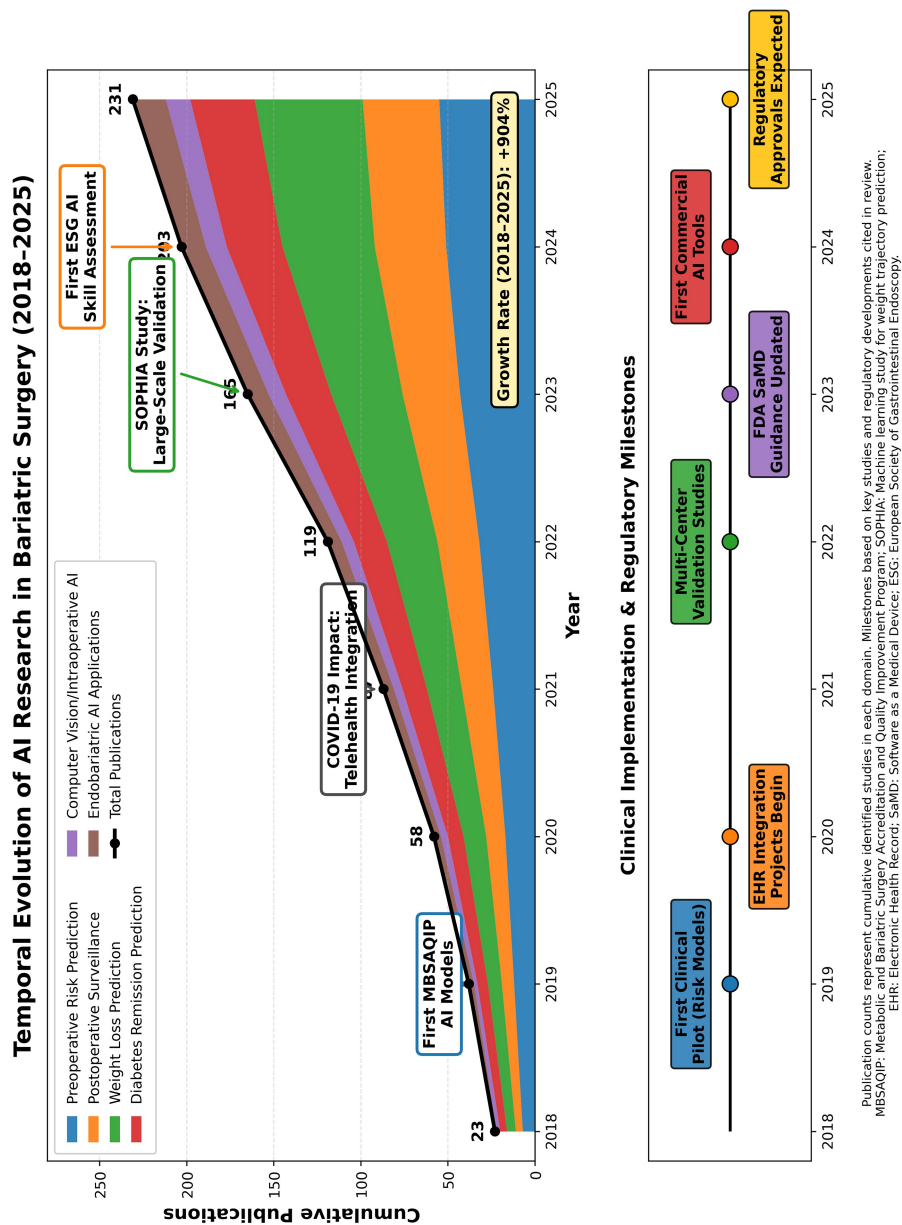


Figure 1. Temporal evolution of artificial intelligence research in bariatric surgery (2018–2025). **Upper panel:** stacked area chart showing cumulative publications across six application domains, demonstrating a 904% increase from 2018 (23 publications) to 2025 (231 publications). Preoperative risk prediction and weight-loss prediction are the most mature domains, while endobariatric applications show the steepest recent growth. Key milestones include early registry-based prediction models, pandemic-related digital integration, a large-scale external validation study for long-term weight outcomes, and the first skill assessment studies in endoscopic bariatric procedures. **Lower panel:** timeline of clinical and regulatory milestones, illustrating a typical three- to five-year lag between research publication and clinical implementation.

Quality Improvement Program (MBSAQIP) and the Scandinavian Obesity Surgery Registry (SOReg). These registries facilitate the development, external validation, and benchmarking of predictive models across diverse clinical settings [8, 11]. In parallel, the growing adoption of ESG and other endobariatric techniques has enabled smaller but methodologically innovative AI studies, often focusing on procedural feasibility, skill assessment, and short-term outcomes [8, 12].

This chapter synthesizes the available evidence across six thematic domains:

1. Preoperative risk prediction and patient selection
2. Postoperative surveillance and detection of complications or readmissions
3. Intraoperative AI and computer vision applications
4. AI applications in endobariatric procedures
5. Prediction of long-term weight loss and metabolic outcomes
6. Ethical, regulatory, and implementation considerations

The objectives of this chapter are to:

- Summarize current methodological approaches and performance metrics of AI/ML models in bariatric and endobariatric care
- Critically evaluate their strengths, limitations, and readiness for clinical implementation
- Identify key research gaps, particularly related to external validation, fairness, calibration, and real-world deployment
- Provide a clinically oriented and practical roadmap to guide future research and development

1.4 Chapter scope, structure, and intended audience

This chapter is a structured, nonsystematic narrative review, providing a high-level, clinically focused overview of AI applications in bariatric and endobariatric care, rather than an exhaustive systematic review or meta-analysis.

It is intended for surgeons and endoscopists, researchers and data scientists, and policymakers or health system leaders seeking insight into clinically relevant AI use cases, methodological limitations, translational challenges, and health-system implications.

Accordingly, the review emphasizes technical foundations and clinical integration of AI tools, alongside ethical and equity considerations, regulatory and governance requirements, and global implementation challenges, particularly in low- and middle-income countries (LMICs).

2. Methods

This narrative literature review employed a structured search across PubMed, Embase, Scopus, Web of Science, and Google Scholar using keywords including “bariatric surgery,” “metabolic surgery,” “endoscopic sleeve gastropasty,” “endobariatric,” “intra-gastric balloon,” “artificial intelligence,” “machine learning,” “deep learning,” “predictive modeling,” “computer vision,” and “risk prediction.” Initial database searches yielded over 300 potentially relevant records. Searches were iterative and exploratory, supplemented by citation tracking and gray literature screening. Studies published in English (2018–2025) that applied AI/ML to surgical and endoscopic weight-loss interventions with clinical validation were prioritized. Purely technical papers without clinical validation and expert opinions lacking empirical data were excluded. The synthesis incorporates findings from 51 studies alongside regulatory guidance and methodological literature. Thematic synthesis organized findings into conceptual domains. Methodological limitations were noted when relevant, although formal risk-of-bias tools were not systematically applied.

