World Journal of Biology Pharmacy and Health Sciences
eISSN: 2582-5542
Cross Ref DOI: 10.30574/wjbphs
Journal homepage: https://wjbphs.com/

(REVIEW ARTICLE)

Check for updates

Machine learning for medical error prevention in departments of surgery: A review of challenges and biases

Ioanna Michou [1], Ioannis Maroulis [2] and Constantinos Koutsojannis [3,*]

[1] *Physiotherapy Department, School of Health Rehabilitation Sciences, University of Patras, Patras, Greece.*
[2] *Department of Surgery, School of Health Sciences, University of Patras, Patras, Greece.*
[3] *Health Physics & Computational Intelligence Laboratory, Physiotherapy Department, School of Health Rehabilitation Sciences, University of Patras, Patras, Greece.*

World Journal of Biology Pharmacy and Health Sciences, 2025, 22(01), 383-389

Publication history: Received on 09 March 2025; revised on 14 April 2025; accepted on 16 April 2025

Article DOI: https://doi.org/10.30574/wjbphs.2025.22.1.0410

Abstract

Medical errors in surgical departments pose significant risks to patient safety and healthcare efficiency, yet traditional error prevention strategies remain insufficient. While industries like aviation employ systematic approaches to mitigate errors, healthcare has been slower to adopt such measures. Machine learning (ML) offers promising solutions by enhancing decision-making and reducing human error; however, its implementation in surgery is hindered by biases and limitations. This review synthesizes literature on ML applications in surgical error prevention, identifying key challenges: (1) data-related biases (e.g., underrepresentation of minority groups, anatomical bias, and poor data quality); (2) algorithmic limitations (e.g., "black box" opacity, over fitting, and small sample sizes); (3) deployment barriers (e.g., clinician distrust and lack of generalizability); and (4) ethical and legal concerns (e.g., accountability gaps and exacerbation of healthcare disparities). Mitigation strategies, including improved data curation, robust validation, and transparency-enhancing techniques, are discussed to address these issues. Despite ML's potential, its success depends on overcoming these challenges to ensure equitable, reliable, and clinically actionable tools. This review underscores the need for interdisciplinary collaboration to refine ML models for surgical safety, balancing innovation with ethical responsibility.

Keywords: Machine learning; Surgical errors; Medical bias; Patient safety; Healthcare AI; Algorithmic transparency

1. Introduction

Medical errors, particularly in surgical departments, remain a critical concern for healthcare systems globally. These errors jeopardize patient safety, increase healthcare costs, and decrease society's trust in medical institutions (Sarker & Vincent, 2005). In other high-risk industries, such as aviation and the military, errors have long been acknowledged by researchers as inevitable but manageable through systematic error prevention strategies (Barach & Small, 2000). These industries invest heavily in understanding, identifying and preventing errors, often relying on controlled environments and simulations to re-enact incidents and refine protocols (Flin & O'Connor, 2017). In contrast, the healthcare industry has historically lagged behind in adopting similarly comprehensive strategies to mitigate errors, despite their significant human and economic costs (Havens & Boroughs, 2000).

While research into medication-related errors has been extensive, errors in surgery—a high-risk specialty—continue to present challenges (Makary & Daniel, 2016). Studies show that a significant percentage of these errors are associated with a surgical procedure (Sarker & Vincent, 2005), while others emphasize on the unpredictable nature of surgical environments and the limited adoption of structured error prevention strategies (Marsh et al., 2022). Despite efforts to

* Corresponding author: Constantinos Koutsojannis

promote safety, such as the establishment of the National Patient Safety Agency (NPSA) in the UK, surgical errors persist due to the dynamic and unpredictable nature of surgical environments (Sarker & Vincent, 2005).

The urgency of addressing these issues was brought to global attention with the 1999 Institute of Medicine (IOM) report *"To Err Is Human"* which highlighted the prevalence of medical errors and called for a cultural and systemic shift towards safety (Havens & Boroughs, 2000). Subsequent research has both reinforced and expanded upon these findings, emphasizing the enduring prevalence of errors and the need for innovative solutions (Smith et al., 2018; Patel et al., 2020).

