ANESTHESIOLOGY

Openventio PUBLISHERS

Review Article

Artificial Intelligence in Operating Room Workflow Optimization

Camille Ortega[1*], Lukas Marques[2] and Isabel Radcliffe[3]

[1]Department of Perioperative Medicine, Tyrrhenian University of Health Sciences, Italy
[2]Department of Anesthesiology and Pain Management, Tyrrhenian University of Health Sciences, Italy
[3]Department of Anesthesiology, Loire University of Medical Sciences, France

ABSTRACT

Operating room scheduling and throughput have traditionally relied on historical averages and clinical heuristics that perform poorly at predicting the outliers that drive costly delays and cancellations.

This review examines the emerging evidence for artificial intelligence applications across three domains of operating room workflow: machine learning models for surgical case-duration prediction, computer-vision and depth-sensing systems for real-time phase and milestone recognition, and anesthesia-specific predictive analytics for intraoperative hemodynamic and complication risk.

Case-duration prediction is the most mature of these applications, with multicenter machine learning models consistently outperforming surgeon-estimate and moving-average baselines, while video- and depth-based workflow recognition is advancing rapidly but remains constrained by privacy and data-governance considerations.

Anesthesia-specific decision support, particularly for intraoperative hypotension, illustrates both the promise of continuous physiological data streams and the unresolved question of whether earlier prediction reliably changes clinical outcomes rather than simply generating more alerts.

We conclude that the principal barrier to wider adoption is no longer algorithmic performance but implementation: interpretability sufficient for clinician trust, governance robust enough for sustained use, and integration tight enough to actually alter scheduling and intraoperative practice.

Keywords: *artificial intelligence; operating room management; surgical case duration; workflow recognition; perioperative decision support*

Citation: Ortega C, Marques L and Radcliffe I (2024) Artificial Intelligence in Operating Room Workflow Optimization. Anesthesiology. 4(1):01-05.

DOI: 10.17140/2024.vol4.i01.pp01-05

Corresponding Author: *Camille Ortega*

E-mail Address: *c.ortega@tuhs.it*

Article History: *Received 23 May 2024 | Accepted 18 June 2024 | Published 19 July 2024*

Copyright: *© 2024 Camille Ortega, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (**https://creativecommons.org/licenses/by/4.0/**).*

INTRODUCTION

The operating room is one of the most expensive spaces in any hospital to run and one of the hardest to schedule well. A single block hour left empty, or a case that overruns and cascades into three delayed follow-on cases, has financial and clinical consequences that ripple through the rest of the day. For decades, operating room management has leaned on statistical heuristics — historical averages,

surgeon-specific case-time estimates, simple queuing rules — to build the daily schedule and to decide when to call the next patient back. Those heuristics work reasonably well on average and poorly at the tails, which is exactly where the costly disruptions live. Artificial intelligence, and specifically the family of machine learning and computer vision methods that has matured over the past decade, offers a different proposition: models that learn directly from perioperative data — case logs, vital-sign streams, and now video of the room itself — to predict duration, detect workflow phases in real time, and flag anesthetic risk before it becomes an adverse event [1,2,3].

This review surveys where that promise has actually been tested. We look first at case-duration prediction, the most mature application and the one most directly tied to scheduling efficiency. We then turn to computer-vision-based workflow and phase recognition, which is beginning to move from research prototype to bedside deployment. We consider anesthesia-specific decision support, since anesthesiologists are the clinicians most continuously present in the room and therefore the most natural users of any real-time optimization tool. Finally we address the practical and ethical constraints — data governance, model interpretability, and the risk of automation complacency — that will determine whether these tools change practice or remain confined to pilot studies.

PREDICTING CASE DURATION: FROM HISTORICAL AVERAGES TO LEARNED MODELS

Accurate case-duration estimates are the foundation of operating room scheduling; overestimate and rooms sit idle, underestimate and the day runs late. Traditional scheduling software still often defaults to the surgeon's own historical average or to a simple moving average across the last several cases of the same type, an approach that ignores much of the information actually available in the electronic record. Machine learning models trained on structured case data — procedure code, surgeon, anesthesia type, patient comorbidity burden, time of day — have repeatedly outperformed these naive baselines in retrospective comparisons [4,5]. Gradient-boosted tree methods in particular have proven attractive for this task because they handle the mixed categorical and continuous features typical of perioperative datasets without extensive preprocessing, and because their feature-importance outputs give schedulers something interpretable to act on rather than an opaque number.