3. Thematic synthesis

3.1 Preoperative risk prediction and patient selection

3.1.1 Neural networks for complication prediction

Risk stratification facilitates shared decision-making and targeted perioperative care. Although tools such as the ACS-NSQIP risk calculator are widely used, studies evaluating their application in bariatric surgery populations have demonstrated only moderate discriminative performance (AUROC ~ 0.65–0.70), highlighting the need for more procedure-specific predictive approaches [8, 13].

Nudel et al. analyzed MBSAQIP data from 436,807 patients to develop artificial neural network (ANN; leak AUROC 0.75, VTE 0.65) and XGBoost models (leak AUROC 0.70, VTE 0.67), compared with LR (leak AUROC 0.63, VTE 0.64) for predicting gastrointestinal leak and VTE after bariatric surgery [14]. The ANN significantly outperformed LR for leak prediction ($p < 0.001$), whereas discrimination for VTE remained modest across all models. Strengths included the massive sample size and internal cross-validation; however, external validation was not performed, limiting generalizability.

Wise et al. developed a back-propagation artificial neural network using MBSAQIP data from 101,721 SG patients to predict 30-day morbidity and mortality, achieving only a modest improvement over LR (AUROC 0.585 vs. 0.572) and demonstrating the feasibility – but limited incremental performance – of neural networks for broader postoperative outcomes [15].

Together, these studies suggest that neural networks may yield clinically meaningful gains in the management of rare, high-impact complications. At the same time, performance improvements for more heterogeneous outcomes remain constrained by outcome prevalence, feature granularity, and data quality.

3.1.2 Comparative machine learning studies

Cao et al. used SOReg data to compare multiple supervised algorithms – including SVM, RF, gradient boosting, and feed-forward neural networks – for predicting severe postoperative complications after bariatric surgery [16]. Although ensemble and deep learning methods achieved training accuracies above 90%, their sensitivity on temporally separated test sets remained below 30%, revealing classic problems of class imbalance and overfitting. To address the rarity of serious complications, the same group subsequently applied the synthetic minority oversampling technique (SMOTE) and deeper neural architectures to the SOReg dataset [11]. SMOTE modestly improved sensitivity at the expense of some specificity, while deep learning offered only incremental performance gains. These findings suggest that simply adopting more complex architectures does not automatically solve the rare-event problem; alternative strategies such as cost-sensitive learning, focal loss, and anomaly-detection approaches warrant further evaluation in this domain.

3.1.3 Explainable AI and clinical integration

Zucchini et al. applied explainable AI techniques, including SHAP analyses, to ML models predicting 30-day complications in a single-center cohort of 424 bariatric surgery patients, with a RF model achieving an AUROC of 0.94 in training and 0.88 in internal testing, significantly outperforming the MBSAQIP risk score (AUROC 0.64, $p < 0.001$) [17]. However, despite strong discriminative performance, the single-center design, limited sample size, and absence of evidence that explainability alters clinician behavior or improves outcomes underscore a critical gap between model interpretability and real-world clinical adoption.

3.1.4 Limitations and clinical implications

Preoperative risk models consistently show low sensitivity for rare but severe events such as anastomotic leaks and VTE [14, 16]. Even when AUROC and specificity appear acceptable, this can lead to underestimation of complications and delayed intervention. Class imbalance often necessitates oversampling or weighting strategies such as SMOTE [11], which provides only partial mitigation. The frequent absence of external validation further constrains generalizability [14, 15, 17], as models developed within single registries or health systems may not perform reliably across diverse populations. In addition, registry-based data often lack critical clinical, nutritional, imaging, and behavioral variables, and many studies emphasize AUROC without adequate calibration assessment [9], resulting in predicted risks that may not reflect observed outcomes.

Consequently, although AUROC improvements of approximately 0.02–0.12 over LR [14, 15] are statistically notable, their clinical impact remains unproven. Translating discrimination gains into estimates of missed complications is essential for assessing real-world clinical relevance. Prospective studies are needed to determine whether AI-based risk stratification meaningfully alters perioperative management, improves patient outcomes, or optimizes resource utilization. Until such evidence emerges, clinicians should continue to rely on established risk assessment tools while remaining informed about evolving AI approaches that may ultimately deliver tangible benefits for personalized care and clinical decision-making.

3.2 Postoperative surveillance and complication detection

3.2.1 Readmission prediction models

Unplanned readmission after bariatric surgery occurs in approximately 3–8% of patients and represents a significant burden for both healthcare systems and patients [18–20]. Torquati et al. analyzed 393,833 patients from the MBSAQIP registry (3.9% readmission rate) and demonstrated that a super learner ensemble outperformed LR in predicting 30-day readmission; however, overall discrimination remained modest and did not reach levels consistent with clinical utility, likely reflecting severe class imbalance and methodological constraints [18].

In contrast, Zhang et al. examined a smaller cohort of 1,262 bariatric patients, integrating preoperative variables with postoperative laboratory results obtained on postoperative days 1 and 3 [19]. Following LASSO-based feature selection, the best-performing SVM achieved an AUROC of 0.784, outperforming LR. Models incorporating early postoperative laboratory trends consistently outperformed preoperative-only approaches.

Similarly, Butler et al. applied ML to a large MBSAQIP cohort of 863,348 patients (4.52% readmission rate) and reported an AUROC of 0.785 using gradient boosting methods [20]. Across studies, consistent predictors included operative time, preoperative BMI, diabetes, obstructive sleep apnea, and early postoperative laboratory abnormalities. Collectively, these findings suggest that while ML can yield statistically meaningful improvements in discrimination across both large registries and smaller laboratory-rich datasets, performance generally remains in the moderate range, external validation is limited, and class imbalance is infrequently addressed, constraining immediate clinical applicability.

3.2.2 Mobile health and patient-generated data

Farinella et al. integrated traditional clinical variables with app-based patient-generated data – including symptoms, activity, dietary adherence, and self-reported intake – collected via the Care4Today platform to develop ML models for predicting postoperative complications after bariatric surgery [21]. In a prospective cohort of 104 patients, models incorporating mobile app alerts achieved an AUROC of 0.715 and an accuracy of 77.4%, outperforming preoperative models based on baseline clinical data alone.

This study illustrates the potential of mobile health-enabled risk stratification, shifting prediction from static preoperative assessment toward iterative postoperative risk updating. By integrating multimodal clinical and behavioral signals, patient risk profiles may evolve, enabling earlier intervention. Although performance remains modest and external validation is needed, the findings support the feasibility of patient-generated data for AI-driven postoperative surveillance.

3.2.3 Complication-specific models

Hsu et al. applied ML to the MBSAQIP registry (n = 159,959) to predict gastrointestinal bleeding, achieving an AUROC of 0.764, with key predictors including anticoagulant use and operative time [22].

Similarly, Shan et al. developed an extreme gradient boosting (XGBoost) model to predict postoperative nausea and vomiting after laparoscopic SG, reporting an AUROC of 0.828 in a cohort of 860 patients; significant risk factors included female sex, operative duration >60 minutes, and remifentanyl use [23].