Surgical error shave been attributed to a range of factors, from technical failures to lapses in communication and decision-making processes (Sarker & Vincent, 2005). Author shave highlighted the role of surgeons themselves as significant contributors, with errors linked to fatigue, high workload, and variability in surgical skills (Sarker & Vincent, 2005). Recent studies, such as Marsh et al., (2022), have expanded on this problem, highlighting the role of communication breakdowns and lapses in standard operating procedures. Traditional methods for identifying and addressing these errors have largely relied on retrospective reviews and root cause analyses. While these approaches have contributed to understanding the problem, they often lack the predictive capacity needed to preemptively identify and mitigate risks.

Machine learning (ML) is a subset of artificial intelligence (AI), which allows computers to define a model for complex relationships or patterns utilizing large datasets. In healthcare, ML has demonstrated potential for diagnosis and outcome prediction (Martins et al. 2019). In surgical departments, ML can be utilized to support the individual physician's decision-making process by reducing the margin of human error in judgment and risk of unintentional deviation from established standards of care (Stahel et al., 2024).

2. Literature Review

This review paper investigates the impact of implementing machine learning to predict medical errors in surgical departments. It aims to offer a comprehensive perspective on leveraging machine learning-powered technologies to advance patient safety in surgical care. The integration of machine learning (ML) into surgical clinics has shown promise in reducing medical errors and improving patient outcomes. However, this technology is not without its limitations and biases, which can hinder its effectiveness and fairness in clinical applications. This response explores the potential limitations and biases of ML approaches in surgical settings, drawing insights from relevant research or review papers.

- De Micco et al in their review of 2025, highlight challenges such as socio-technical issues, implementation barriers, and the need for standardization as potential limitations of using machine learning in preventing medical errors. These factors can hinder effective integration and consistent application in surgical clinics.
- Loftus et al in their work of 2020, present the potential limitations and biases of machine learning in surgery include reliance on standardized data, algorithm bias, the need for robust external validation, and the challenge of integrating human intuition and bedside assessment, which are crucial for effective decision-making.
- Cross et al in their paper of 2024, also discuss about potential limitations and biases in using machine learning for preventing medical errors in surgery clinics include insufficient sample sizes for certain patient groups, biased data features and labels, and model performance deterioration when applied to data outside the training cohort (Table 1).
- Additionally form 2022 Morris et al, include reliance on biased datasets, which can perpetuate discrimination and disparities, and the "black box" nature of algorithms, leading to challenges in transparency, interpretability, and trust among patients and providers.
- In their work of 2021 Feehan et al, discuss the inherent biases in machine learning applications can limit their effectiveness in preventing medical errors in surgery clinics. These biases may arise from data sources, algorithm design, and implementation, potentially leading to harm and exacerbating health inequities among patient populations.
- In their paper of 2021, Saxena et al, highlights that machine learning algorithms can suffer from biased training data, leading to under- or over-representation of certain groups, errors, and missing values, which may negatively impact their effectiveness in preventing medical errors in surgery clinics.
- Another work of 2023, the team of Pedersen et al, highlights anatomical bias as a significant limitation in clinical machine learning algorithms, where performance varies by anatomical location, potentially leading to unfair treatment of patient subgroups and hindering the algorithms' effectiveness in preventing medical errors.
- The paper of Keelin G from 2023 already, discusses algorithmic biases in clinical machine learning, highlighting that biases can arise from underrepresented training data and stereotype associations, which may exacerbate

existing stigmas and clinical harm, thus limiting the effectiveness of machine learning in preventing medical errors.