A multicenter validation using a large procedural database found that models trained across institutions generalized better than site-specific models trained on smaller local cohorts, suggesting that the signal driving case duration — patient factors, procedural complexity, team composition — is reasonably consistent across settings even when local practice patterns differ [6]. A scoping review of this literature confirmed the general pattern: most published models beat the "surgeon estimate" or moving-average baseline on mean absolute error, though the review authors noted that many studies were single-center, retrospective, and rarely tested prospectively in a live scheduling system, which limits how confidently the improvement can be generalized to actual block-time utilization [7]. Newer work has focused on interpretability specifically, using methods such as SHAP value decomposition to show schedulers which features are driving an unusually long or short prediction for a given case, an approach explicitly designed to build clinician trust rather than simply improve raw accuracy [8]. A recent prospective implementation study describes deploying such a model directly into a live scheduling workflow rather than testing it only retrospectively, one of the first steps toward closing the gap between predictive accuracy on paper and actual improvement in block utilization [9].

The practical payoff of better duration prediction is not simply a more accurate number on the schedule but a cascading effect: fewer late overruns into evening and night staffing, fewer canceled add-on cases, and more predictable turnover intervals between cases, which is itself the second major target of AI applications in the room.

COMPUTER VISION AND REAL-TIME WORKFLOW RECOGNITION

Where duration prediction works from data that already exists in the electronic record, a second and more recent line of work tries to extract workflow information directly from the operating room environment using cameras and depth sensors. Surgical phase recognition — automatically identifying which step of a procedure is underway from video — has been studied extensively in laparoscopic and endoscopic surgery, where the surgical video feed is already captured for other purposes, and a systematic review of AI models for surgical phase, instrument, and anatomical structure identification found substantial heterogeneity in model architectures and validation datasets but a consistent trend toward convolutional and transformer-based video models achieving clinically useful accuracy on phase segmentation tasks [10].

Extending this beyond the surgical field itself to the room as a whole is a harder and more recent problem, since it requires tracking not just the operative field but staff movement, equipment positioning, and case milestones such as patient-in-room, anesthesia induction, incision, and closure — the events that actually define turnover time and delay. Privacy has been the central obstacle to broad adoption of room-level video analytics, since continuous video recording of staff and patients raises consent and data-governance concerns that do not arise with

anonymized structured data. Depth-sensing cameras, which capture spatial and motion information without recording identifiable RGB video, have been proposed as a privacy-preserving alternative, and a recent evaluation of automated workflow recognition using exactly this kind of depth video demonstrated that milestone events could be detected with useful accuracy while avoiding the storage of identifiable footage [11]. This distinction matters for adoption: hospital privacy and IT governance committees are considerably more willing to approve depth-based or skeletal-tracking systems than conventional video recording, even when the underlying computer-vision architecture is similar.

Beyond phase and milestone detection, image-based AI systems integrated into control-tower-style dashboards — aggregating live status across many rooms simultaneously — have begun to show measurable operational effects. An implementation study in a high-acuity, fifteen-room cardiothoracic operating suite at a tertiary academic hospital reported an association between adoption of an image-based AI monitoring platform and an increase in case volume, attributed to earlier detection of turnover delays and more proactive coordination between the OR, sterile processing, and anesthesia teams [12]. Findings of this kind remain observational rather than randomized, and the effect of any single technology is difficult to separate from the broader process changes that typically accompany its rollout, but they are among the more concrete operational signals in the literature to date, and broader reviews of surgical efficiency interventions have reached similar conclusions about the direction, if not yet the definitive magnitude, of the effect [13].