Beyond structured data, natural language processing (NLP) methods have been used to extract postoperative complications from operative notes and discharge summaries across multiple surgical domains [24]. Although bariatric-specific NLP applications remain limited, this approach offers a scalable strategy for automated complication surveillance and the generation of richer datasets for future ML models.

3.2.4 Summary and gaps

Postoperative ML models typically achieve moderate discrimination (AUROC ~ 0.75–0.80) for predicting readmission and selected complications, particularly when early postoperative data are included. However, most evidence remains retrospective, derived from single-center studies or large registries, and no models have yet demonstrated improved clinical outcomes or efficiency in prospective real-world use. Integration into electronic health records and clinical workflows is uncommon, and data on downstream clinical impact are largely absent. The principal gap, therefore, lies not in predictive accuracy but in actionability. Future work must focus on real-time, EHR-embedded decision support systems evaluated in pragmatic or randomized trials, with predefined interventions linked to risk thresholds and supported by decision curve or net benefit analyses to determine whether model-guided care meaningfully benefits patients.

3.3 Intraoperative AI and computer vision

3.3.1 Current state of evidence

Computer vision enables automated surgical video analysis for phase recognition, anatomical identification, instrument tracking, and skill assessment, with recent systems demonstrating real-time scene understanding across procedures [25, 26]. In bariatric surgery, proposed applications include landmark identification, intraoperative quality checks (e.g., staple line assessment), and objective performance evaluation; however, validated bariatric-specific systems remain scarce. Bellini et al. identified only a limited number of bariatric-focused tools, including SleeveNet (~82% phase-recognition accuracy) and Remaining Surgery Duration Network (RSDNet) for gastric bypass duration prediction [27]. Most mature systems originate from cholecystectomy and colorectal surgery, such as real-time anatomical segmentation [28] and automated assessment of the critical view of safety (~93% accuracy) [29], illustrating a largely unvalidated translation pathway for bariatric applications.

3.3.2 Barriers to real-time intraoperative deployment

Real-time deployment of computer vision in bariatric surgery remains limited by the scarcity of large, well-annotated bariatric video datasets, heterogeneity in

operative techniques, and the need for stable, low-latency performance. Additional barriers include unresolved regulatory and medico-legal issues, as well as surgeon acceptance and trust. Procedure-specific visual challenges – such as excess intraabdominal fat, variable exposure, and smoke or blood contamination – further complicate reliable intraoperative computer-vision analysis.

3.3.3 Future directions

Future work should focus on developing large, multicenter bariatric video datasets with standardized labeling and on building robust phase-recognition and anatomy-segmentation models for SG, Roux-en-Y gastric bypass, and revisional procedures. Integration of multimodal operating room data may further enhance intraoperative safety monitoring. Prospective validation, clearly defined performance endpoints, and regulatory approval will be required before routine clinical adoption.

3.4 AI in endobariatric and endoscopic procedures

3.4.1 Overview and evidence scope

Endoscopic bariatric therapies – including ESG, intragastric balloons, TORe, and POSE – offer minimally invasive alternatives to surgery. Current AI applications remain limited and predominantly ESG-focused, addressing outcome prediction [30], reintervention risk [31], computer vision-based skill assessment [32], cross-procedure VTE modeling [31], and endoscopic stent outcomes [33]. Small cohorts, retrospective designs, and a lack of prospective validation constrain most studies [27, 34]. Notably, no dedicated ML models currently exist for intragastric balloons, TORe, or POSE as primary procedures.

3.4.2 Computer vision guidance

Computer vision efforts in endobariatric procedures primarily focus on simulated skill classification rather than real-time intraoperative guidance [32]. Dials et al. applied SMOTE augmentation and multiple classifiers (SVM, AdaBoost, KNN, RF) and reported very high accuracies for expert–novice discrimination in simulated ESG; however, these results were derived from small datasets without live clinical validation [32]. Subsequent work on stomach location and task classification similarly lacked real-time deployment [32]. To date, validated systems for real-time anatomical landmarking, instrument tracking, or automated safety alerts remain unavailable [27, 34].

3.4.3 ESG outcome and complication prediction

Vannucci et al. analyzed 404 ESG cases across 12 centers and predicted $\geq 25\%$ total weight loss or $\geq 50\%$ excess weight loss at 12 months [30]. Models based solely on preoperative variables showed poor discrimination (AUROC 0.50–0.60), whereas the incorporation of early follow-up data substantially improved performance, achieving AUROC 0.79–0.81 at 3 months and 0.87 (TWL) and 0.86 (EWL) on 6-month external validation [30].

Ali et al. developed the ESG Reintervention Score using LASSO-LR in 3,583 patients, achieving an AUROC of 0.74 for predicting 30-day reintervention. Key predictors included prior foregut surgery, anticoagulation, female sex, ASA $\geq$ II and hypertension [31]. A separate ML-based VTE model demonstrated moderate discrimination in ESG compared with RYGB, supporting shared risk factors while underscoring the need for procedure-specific calibration [31].

3.4.4 Endoscopic complication management

Hany et al. developed the Alexandria-Bari-Stent model to predict endoscopic stent failure for postbariatric leaks, achieving an external AUROC of 0.85 with balanced sensitivity and specificity. Predictors included obstructive sleep apnea, cardiometabolic comorbidities, BMI, and stent characteristics [33]. No existing models address primary endobariatric adverse events, including balloon migration, perforation, and early procedure-specific complications [31, 34].

3.4.5 Evidence gaps in endobariatric AI

Despite promising early ESG work, several gaps remain:

- Procedure concentration: most AI evidence centers on ESG; little to no ML work exists for intragastric balloons, TORe, POSE, or balloon-specific complications (migration, pancreatitis) [27, 34].
- Outcome diversity: models focus on weight loss and reintervention, neglecting quality of life and metabolic response [27, 34].
- Real-time guidance: no clinical trials exist for real-time endoscopic AI tools guiding suture placement, pattern recognition, or complication avoidance [27, 32, 34].
- Sample size and heterogeneity: ESG studies typically involve a few hundred patients, limiting generalizability [30].
- External validation: Vannucci et al. provide a vital exception; most work remains single-center and internally validated [27, 34].
- Skill integration: assessment models are proof-of-concept, not tied to training curricula and credentialing [32].

Endobariatric AI remains experimental – best utilized for research, hypothesis generation, and training advancement rather than routine clinical decision-making. Clinicians should track developments, critically appraise evidence, and participate in pilot studies and registry initiatives until large-scale validation confirms clinical utility.

To contextualize these gaps across procedures and application domains, the overall maturity of AI research in bariatric and endobariatric care is summarized in **Figure 2**.