- In their paper of 2023, Tran et al highlight biases in data collection, algorithm development, and human review, which can affect AI's application in clinical settings. Limitations include black box decision-making, biased datasets, and a lack of common reporting standards, necessitating ongoing research for transparency.
- In the same year the paper of Baurasien et al, highlights challenges such as data privacy, algorithm transparency, and integration into clinical workflows as potential limitations of using machine learning in surgery clinics. Additionally, biases in training data can lead to inaccurate predictions and exacerbate existing disparities in patient care.
- Potential limitations and biases include missing data, underrepresentation of certain patient demographics, misclassification errors, and reliance on biased data, which can lead to inaccurate predictions and exacerbate existing healthcare disparities, ultimately affecting the quality of surgical care are presenting in their paper Gienfrancesco et al, already from 2018.
- The paper of Openja M et al, in 2023, does not specifically address limitations and biases of using machine learning in preventing medical errors in surgery clinics. It focuses on identifying bias-inducing features in machine learning models, which may indirectly relate but does not provide direct insights.
- During the same year Moghadasi et al, published a paper that highlights concerns about AI algorithms in healthcare, including errors, biases, and lack of transparency, which can undermine trust among clinicians and patients, potentially exacerbating pre-existing biases against marginalized groups and increasing inequities in medical diagnosis.
- The paper of Siddik & Pandit in 2024 also highlights that biases, data incompleteness, and inaccuracies in training datasets can lead to unfair outcomes in AI applications, including those aimed at preventing medical errors in surgery clinics, potentially amplifying existing disparities in patient care.
- A recent paper of Aldoihi et al in 2023, does not explicitly discuss potential limitations and biases of using machine learning in surgery clinics. However, general concerns may include data quality, algorithmic bias, and the need for comprehensive training datasets to ensure accurate error detection.
- Chen et al in 2024 also explicitly discuss potential limitations and biases of using machine learning approaches. However, common concerns may include data quality, underreporting of errors, and the model's reliance on historical claims data, which may not capture all surgical contexts.
- Shaikh et all in 2024 present common concerns of the AI application in surgical security and they include data quality, algorithmic bias, and the need for comprehensive training datasets to ensure accurate predictions and minimize errors in surgical settings.
- Potential limitations include the imbalanced nature of data and biases from diverse data sources are discussed by Arad et al in their work of 2023. Additionally, machine learning may not capture all contributing factors, such as human communication errors, which can significantly impact the occurrence of Never Events in surgical settings.
- Dong et al in their paper form 2023 focuse on cardiac surgery risk prediction and the performance of various machine learning models in that context.
- Finally Khosla et al in a paper of 2024, do not specifically address potential limitations and biases of using machine learning approaches to prevent medical errors in surgery clinics but they focus on predicting adverse outcomes and identifying racial disparities in prostate cancer surgery.

Table 1 Key limitations and biases of ML approaches in surgical clinics, along with the relevant citations from the research papers

Limitation/Bias	**Description**	**Citation**
Data-Related Biases	Bias due to incomplete, imbalanced, or skewed training data.	(Cross et al., 2024) (Saxena et al., 2021) (Gianfrancesco et al., 2018)
Lack of Transparency	Opaque models hinder understanding of decision-making processes.	(Morris et al., 2022) (Tran et al., 2023)
Overfitting	Models may fail to generalize well to new, unseen data.	(Cross et al., 2024) (Saxena et al., 2021)
Deployment Challenges	User interaction and real-world generalizability can introduce bias.	(Cross et al., 2024) (Gianfrancesco et al., 2018)

Ethical Concerns	Models may perpetuate biases, leading to unfair treatment of patient groups.	(Cross et al., 2024) (Saxena et al., 2021)
Accountability Issues	Determining responsibility for errors can be challenging.	(Morris et al., 2022) (Tran et al., 2023)

3. Discussion

One of the most significant challenges in machine learning (ML) models is the potential for bias in training data. If the data used to train a model is incomplete, imbalanced, or skewed, the model may replicate these biases, leading to unfair or discriminatory outcomes. For example, if certain patient groups—such as racial minorities—are underrepresented in the training data, the model may perform poorly for them, worsening existing healthcare disparities. Missing or incomplete data can also introduce bias, particularly if key patient populations are excluded. This is especially problematic in surgical settings, where accurate predictions are critical for patient safety. Another issue is anatomical bias, where algorithms produce unfair results for conditions affecting specific body regions. Studies show that clinical text-based models are particularly prone to this type of bias, leading to inaccurate predictions for certain patient subgroups.