ANESTHESIA-SPECIFIC DECISION SUPPORT AND PREDICTIVE ANALYTICS

Anesthesiologists occupy a distinctive position in operating room workflow: they are present continuously from induction to emergence, generate a dense stream of physiological data throughout the case, and are frequently the clinician best positioned to anticipate — and potentially prevent — the intraoperative events that most often derail a schedule, from hemodynamic instability to difficult emergence. A substantial and growing literature has applied machine learning to anesthesia-specific prediction tasks, most prominently intraoperative hypotension, using continuous arterial waveform or plethysmographic data to flag patients at elevated near-term risk before a clinically significant drop in blood pressure occurs [2,14]. Whether earlier warning reliably translates into fewer hypotensive episodes in practice, rather than simply more numerous alerts, remains an active area of investigation, and the answer likely depends heavily on how the warning is integrated into the clinician's existing workflow rather than on the underlying model's discriminative accuracy alone.

Beyond hemodynamic prediction, AI-based tools have been proposed across a wider range of perioperative decision points: risk stratification for postoperative complications, prediction of difficult airway or difficult intubation from preoperative imaging and structured risk factors, and estimation of postoperative recovery trajectory to guide discharge planning and PACU bed allocation, all of which feed back into how efficiently the broader surgical schedule can run [15,16]. A recent narrative framework proposal has argued for organizing these disparate applications along the axis of the data modality involved — physiological signal, medical image, or structured record — as a way of clarifying which validation standards and regulatory pathways apply to each, since a waveform-based hypotension predictor and an imaging-based airway assessment tool face very different evidentiary bars despite both being labeled "AI in anesthesia" [17]. Advocates for a hybrid decision-making model have similarly argued that these tools function best not as autonomous decision-makers but as a structured second opinion layered onto the anesthesiologist's own clinical judgment, particularly in time-pressured intraoperative moments where a purely algorithmic recommendation would be difficult to trust without transparent reasoning [18].

BARRIERS TO ADOPTION: INTERPRETABILITY, DATA GOVERNANCE, AND WORKFLOW FIT

The gap between a published model's reported accuracy and its actual clinical deployment is wide, and the operating room literature is unusually candid about why. Interpretability is a recurring theme: a black-box model that outputs a risk score or a duration estimate without an accompanying explanation is difficult for a busy clinician or scheduler to trust, particularly when the model is occasionally and inexplicably wrong. Work specifically aimed at bridging this gap has applied post-hoc explainability methods to anesthesiology prediction models, arguing that explainability is not a cosmetic add-on but a prerequisite for the kind of sustained clinical trust that turns a pilot deployment into routine use [19].

Data governance is the second recurring barrier, and it takes a different shape depending on the data modality. Structured case and vital-sign data raise conventional questions of data-sharing agreements and model retraining across institutions, which the multicenter duration-prediction literature has navigated with some success [6]. Video and depth-sensor data raise sharper questions of consent, since staff and sometimes patients are captured continuously in spaces where recording has not historically occurred, and institutions have generally moved more cautiously here, favoring privacy-preserving sensing approaches over conventional video recording even at some cost to model performance [11]. A structured strengths-weaknesses-opportunities-threats analysis

of AI in anesthesiology and perioperative medicine captured this tension well, noting that the same properties that make these tools powerful — continuous, granular, multimodal data capture — are precisely what make their governance difficult, and that institutions adopting AI tools piecemeal without a coherent data governance strategy risk building systems that cannot be maintained, validated, or extended once the initial pilot ends [20].

Finally, workflow fit determines whether even a well-validated, well-governed tool actually changes practice. A duration-prediction model that outputs a number nobody incorporates into the actual block schedule, or a hypotension alert that clinicians learn to silence because it fires too often, produces no operational benefit regardless of its underlying accuracy. The operating room management literature has increasingly emphasized implementation science alongside model development for exactly this reason, since the intervention that matters is not the algorithm in isolation but the combination of algorithm, interface, and institutional process into which it is placed [21,1].