	Preoperative Risk Prediction	Postoperative Surveillance	Intraoperative AI / Computer Vision	Outcome Prediction (Weight / Metabolic)	Ethical & Regulatory
Roux-en-Y Gastric Bypass (RYGB)	Mature	Emerging	Limited	Mature	Emerging
Sleeve Gastrectomy	Mature	Emerging	Limited	Mature	Emerging
Endoscopic Sleeve Gastroplasty (ESG)	Limited	Limited	Absent	Emerging	Absent
Intragastric Balloons	Absent	Absent	Absent	Absent	Absent
(TORe)	Absent	Absent	Absent	Absent	Absent
(POSE)	Absent	Absent	Absent	Absent	Absent

Figure 2.
AI research maturity by procedure. Heat map summarizing AI evidence across seven clinical domains and six bariatric and endobariatric procedures. Colors indicate maturity from minimal evidence to multicenter validation; study counts are shown, and stars denote external validation. The figure highlights mature AI research in surgical bariatrics compared with limited ESG-focused endobariatric evidence, with significant gaps in intragastric balloons, transoral outlet reduction, and other primary endoluminal procedures.

3.4.6 Summary

AI applications in endobariatric therapy remain in their early stages but show emerging promise, particularly in ESG outcome prediction, as shown in **Figure 3**, where incorporating early follow-up data yielded 12-month AUROCs of 0.86–0.87 [30]. Computer vision-based skill assessment is feasible but lacks clinical validation [32]. Broader application beyond ESG, integration of real-time guidance, and prospective multicenter validation are essential for clinical translation.

3.5 Long-term outcomes and weight loss prediction

3.5.1 Heterogeneity of long-term outcomes

Long-term outcomes after bariatric surgery are highly variable. Although most patients achieve substantial weight loss and metabolic improvement within one to two years, 20–30% experience weight regain by five years [35], and diabetes remission rates range widely (40–80%) depending on patient and procedural factors [36]. Accurate individual-level prediction could improve preoperative counseling, procedure selection, and postoperative management.

Comparative Performance: AI/ML vs. Traditional Risk Prediction Models in Bariatric and Endobariatric Surgery

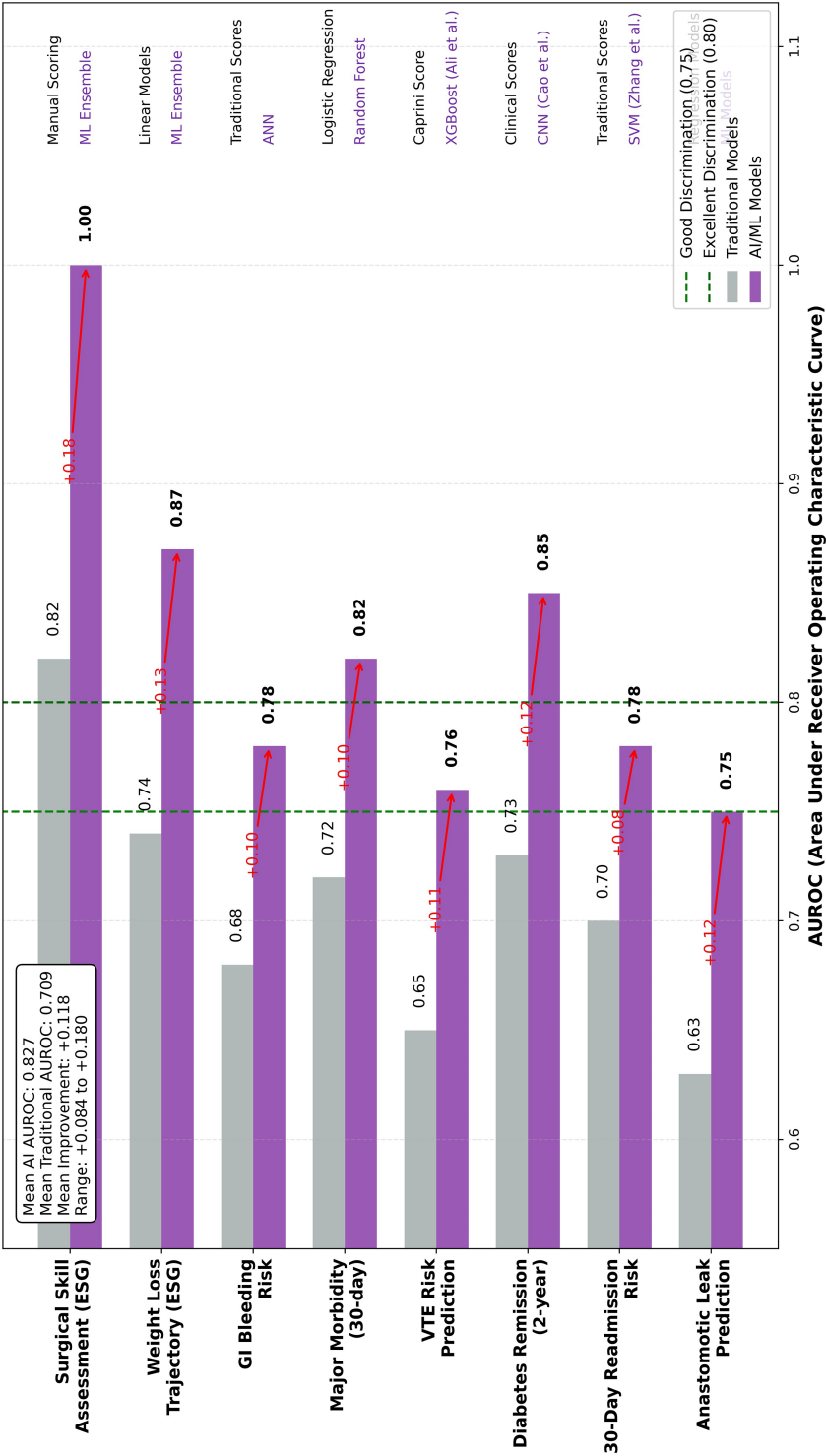


Figure 3. AI/ML versus traditional risk models. Horizontal bar chart comparing AUROC performance of AI/ML (purple) and traditional models (gray) across eight bariatric and endobariatric outcomes. Red arrows mark clinically meaningful gains ($\Delta\text{AUROC} \geq 0.10$), with reference lines for good (0.75) and excellent (0.80) discrimination. Overall, AI/ML shows a mean AUROC improvement of + 0.118, with model types and representative studies indicated.

3.5.2 Interpretable weight trajectory prediction: The SOPHIA study

The SOPHIA study by Saux et al. represents one of the most rigorous efforts to predict long-term outcomes. Using data from 10,231 patients across multiple European centers, the authors developed an interpretable ML model combining LASSO feature selection with regression-tree ensembles to predict 5-year BMI trajectories. Key predictors included baseline BMI, age, sex, diabetes status, and procedure type [37].

The model underwent multinational external validation, achieving MAD 2.8 kg/m² and RMSE 4.7 kg/m², indicating clinically useful accuracy for counseling. A web-based calculator was also developed, making SOPHIA a benchmark example of externally validated, interpretable, and clinically deployable ML in bariatric surgery.

