

EDITORIAL

Open Access



Machine learning approaches for improvement of patient safety in surgery

Philip F. Stahel^{1,2,3*}, Kathryn Holland¹ and Roy Nanz¹

Keywords Artificial intelligence, Machine learning, Predictive analytics, Patient safety in surgery

Traditional patient safety protocols are based on the historic principle of the Hippocratic oath intended to abstain from inflicting harm on our patients (*"Primum non nocere"*) [1]. In spite of this noble vision, patients frequently remain caught in the ‘friendly fire’ of preventable surgical complications and hospital-acquired adverse conditions, which are widely regarded as an unfortunate side-effect of modern healthcare [2]. Ironically, the plethora of stringent regulatory compliance-mandated protocols and globally disseminated patient safety checklists still fail to protect our patients from medical errors in the modern age of patient safety today [3]. Our continued quest towards “Goal Zero” for preventable harm requires a pragmatic reassessment of the historically failed approaches in the patient safety arena [4, 5]. It is time for a paradigm shift from the traditional approach of counting and analyzing medical errors, with the intent of preventing similar occurrences to harm a different patient in the future, to effectively predicting and preventing an adverse event before harm reaches the patient [6]. The new age of artificial intelligence allows for novel machine learning approaches to support predictive analytics tools which can identify

a patient at risk of sustaining harm before the adverse event occurs [7]. The latest special collection in *Patient Safety in Surgery* is dedicated to covering a variety of aspects pertinent to machine learning approaches for the improvement of surgical patient safety (www.biomedcentral.com/collections/MLPS).

A total of eight peer-reviewed articles were published in the collection from 2022 to 2024. A study from São Paulo University in Brazil [8] investigated the hypothesis that an AI approach would improve the quality and sensitivity of prostate cancer detection and grading in 6,982 image sections from 32 radical prostatectomy specimens (Fig. 1). The authors found that a “deep learning” approach for AI-supported analysis showed a high sensitivity and specificity for detection and grading of cancer stages beyond the baseline inter-pathologist variance in the the interpretation of human prostate cancer specimens [8]. A different study from Bar-Ilan University in Israel assessed the value of machine learning for predicting risk factors which contribute to “never events” in the operating, including wrong-site surgery and unintentionally retained foreign objects [9]. These results demonstrated impressively that selected variables had an impact of >900% on predicting the occurrence of a “never event” by machine learning analysis, including discrepancies in the second and third count of surgical instruments and sponges [9]. Conversely, standard elements in the surgical safety checklist appeared to have minimal impact or no impact at all on adverse event prediction (Fig. 2). The third paper in the collection

*Correspondence:

Philip F. Stahel

Philip.Stahel@HCAhealthcare.com

¹Mission Health, 509 Biltmore Ave., Asheville, NC 28801, USA

²Department of Surgery, East Carolina University, Brody School of Medicine, Greenville, NC, USA

³Department of Specialty Medicine, Rocky Vista University, College of Osteopathic Medicine, Parker, CO, USA