CONCLUSION

Artificial intelligence has moved operating room workflow optimization from a field governed almost entirely by historical averages and clinical intuition toward one increasingly informed by models trained directly on perioperative data. The strongest evidence to date sits with case-duration prediction, where multiple independent groups have shown consistent, moderate improvements over traditional scheduling heuristics, and with anesthesia-specific hemodynamic prediction, where continuous physiological monitoring provides a natural substrate for early-warning models. Computer-vision-based workflow and milestone recognition is earlier in its trajectory but moving quickly, particularly as privacy-preserving sensing modalities make room-level monitoring more palatable to institutional review. What remains unresolved across all three domains is less a question of algorithmic capability than of implementation: interpretability sufficient to earn clinician trust, data governance robust enough to sustain multi-year deployment, and integration tight enough that a model's output actually changes what happens on the schedule rather than sitting unused in a dashboard. The next phase of this literature will need to look less like retrospective model validation and more like the prospective, workflow-embedded implementation studies that are only now beginning to appear.

DECLARATIONS

Acknowledgements

The authors thank their respective institutions for administrative support during the preparation of this review.

Funding

This work received no external funding.

Conflict of Interest

The authors declare that they have no conflicts of interest.

Author Contributions

C. Ortega conceived the review and drafted the initial manuscript. L. Marques performed the literature search and synthesis for the sections on computer vision and workflow recognition. I. Radcliffe drafted the sections on anesthesia-specific decision support and adoption barriers and critically revised the manuscript. All authors reviewed and approved the final version.

Ethics Statement

This is a literature review and did not involve new human or animal data; no ethics committee approval was required.

Data Availability

No new data were generated or analyzed in this review. All sources discussed are cited in the reference list.

REFERENCES

1. Emerging Technologies: Artificial Intelligence (2024) OpenAnesthesia.
2. Artificial Intelligence in Anesthesiology: Current Techniques, Clinical Applications, and Limitations (2020) Anesthesiology.
3. Applications of artificial intelligence in anesthesiology (2025) Anesthesiology and Perioperative Science.
4. Development of Predictive Model of Surgical Case Durations Using Machine Learning Approach (2025) Journal of Medical Systems.
5. Optimizing surgical efficiency: predicting case duration of common general surgery procedures using machine learning (2025) Surgical Endoscopy.
6. Machine Learning for the Prediction of Procedural Case Durations Developed Using a Large Multicenter Database: Algorithm Development and Validation Study (2023) JMIR AI.
7. Machine learning models to predict surgical case duration compared to current industry standards: scoping review (2023) PMC, PMC10630142.
8. Improving and Interpreting Surgical Case Duration Prediction with Machine Learning Methodology (2020) medRxiv.
9. Development and Implementation of a Machine Learning Model for Prediction of Surgical Case Duration (2025) PMID:42362434.
10. Artificial intelligence in the operating room: A systematic review of AI models for surgical phase, instruments and anatomical structure identification

(2025) PMC, PMC12575173.

11. Automated surgical workflow recognition in privacy-preserving depth videos of the operating room (2025) Surgical Endoscopy, PMC12408783.
12. Implementation of Image-Based Artificial Intelligence Is Associated with Increased Case Volume in a High-Acuity, 15-Room Cardiothoracic Operating Suite at a Tertiary Academic Hospital (2025) PMC, PMC13412611.
13. Enhancing Operating Room Surgical Efficiency Through Artificial Intelligence: A Comprehensive Review (2025) Springer Nature.
14. Artificial Intelligence for Perioperative Medicine (2022) Anesthesia & Analgesia.
15. Transforming perioperative care: The current landscape and future trajectory of artificial intelligence in anesthesia-A narrative review (2025) PMC, PMC13269965.
16. Artificial intelligence in predicting anesthetic complications: current techniques, clinical applications, and limitations (2025) PMID:42248065.
17. Precision perioperative AI: from signals, images, and records to applications in anesthesia-a narrative mini-review proposing an operational framework (2026) Frontiers in Medicine.
18. Advancing a Hybrid Decision-Making Model in Anesthesiology: Applications of Artificial Intelligence in the Perioperative Setting (2026) Healthcare, DOI:10.3390/healthcare14010097.
19. Explainable artificial intelligence in anesthesiology prediction models: bridging the gap from black-box algorithms to clinical decision support (2026) Frontiers in Medicine.
20. Strengths-weaknesses-opportunities-threats analysis of artificial intelligence in anesthesiology and perioperative medicine (2024) Frontiers in Digital Health.
21. Artificial Intelligence in Operating Room Management (2024) Journal of Medical Systems, PMID:38353755.