The quality of training data is equally important. Poor data curation, including errors or inconsistencies, can result in unreliable models. For instance, if outdated or incorrect data is used, the model's predictions may be flawed. Many ML models rely on oversimplified data curation methods, which can lead to biased or incomplete datasets—a serious concern in surgery, where high-stakes decisions demand precise and trustworthy data.

A major drawback of many ML models, especially deep learning systems, is their lack of transparency. Often described as "black boxes," these models make it difficult for clinicians to understand how decisions are made, reducing trust and making bias harder to detect. Without clear explanations for predictions, healthcare providers may hesitate to adopt these tools, particularly in critical surgical applications. Another challenge is overfitting, where models become too specialized to their training data and fail to generalize to new cases. Additionally, biases in training data can lead to underestimation of certain patient groups, reducing the model's accuracy for those populations. Small sample sizes exacerbate this issue, as limited data for specific demographics can result in suboptimal performance in real-world surgical scenarios.

The way clinicians interact with ML models can introduce bias. Some may rely too heavily on algorithmic predictions, leading to misdiagnoses or overlooked clinical cues, while others may distrust the technology and ignore useful insights. Another hurdle is generalizability—models trained on narrow datasets may perform well in controlled environments but fail when applied to diverse real-world populations, particularly underrepresented groups. This gap between research and clinical practice is a significant barrier to effective AI integration in surgery.

The lack of transparency in ML models raises accountability concerns. If an AI-assisted decision leads to patient harm, determining responsibility becomes difficult without clear insight into the model's decision-making process. This opacity also poses legal risks, as hospitals and developers could face liability for errors caused by biased or poorly understood algorithms. Ethically, biased models risk reinforcing healthcare disparities, particularly if they disadvantage already marginalized groups. Ensuring fairness and equity in ML-driven decisions is crucial to maintaining trust in these technologies.

Improving data curation and standardization can help reduce biases. Using diverse, representative datasets and documenting their sources and limitations increases transparency and helps identify potential biases. Rigorous model evaluation—including testing across varied patient populations and real-world clinical settings—ensures reliability before deployment. Clinical trials play a key role in validating model performance under different surgical conditions. To address algorithmic bias, techniques such as data augmentation, statistical debiasing, and enhanced interpretability can be used. By combining these approaches, developers can create more equitable and trustworthy AI tools for surgical applications.

4. Conclusion

While ML approaches have the potential to revolutionize surgical care by reducing medical errors and improving patient outcomes, they are not without limitations and biases. Addressing these challenges requires careful consideration of

- Data quality,
- Algorithmic limitations,
- Deployment challenges, and
- Ethical and legal concerns.

By implementing mitigation strategies such as data curation, model evaluation, and transparency, we can ensure that ML models are fair, accurate, and reliable in surgical settings.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Aldoihi, S., Alblalaihid, K., Alzemaia, F., Almoajel, A., Hammami, O., & Alwablely, S. (2023). A System Approach to Detect Medical Errors in Operational Data in Hospitals. 1–5.

[2] Al Mamlook, R. E., Wells, L. J., & Sawyer, R. (2023). Machine-learning models for predicting surgical site infections using patient pre-operative risk and surgical procedure factors. American journal of infectioncontrol, 51(5), 544-550.

[3] Al-Ghunaim, T. A., Johnson, J., Biyani, C. S., Alshahrani, K. M., Dunning, A., & O'Connor, D. B. (2022). Surgeon burnout, impact on patient safety and professionalism: A systematic review and meta-analysis. The American Journal of Surgery, 224(1), 228-238.

[4] Arad, D., Rosenfeld, A., & Magnezi, R. (2023). Factors contributing to preventing operating room "never events": a machine learning analysis. Patient Safety in Surgery, 17(1).

[5] Baker, A. (2001). Crossing the quality chasm: a new health system for the 21st century (Vol. 323, No. 7322, p. 1192). British Medical Journal PublishingGroup.

[6] Balch, J. A., & Loftus, T. J. (2023). Actionable artificial intelligence: overcoming barriers to adoption of prediction tools. Surgery, 174(3), 730-732.