3.5.3 Deep learning for diabetes remission

Using SOReg, Cao et al. developed a 7-layer CNN to predict diabetes remission after Roux-en-Y gastric bypass. The model achieved AUROC $\approx$ 0.85 in validation and 0.83 on external testing, outperforming established clinical scores by 9–12% [11]. While demonstrating the potential of deep learning to capture complex metabolic interactions, the model's limited interpretability constrains direct clinical adoption.

3.5.4 Other predictive approaches

Smaller exploratory studies have reported mixed results. Nadal et al. achieved 71.4% accuracy in classifying one-year total weight loss < 30% after Roux-en-Y gastric bypass in a limited cohort (n = 118), demonstrating feasibility but requiring larger-scale validation [38].

3.5.5 Methodological challenges

Long-term outcome prediction is complicated by loss to follow-up [35], heterogeneous outcome definitions (%TWL, %EWL, BMI change), varying time horizons, and infrequent calibration assessment despite widespread reporting of AUROC [37]. Moreover, clinical utility depends on whether predictions can meaningfully guide actionable interventions, not accuracy alone.

3.5.6 Summary and clinical implications

ML models show clinically meaningful performance for predicting long-term weight and metabolic outcomes, as shown in **Figure 3**, particularly when externally validated across large, multinational cohorts. Interpretable models, such as SOPHIA, currently offer the best balance among accuracy, transparency, and usability, while deep learning approaches provide greater discrimination for specific outcomes, such as diabetes remission, at the cost of interpretability. Future priorities include identifying modifiable risk factors, integrating behavioral and patient-reported data, and conducting implementation studies to determine whether outcome prediction tools improve shared decision-making and patient-centered outcomes.

3.6 Ethical, regulatory, and implementation considerations

3.6.1 Algorithmic bias and health equity

ML models trained on historical data may perpetuate health disparities through underrepresentation of racial, socioeconomic, or geographic groups, or by inheriting biases arising from inequitable access to bariatric surgery and perioperative care [39, 40]. Kang et al. reported that LR demonstrated the most favorable racial fairness in predicting postoperative complications, while undersampling improved fairness in boosted tree models. Routine fairness assessments (e.g., equalized odds, demographic parity), together with bias-mitigation strategies and established ethical frameworks [41], are essential for the equitable implementation of bariatric AI.

3.6.2 Explainability and interpretability

“Black box” ML models, particularly deep neural networks, can achieve high predictive accuracy but offer limited transparency, which hinders clinicians' trust and clinical adoption. Increasingly, regulatory and medicolegal frameworks require explainability in clinical decision support systems.

Explainable AI (XAI) approaches seek to address this gap. Zucchini et al. applied SHAP values to quantify feature contributions in complication prediction, enabling case-level interpretation of risk drivers such as low albumin and advanced age [17]. Similarly, the SOPHIA study emphasized interpretability through regression-tree ensembles and a transparent web-based calculator [37].

Despite these advances, the clinical impact of XAI remains uncertain. Whether explanations improve clinician trust, influence decision-making, or enhance outcomes – and how explanations should be tailored to different users – requires further human factors research and prospective evaluation.

3.6.3 Regulatory pathways and clinical deployment

The regulatory landscape for AI-based clinical decision support is evolving. In the United States, FDA guidance addresses software as a medical device (SaMD) [42] and adaptive algorithms, while in the European Union, the Medical Device Regulation (MDR) [43] and the emerging AI Act mandate requirements for safety, efficacy, and transparency.

Few AI tools in bariatric surgery have achieved regulatory approval or routine clinical deployment; most remain retrospective research prototypes. Translation to clinical use requires prospective validation, usability testing, EHR integration, postdeployment monitoring, and clear evidence of clinical utility (improved outcomes, reduced costs, or enhanced patient satisfaction).

Bellini et al. highlighted persistent barriers to deployment, including data privacy, accountability for AI-generated recommendations, and the inadequacy of retrospective performance metrics in the absence of prospective real-world evidence [27].

3.6.4 Practical implementation barriers

Beyond regulation, real-world adoption faces several pragmatic obstacles [44].

1. Workflow integration: tools must deliver predictions at the right time, in the correct format, within the platforms clinicians already use.
2. Clinician trust and training: surgeons and endoscopists must understand the model's scope, limitations, and appropriate use.
3. Reimbursement and business models: without clear financial incentives, health systems may be slow to invest in development and maintenance.
4. Model maintenance: data distributions and practice patterns evolve; AI models require periodic updates and recalibration.
5. Digital divide: concentrating AI deployment in highly resourced centers risks widening disparities.

Emerging reporting frameworks, such as TRIPOD-AI, and risk-of-bias tools, such as PROBAST-AI, are beginning to set higher standards for model development, validation, and reporting in exactly these contexts.

3.6.5 Data privacy, security, and federated learning

AI deployment in bariatric care often requires aggregating sensitive patient data across institutions and jurisdictions, raising significant privacy and security concerns [45]. Privacy-preserving methods, including federated learning – which enables model training without centralizing raw data – and differential privacy, offer potential solutions but remain technically complex and resource-intensive [46]. As such, large-scale, routine clinical implementation of these approaches is still largely aspirational.

3.6.6 Implementation and equity in low- and middle-income countries (LMICs)

In LMICs, where bariatric services are expanding amid constrained resources, AI presents both strategic opportunities and substantial risks.

Potential benefits include optimizing patient selection for limited operating room and endoscopy capacity, enabling telemedicine-supported follow-up, and leveraging regional registries for benchmarking and quality improvement.

However, significant barriers persist, including limited digital infrastructure and interoperability, fragmented and incomplete data capture, scarcity of local data science expertise, and underdeveloped regulatory and governance frameworks.

Successful AI screening programs for diabetic retinopathy and tuberculosis in LMICs demonstrate feasibility despite these constraints [47, 48]. In contrast, bariatric AI applications remain largely undeveloped. Future progress will require equitable high- and low-resource collaborations, capacity building, and privacy-preserving approaches, such as federated learning, to ensure a sustainable, nonextractive implementation.

3.6.7 Summary and recommendations

Ethical, regulatory, and implementation considerations are as important as technical performance for the successful clinical integration of AI in bariatric surgery. Key recommendations include routine assessment of algorithmic fairness across demographic groups with active bias-mitigation strategies; prioritization of interpretability through XAI to support clinician trust and accountability; early engagement with regulatory authorities to navigate SaMD and adaptive AI pathways; and prospective implementation studies to evaluate real-world impact, usability, and unintended consequences. The development and adoption of standardized reporting frameworks, such as TRIPOD-AI for surgical AI, are essential to improve transparency and reproducibility. Finally, meaningful involvement of diverse stakeholders – including patients, surgeons, ethicists, and policymakers – should guide the design, evaluation, and deployment of bariatric AI systems.

4. Discussion

4.1 Summary of key findings

This narrative literature review synthesized AI and ML applications across the bariatric and metabolic surgery care continuum. Across five thematic domains, substantial research activity has yielded promising results but also highlighted methodological limitations and implementation barriers.