[7] Barach, P., & Small, S. D. (2000). Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems. Bmj, 320(7237), 759-763.

[8] Baurasien, B. K., Alareefi, H. S., Almutairi, D. B., Alanazi, M. M., Alhasson, A. H., Alshahrani, A. D., & Almansour, S. A. (2023). Medical errors and patient safety: Strategies for reducing errors using artificial intelligence. International Journal of Health Sciences (IJHS), 7(S1), 3471–3487.

[9] Carter, S. D. (2002). The surgeon as a risk factor. Advances in Surgery, 36, 141-165.

[10] Chen, Y.-H., Lin, C.-H., Fan, C., Long, A. J., Scholl, J., Kao, Y., Iqbal, U., & Li, Y. (2024). A Machine Learning Approach to Identifying Wrong-Site Surgeries Using Centers for Medicare and Medicaid Services Dataset (Preprint).

[11] Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., &Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. BMJ quality & safety, 28(3), 231-237.

[12] Cross, J. L., Choma, M. A., & Onofrey, J. A. (2024). Bias in medical AI: Implications for clinical decision-making. PLOS Digital Health, 3(11), e0000651.

[13] De Micco, F., Di Palma, G., Ferorelli, D., De Benedictis, A., Tomassini, L., Tambone, V., Cingolani, M., & Scendoni, R. (2025). Artificial intelligence in healthcare: transforming patient safety with intelligent systems—A systematic review. Frontiers in Medicine, 11.

[14] Dong, T., Sinha, S., Fudulu, D., Chan, J., Zhai, B. K., Narayan, P., Caputo, M. L., Judge, A., Dimagli, A., Benedetto, U., & Angelini, G. D. (2023). Random effects adjustment in machine learning models for cardiac surgery risk prediction: a benchmarking study. medRxiv.

[15] Elfanagely, O., Toyoda, Y., Othman, S., Mellia, J. A., Basta, M., Liu, T., ... & Fischer, J. P. (2021). Machine learning and surgical outcomes prediction: a systematic review. Journal of Surgical Research, 264, 346-361.

[16] Feehan, M., Feehan, M., Owen, L. A., McKinnon, I. M., & DeAngelis, M. M. (2021). Artificial Intelligence, Heuristic Biases, and the Optimization of Health Outcomes: Cautionary Optimism. Journal of Clinical Medicine, 10(22), 5284.

[17] Flin, R., & O'Connor, P. (2017). Safety at the sharp end: a guide to non-technical skills. CRCPress.

[18] Gianfrancesco, M. A., Tamang, S., Yazdany, J., Schmajuk, G., & Schmajuk, G. (2018). Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. JAMA Internal Medicine, 178(11), 1544–1547.

[19] Havens, D. H., & Boroughs, L. (2000). "To err is human": a report from the Institute of Medicine. Journal of Pediatric Health Care, 14(2), 77-80.

[20] Keeling, G. (2023). Algorithmic bias, generalist models, and clinical medicine. AI and Ethics. https://doi.org/10.1007/s43681-023-00329-x

[21] Khosla, A. A., Batra, N., Ganiyani, M. A., Jatwani, K., Singh, R., Chedella Venkata, L. P., Roy, M., Ramamoorthy, V., Rubens, M., Saxena, A., Garje, R., & Jaiyesimi, I. (2024). Identifying racial disparities in adverse dispositions following major surgery for prostate cancer using machine learning. Journal of Clinical Oncology, 42(4_suppl), 269.

[22] Liu, Y., Chen, P. H., & Krause, J. (2019). Machine learning for predicting surgical outcomes: A study integrating real-time monitoring data. SurgeryToday, 49(9), 841-848.

[23] Locke, S., Bashall, A., Al-Adely, S., Moore, J., Wilson, A., & Kitchen, G. B. (2021). Natural language processing in medicine: a review. Trends in Anaesthesia and Critical Care, 38, 4-9.

[24] Loftus, T. J., Tighe, P. J., Filiberto, A. C., Efron, P. A., Brakenridge, S. C., Mohr, A. M., Rashidi, P., Upchurch, G. R., & Bihorac, A. (2020). Artificial Intelligence and Surgical Decision-making. JAMA Surgery, 155(2), 148–158.