- **Preoperative risk prediction:** neural networks and gradient boosting outperform LR (AUROC gains of 0.02–0.12) for complications like anastomotic leak and VTE, but low sensitivity for rare events and limited external validation limit utility [11, 14–16].
- **Postoperative surveillance:** ML achieves moderate to good readmission prediction (AUROC 0.75–0.80) using dynamic labs; mobile health integration shows promise but lacks prospective validation [19, 20].
- **Intraoperative computer vision** in bariatric surgery is still limited compared to other specialties; no clinically deployed systems for real-time guidance exist [25–27, 29].
- **Endobariatric AI** is focused on ESG, with 12-month weight-prediction models in a single multicenter study achieving AUROCs of 0.86–0.87 when early follow-up data are integrated [30]. Still, evidence for other endoscopic modalities is limited and based on cohorts of only a few hundred patients.
- **Long-term outcome models**, such as SOPHIA, demonstrate clinically useful accuracy for 5-year BMI trajectories, and deep learning improves diabetes remission prediction relative to traditional scores [11, 37].

- **Ethical, regulatory, and implementation risks:** algorithmic bias, explainability gaps, regulatory hurdles, and workflow barriers persist; no widely deployed, approved bariatric AI tools exist [39, 42–46].

4.2 Surgical versus endoscopic AI maturity

Surgical bariatric AI is more advanced due to access to large, validated registries (MBSAQIP, SOReg), enabling robust model development and external validation [11, 14–16, 37]. In contrast, endoscopic AI – primarily focused on ESG – relies on smaller, single-center or simulation-based cohorts, prioritizing technical feasibility over generalizability [30–32, 49].

Methodologically, surgical studies predominantly use structured, multiinstitutional registry data [11, 14–16], whereas endoscopic studies are constrained by limited case volumes and heterogeneous techniques [30–32, 49].

Implementation challenges differ: surgical AI must integrate with operating room logistics and perioperative pathways [14, 15], while endoscopic AI must address dynamic video analysis, operator variability, and device heterogeneity [32, 49]. Cross-domain collaboration in video analysis and skill assessment may accelerate convergence.

4.3 Comparison with traditional approaches

A key issue is whether AI/ML models provide sufficient advantage over traditional statistical approaches to justify their complexity, data demands, and implementation costs. As shown in **Figure 3**, most studies report modest AUROC improvements of approximately 0.02–0.12 compared with conventional risk models [11, 14–16]. While statistically significant, the clinical relevance of these gains is highly context-dependent. For rare but catastrophic complications, such as anastomotic leak, even minor improvements in discrimination may support targeted surveillance or preventive strategies. In contrast, for common outcomes with established and effective treatments, incremental predictive gains may offer limited practical benefit.

Traditional risk scores (e.g., MBSAQIP, DiaRem) [50] remain attractive due to their simplicity, transparency, and broad accessibility, whereas most AI/ML models require specialized computational infrastructure and technical expertise. The SOPHIA study represents a pragmatic exception, combining improved performance with interpretability, external validation, and a user-friendly interface, and provides a benchmark for clinically deployable AI in bariatric surgery [37].

4.4 Methodological challenges and reporting standards

Despite promising performance, methodological limitations continue to hinder the clinical translation of AI models in bariatric surgery. These challenges largely reflect data constraints and study designs typical of surgical research but can be mitigated through standardized approaches.

Class imbalance is a significant limitation, as severe postoperative complications occur in only 1–5% of cases, biasing models toward nonevents and inflating apparent accuracy [2]. Techniques such as SMOTE oversampling, class weighting, and threshold optimization can improve sensitivity, but robust mitigation ultimately requires larger, multicenter datasets [8, 13].

External validation remains insufficient. Models validated internally often overestimate performance, with AUROC declines observed during cross-institutional testing due to data heterogeneity [14, 16]. Registry-based collaborations, such as MBSAQIP, are essential for assessing generalizability across diverse bariatric populations [4].

Reporting standards are inconsistent, with variable outcome definitions and an overreliance on discrimination metrics alone. Adoption of TRIPOD-AI and PROBAST-AI frameworks promotes transparency, bias assessment, and reproducibility, particularly when hyperparameters and calibration methods are fully reported [9, 10].

Prospective validation is rare, as most studies remain retrospective and prone to hindsight bias. Prospective trials comparing AI-assisted decision-making with standard care are necessary to establish clinical utility, safety, and regulatory readiness [7, 12, 42].

Finally, **calibration** is frequently overlooked. At the same time, AUROC reflects discrimination, and poor alignment between predicted and observed risks limits decision support. Calibration plots and Brier scores should routinely accompany performance reporting [11, 21, 37].

To improve transparency, comparability, and methodological rigor in bariatric and endobariatric AI research, adoption of standardized reporting frameworks – such as TRIPOD and its AI extension (TRIPOD-AI) – along with formal risk-of-bias assessment using PROBAST and PROBAST-AI, is essential [9]. Future studies should prospectively register protocols, clearly describe development and validation cohorts, and report results in TRIPOD-AI-compliant formats with systematic bias evaluation. In addition to discrimination metrics, calibration plots and decision-curve or net benefit analyses should be presented to determine whether predictive gains translate into meaningful clinical benefit. Collectively, these practices will support more robust evidence generation and responsible clinical integration of AI models.

4.5 Clinical implications, ROI, and future directions

When rigorously developed and responsibly implemented, AI has the potential to transform the bariatric care continuum by enabling personalized risk stratification, optimizing procedure selection, enhancing early detection of postoperative complications, and supporting objective assessment of technical performance for training and quality improvement. From a health-system perspective, these capabilities translate into reduced readmissions and emergency visits, improved operating room utilization, more efficient allocation of multidisciplinary resources, and strengthened benchmarking for accreditation and quality programs. Realizing this value requires large, multiinstitutional datasets, prospective implementation trials evaluating clinical impact and cost-effectiveness, fairness and equity audits, and explainable, clinician-centered designs that integrate seamlessly into clinical workflows. Additional priorities include endoscopy-specific and real-time guidance tools, clear regulatory pathways developed through interdisciplinary co-design, and context-appropriate deployment strategies for low- and middle-income settings, including federated learning approaches.

4.6 Review limitations

This narrative review has several limitations. First, the search strategy, while structured and comprehensive, was not exhaustive and did not follow systematic review methodology (e.g., PRISMA). Relevant studies may have been missed, and publication bias (preferential publication of positive results) likely influences the literature. Second, a formal risk-of-bias assessment was not performed; methodological quality varies substantially across the included studies. Third, the rapid pace of AI research means that new studies emerge continuously; this review reflects the state of knowledge as of late 2024/early 2025. Fourth, thematic synthesis involves subjective interpretation; alternative organizational frameworks could yield different emphases. Finally, this review focused on clinical applications; technical details of ML algorithms, computational infrastructure, and data preprocessing were not exhaustively covered.

5. Conclusion

AI and ML are poised to transform bariatric and metabolic surgery, offering improved risk prediction, enhanced surveillance, and personalized outcome forecasting. Current evidence demonstrates statistically significant improvements in predictive performance over traditional approaches, particularly for predicting preoperative complication risk and long-term weight trajectory. However, considerable challenges persist: low sensitivity for rare events, limited external validation, sparse intraoperative applications, concerns about algorithmic bias, and the absence of regulatory and implementation pathways.

The field stands at a critical juncture. To advance from promising prototypes to clinically deployed, evidence-based tools, the bariatric surgery and AI communities must prioritize rigorous prospective validation, standardized reporting (TRIPOD-AI), fairness and interpretability assessments, and interdisciplinary collaboration. With sustained methodological rigor, AI can enhance surgical safety, optimize patient outcomes, and advance bariatric surgery practice.

Acronyms and abbreviations

AI	Artificial Intelligence
ML	Machine Learning
AUROC	Area Under the Receiver Operating Characteristic Curve
BMI	Body Mass Index
CI	Confidence Interval
CNN	Convolutional Neural Network
ESG	Endoscopic Sleeve Gastropasty
EWL	Excess Weight Loss
LMICs	Low- and Middle-Income Countries
LR	Logistic Regression
MBSAQIP	Metabolic and Bariatric Surgery Accreditation and Quality Improvement Program
NLP	Natural Language Processing

OR	Operating Room
PROBAST-AI	Artificial Intelligence extension of PROBAST
RF	Random Forest
RYGB	Roux-en-Y Gastric Bypass
SG	Sleeve Gastrectomy
SHAP	Shapley Additive exPlanations
SMOTE	Synthetic Minority Oversampling Technique
SOPHIA	Sorbonne/Paris-Lille International Study on Bariatric Surgery
SOReg	Scandinavian Obesity Surgery Registry
SVM	Support Vector Machine
TORe	Transoral Outlet Reduction
TRIPOD-AI	Artificial Intelligence extension of TRIPOD
TWL	Total Weight Loss
VTE	Venous Thromboembolism
XGBoost	Extreme Gradient Boosting

Acknowledgments

The author thanks Prof. Rodolfo Oviedo, Academic Editor, for his guidance, and the IntechOpen Publishing Team for their editorial support. Appreciation is extended to Al-Makassed Hospital and Al-Quds University for their academic and clinical environment. The author acknowledges the use of Grammarly for language polishing of the manuscript and ChatGPT for assistance in figure generation. No external funding was received for this work.

Conflict of Interest

The author declares no conflict of interest related to the content of this chapter. No financial relationships, commercial affiliations, or external influences affected the development or interpretation of this work.

Author details

Mohammed Maree
Consultant General and Laparoscopic Surgeon, Al-Makassed Hospital, Jerusalem,
Palestine

*Address all correspondence to: mareem@almakassed.org

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Obesity and overweight [Internet]. Available from: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight> [Accessed: 2025-December-28]
- [2] Kim JC, Kim MG, Park JK, Lee S, Kim J, Cho YS, et al. Outcomes and adverse events after bariatric surgery: An updated systematic review and meta-analysis, 2013–2023. *Journal of Metabolic and Bariatric Surgery* 2023;**12**:76
- [3] Reja D, Zhang C, Sarkar A. Endoscopic bariatrics: Current therapies and future directions. *Translational Gastroenterology and Hepatology*. 2022;**7**:21–21
- [4] Flum D, Belle S, King W, Wahed A. Perioperative safety in the longitudinal assessment of bariatric surgery. *New England Journal of Medicine*. 2009;**361** (5):445–454
- [5] Arterburn DE, Telem DA, Kushner RF, Courcoulas AP. Benefits and risks of bariatric surgery in adults: A review. *JAMA*. 2020;**324**(9):879
- [6] Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*. 2019;**25**(1):44–56
- [7] Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: Promises and perils. *Annals of Surgery*. 2018;**268**(1):70–76
- [8] Enodien B, Taha-Mehlitz S, Saad B, Nasser M, Frey DM, Taha A. The development of machine learning in bariatric surgery. *Frontiers in Surgery*. 2023;**10**:1102711
- [9] 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**: e048008
- [10] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*. 2019;**110**:12–22
- [11] Cao Y, Näslund I, Näslund E, Ottosson J, Montgomery S, Stenberg E. Using a convolutional neural network to predict remission of diabetes after gastric bypass surgery: Machine learning study from the Scandinavian Obesity Surgery Register. *JMIR Medical Informatics*. 2021;**9**(8):e25612
- [12] Elzayyat M, Kermansaravi M, Fakhro J, Kassir R. Artificial intelligence in bariatric surgery: Optimizing personalized decision-making, predictive monitoring, and postoperative outcomes. *Obesity Surgery*. 2025;**35**(10):4531–4533
- [13] Hassan Mukhtar MA, Babiker Ahmed AU, Siddig Mohammed MA, Ibrahim Omer NO, Altom DS, Elnour MAA. The role of artificial intelligence in the prediction of bariatric surgery complications: A systematic review. *Cureus*. 2025;**17**(4):e82461
- [14] Nudel J, Bishara AM, De Geus SWL, Patil P, Srinivasan J, Hess DT, et al. Development and

validation of machine learning models to predict gastrointestinal leak and venous thromboembolism after weight loss surgery: An analysis of the MBSAQIP database. *Surgical Endoscopy* 2021;**35**:182–191

[15] Wise ES, Amateau SK, Ikramuddin S, Leslie DB. Prediction of thirty-day morbidity and mortality after laparoscopic sleeve gastrectomy: Data from an artificial neural network. *Surgical Endoscopy*. 2020;**34**(8): 3590–3596

[16] Cao Y, Fang X, Ottosson J, Näslund E, Stenberg E. A comparative study of machine learning algorithms in predicting severe complications after bariatric surgery. *Journal of Clinical Medicine*. 2019;**8**(5):668

[17] Zucchini N, Capozzella E, Giuffrè M, Mastronardi M, Casagrande B, Crocè SL, et al. Advanced non-linear modeling and explainable artificial intelligence techniques for predicting 30-day complications in bariatric surgery: A single-center study. *Obesity Surgery* 2024;**34**:3627–3638

[18] Torquati M, Mendis M, Xu H, Myneni AA, Noyes K, Hoffman AB, et al. Using the super learner algorithm to predict risk of 30-day readmission after bariatric surgery in the United States. *Surgery* 2022;**171**:621–627

[19] Zhang M, Chen R, Yang Y, Sun X, Shan X. Machine learning analysis of lab tests to predict bariatric readmissions. *Scientific Reports*. 2024;**14**(1):16845

[20] Butler LR, Chen KA, Hsu J, Kapadia MR, Gomez SM, Farrell TM. Predicting readmission after bariatric surgery using machine learning. *Surgery for Obesity and Related Diseases*. 2023;**19**(11):1236–1244