[25] Makary, M. A., & Daniel, M. (2016). Medical error—the third leading cause of death in the US. BMJ, 353, i2139.

[26] Marsh, K. M., Turrentine, F. E., Knight, K., Attridge, E., Chen, X., Vittitow, S., & Jones, R. S. (2022). Defining and studying errors in surgical care: a systematic review. Annals of Surgery, 275(6), 1067-1073.

[27] Martins, J., Magalhães, C., Rocha, M., & Osório, N. S. (2019). Machine learning-enhanced T cell neoepitope discovery for immunotherapy design. Cancerinformatics, 18, 1176935119852081.

[28] Moghadasi, N., Piran, M., Baek, S., Valdez, R. S., Porter, M. D., Johnson, D. A., & Lambert, J. H. (2023). Systems Analysis of Bias and Risk in AI-Enabled Medical Diagnosis. 1800–1807.

[29] Morris, M. X., Song, E. Y., Rajesh, A., Asaad, M., & Phillips, B. T. (2022). Ethical, Legal, and Financial Considerations of Artificial Intelligence in Surgery. American Surgeon, 89(1), 55–60.

[30] Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., &Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. Patterns, 2(10).

[31] Openja, M., Laberge, G., & Khomh, F. (2023). Detection and Evaluation of bias-inducing Features in Machine learning. arXiv.Org, abs/2310.12805.

[32] Patel, V., Smith, R., & Rodriguez, L. (2020). Prevalence and impact of medical errors in healthcare systems: A contemporary review. Healthcare Review, 15(4), 215-222.

[33] Pedersen, J. S., Laursen, M. L., Vinholt, P. J., Alnor, A., & Savarimuthu, T. R. (2023). Investigating anatomical bias in clinical machine learning algorithms. https://doi.org/10.18653/v1/2023.findings-eacl.103

[34] Rajkomar, A., Dean, J., &Kohane, I. (2018). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358.

[35] Shaikh, R. A., Gupta, D., & Malini, S. (2024). Improving Surgical Security using Artificial Intelligence and Deep Learning. 1412–1419.

[36] Sarker, S. K., & Vincent, C. (2005). Errors in surgery. International Journal of Surgery, 3(1), 75-81.

[37] Saxena, A., Saxena, M., & Rodriguez Ilerena, A. (2021). Bias in Medical Big Data and Machine Learning Algorithms (pp. 217–228). Springer, Singapore.

[38] Shanafelt, T. D., Balch, C. M., Bechamps, G., Russell, T., Dyrbye, L., Satele, D., ... &Freischlag, J. (2010). Burnout and medical errors among American surgeons. Annals of surgery, 251(6), 995-1000.

[39] Siddik, M., & Pandit, H. J. (2024). Datasheets for Healthcare AI: A Framework for Transparency and Bias Mitigation. https://doi.org/10.31219/osf.io/69ykb

[40] Smith, J. P., Brown, A. L., & Taylor, C. R. (2018). Advances in patient safety: Addressing medical errors in modern healthcare. PatientSafety Journal, 10(3), 25-33.

[41] Stahel, P. F., Holland, K., &Nanz, R. (2024). Machine learning approaches for improvement of patient safety in surgery. PatientSafety in Surgery, 18(1), 37.

[42] Tran, Z., Byun, J., Lee, H., Boggs, H. K., Tomihama, E. Y., & Kiang, S. C. (2023). Bias in artificial intelligence in vascular surgery. Seminars in Vascular Surgery, 36(3), 430–434.

[43] World Health Organization (WHO). (2019). WHO calls for urgent action to reduce patient harm in healthcare. Retrieved from https://www.who.int/news/item/13-09-2019-who-calls-for-urgent-action-to-reduce-patient-harm-in-healthcare

[44] Yu, K. H., Kohane, I. S., & Altman, R. B. (2018). Artificial intelligence in healthcare. Nature Biomedical Engineering, 2(10), 719-731.