[21] Farinella E, Papakonstantinou D, Koliakos N, Maréchal MT, Poras M, Pau L, et al. Integrating machine learning and dynamic digital follow-up for enhanced prediction of postoperative complications in bariatric surgery. *Obesity Surgery* 2025;**35**: 2202–2209

[22] Hsu JL, Chen KA, Butler LR, Bahraini A, Kapadia MR, Gomez SM, et al. Application of machine learning to predict postoperative gastrointestinal bleed in bariatric surgery. *Surgical Endoscopy* 2023;**37**:7121–7127

[23] Shan X, Zhang M, Liang Y, Yang Y, Xiao X, Chen R, et al. Application of machine learning to predict postoperative nausea and vomiting in laparoscopic sleeve gastrectomy. *Obesity Surgery* 2025;**35**:2663–2672

[24] Kdr L, Tay SBP, Choy KT, Verjans J, Sasanelli N, Kong JCH. Applications of natural language processing tools in the surgical journey. *Frontiers in Surgery*. 2024;**11**:1403540

[25] Maier-Hein L, Vedula SS, Speidel S, Navab N, Kikinis R, Park A, et al. Surgical data science for next-generation interventions. *Nature Biomedical Engineering* 2017;**1**: 691–696

[26] Kitaguchi D, Takeshita N, Hasegawa H, Ito M. Artificial intelligence-based computer vision in surgery: Recent advances and future perspectives. *Annals of Gastroenterological Surgery*. 2022;**6** (1):29–36

[27] Bellini V, Valente M, Turetti M, Del Rio P, Saturno F, Maffezzoni M, et al. Current applications of artificial intelligence in bariatric surgery. *Obesity Surgery* 2022;**32**:2717–2733

- [28] Madani A, Grover K, Watanabe Y. Measuring and teaching intraoperative decision-making using the visual concordance test: Deliberate practice of advanced cognitive skills. *JAMA Surgery*. 2020;**155**(1):78
- [29] Mascagni P, Vardazaryan A, Alapatt D, Urade T, Emre T, Fiorillo C, et al. Artificial intelligence for surgical safety: Automatic assessment of the critical view of safety in laparoscopic cholecystectomy using deep learning. *Annals of Surgery* 2022;**275**:955–961
- [30] Vannucci M, Niyishaka P, Collins T, Rodríguez-Luna MR, Mascagni P, Hostettler A, et al. Machine learning models to predict success of endoscopic sleeve gastropasty using total and excess weight loss percent achievement: A multicentre study. *Surgical Endoscopy* 2024;**38**:229–239
- [31] Ali H, Patel P, Malik TF, Pamarthy R, Mohan BP, Asokkumar R, et al. Endoscopic sleeve gastropasty reintervention score using supervised machine learning. *Gastrointestinal Endoscopy* 2023;**98**:747–754.e5
- [32] Dials J, Demirel D, Sanchez-Arias R, Halic T, Kruger U, De S, et al. Skill-level classification and performance evaluation for endoscopic sleeve gastropasty. *Surgical Endoscopy* 2023;**37**:4754–4765
- [33] Hany M, Elmongui E, El Ansari W. Development and external validation of a machine learning-based risk score for stent outcomes in post-bariatric leak management: The “Alexandria-Bari-Stent. *Obesity Surgery*. 2025;**35** (12):5120–5135
- [34] Bektaş M, Reiber BMM, Pereira JC, Burchell GL, Van Der Peet DL. Artificial intelligence in bariatric surgery: Current status and future perspectives. *Obesity Surgery*. 2022;**32**(8):2772–2783
- [35] Hatami M, Pazouki A, Hosseini-baharanchi FS, Kabir A. Bariatric surgeries, from weight loss to weight regain: A retrospective five-years cohort study. *Obesity Facts*. 2023;**16**(6):540–547
- [36] Jans A, Näslund I, Ottosson J, Szabo E, Näslund E, Stenberg E. Duration of type 2 diabetes and remission rates after bariatric surgery in Sweden 2007–2015: A registry-based cohort study. Pietiläinen KH, editor. *PLOS Medicine*. 2019;**16**(11):e1002985
- [37] Saux P, Bauvin P, Raverdy V, Teigny J, Verkindt H, Soumphonphakdy T, et al. Development and validation of an interpretable machine learning-based calculator for predicting 5-year weight trajectories after bariatric surgery: A multinational retrospective cohort SOPHIA study. *The Lancet Digital Health* 2023;**5**:e692–702
- [38] Nadal E, Benito E, Ródenas-Navarro AM, Palanca A, Martínez-Hervas S, Civera M, et al. Machine learning model in obesity to predict weight loss one year after bariatric surgery: A pilot study. *Biomedicines* 2024;**12**:1175
- [39] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;**366** (6464):447–453
- [40] Kang DW, Zhou S, Torres R, Chowdhury A, Niranjana S, Rogers A, et al. Predicting serious postoperative complications and evaluating racial fairness in machine learning algorithms for metabolic and bariatric surgery. *Surgery for Obesity and Related Diseases* 2024;**20**:1056–1064

- [41] Char DS, Shah NH, Magnus D. Implementing machine learning in health care — addressing ethical challenges. *New England Journal of Medicine*. 2018;**378**(11):981–983
- [42] Health C for D and R. Artificial intelligence in software as a medical device. FDA [Internet]. 2025 July 10. Available from: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device> [Accessed: 2025-December-27]
- [43] AI Act | Shaping Europe's digital future [Internet]. 2025. Available from: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> [Accessed: 2025-December-27]
- [44] Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*. 2019;**17**(1):195
- [45] Price WN, Cohen IG. Privacy in the age of medical big data. *Nature Medicine*. 2019;**25**(1):37–43
- [46] Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *Npj Digital Medicine* 2020;**3**:119
- [47] Cleland CR, Rwiza J, Evans JR, Gordon I, MacLeod D, Burton MJ, et al. Artificial intelligence for diabetic retinopathy in low-income and middle-income countries: A scoping review. *BMJ Open Diabetes Research & Care* 2023;**11**:e003424
- [48] Hwang EJ, Jeong WG, David PM, Arentz M, Ruhwald M, Yoon SH. AI for detection of tuberculosis: Implications for global health. *Radiology: Artificial Intelligence*. 2024;**6**(2):e230327
- [49] Dials J, Demirel D, Sanchez-Arias R, Halic T, De S, Gromski MA. Endoscopic sleeve gastroplasty: Stomach location and task classification for evaluation using artificial intelligence. *International Journal of Computer Assisted Radiology and Surgery*. 2024;**19**(4):635–644
- [50] Sjöholm K, Carlsson LMS, Taube M, Le Roux CW, Svensson PA, Peltonen M. Comparison of preoperative remission scores and diabetes duration alone as predictors of durable type 2 diabetes remission and risk of diabetes complications after bariatric surgery: A post hoc analysis of participants from the Swedish Obese Subjects Study. *Diabetes Care*. 2020;**43**(11):2804–2811