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

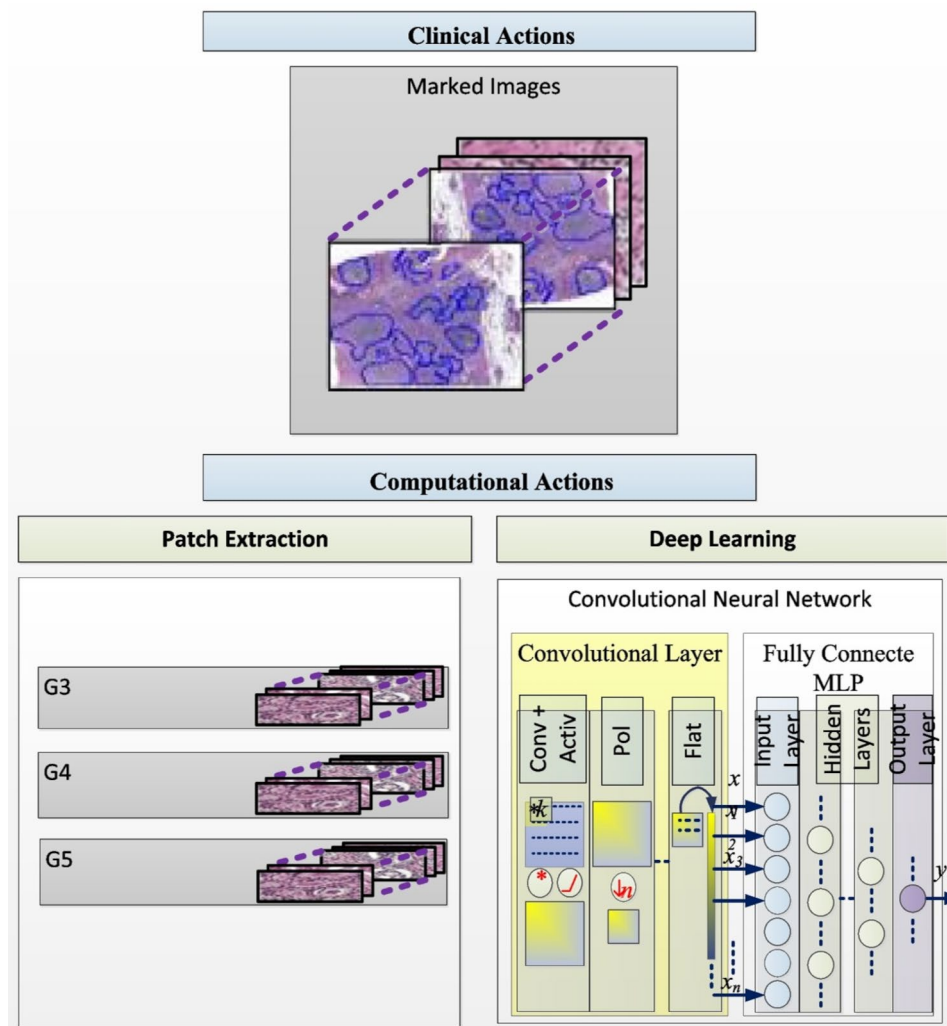


Fig. 1 Artificial intelligence approach for detection and grading of prostate cancer in human prostatectomy specimens. © Kudo MS, et al., 2022. Reproduced with permission by the Creative Commons Attribution 4.0 International License, from: Kudo MS, et al., The value of artificial intelligence for detection and grading of prostate cancer in human prostatectomy specimens: a validation study. *Patient Saf. Surg.* 2022, 16:36

represents a retrospective propensity-matched cohort study from a tertiary referral academic medical center in the United States [10]. The authors investigated the “Rothman Index” (RI) as a predictive analytics tool to predict and prevent unsafe patient downgrades with subsequent unplanned readmissions to the intensive care unit (ICU) [10]. The RI represents a real-time composite measure of medical acuity in hospitalized patients and provides a predictive analytics model for continuous monitoring of a patient’s clinical status and improvement or deterioration over time (Fig. 3). During a 12-months study time-window, a total of 5,261 ICU patients were included in analysis, of which 212 patients (4.0%) had an unplanned readmission to the ICU within 7 days [10]. The results from this study identified the RI as a sensitive predictor of unanticipated readmissions to ICU which were associated with significantly increased mortality and hospital length

of stay. The authors therefore concluded that the RI should be considered as a real-time objective measure for prediction of a safe downgrade from ICU to a lower level of care [10].

A narrative review article from Kermanshah University of Medical Sciences in Iran provided a compelling state-of-the-art overview on our current knowledge and clinical applications of machine learning technology and artificial intelligence for improving patient safety in spine surgery [11]. The authors provided intriguing insights on how predictive analytics can be leveraged to improve surgical outcomes, how spine surgery practices may design personalized care through machine learning approaches, and how peri-operative risks may be mitigated through artificial intelligence for patients undergoing spinal procedures [11]. Importantly, the paper also discussed the potential utility for artificial intelligence technologies to

Feature	Description	Operation Stage	Ophthalmology	General Surgery	Orthopedics	Cardiology	Urology	Gynecology
A	In case of discrepancy in the count of absorbable items, their package is taken out from the OR (verified out loud)	First Count	-2.78	11				
B	Surgery is paused because of discrepancy in third count	Third count	269	1125		1540	1283	
C	Discrepancy in second count	Second Count	269	1150	160	1950	1168	
D	Length of surgery 1-2 hours	General		-100				
E	Length of surgery >4 hours	General		577				
F	Surgeon scans the cavity/fascia before closure	Second Count	-69		-100	-65	-100	-100
G	Length of surgery <1 hour	General			128			
H	Second count is performed before closure of fascia/cavity	Second Count				-87	-81	
I	Procedure's type is compared to the one written patient's file (verified out loud)	Time out				-87		-100
J	Second count is performed before closure of fascia/cavity	Second Count					-81	
K	Closure of fascia/cavity begins after verification of correct count	Second Count	-77		-100			
L	Third count starts after removing all items from surgical field	Third Count			-100			
M	All staff members agreed to the details mentioned during time out (verified out loud)	Time out						-100
N	Small rural hospital	General						-100
O	Anesthesia consent forms are completed and signed (verified out loud)	Time out						-100

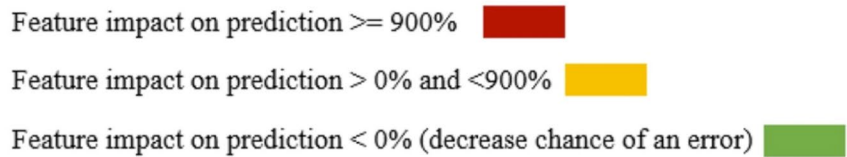


Fig. 2 Machine learning approach for prediction of “never events” in the operating room. © Arad D, et al., 2023. Reproduced with permission by the Creative Commons Attribution 4.0 International License, from: Arad D, et al., Factors contributing to preventing operating room “never events”: a machine learning analysis. *Patient Saf. Surg.* 2023, 17:6

optimize the shared decision-making process with surgical patients. The same research group further investigated the predictive capability of machine learning models in determining outcomes and complications in 329 patients undergoing cervical spine decompression and fusion surgery for spondylotic myelopathy [12]. The authors concluded that their machine learning model represented a highly sensitive predictive tool for surgical complications and postoperative outcomes in the vulnerable cohort of patients with cervical myelopathy [12]. With a similar intent, a prospective proof-of-concept pilot study from Goethe University in Frankfurt, Germany, investigated the “Prehab App” as a novel digital risk calculator to predict patient outcomes after major surgery [13]. The investigator-initiated Protego Maxima Trial showed that risk assessment using the “Prehab App” represented a sensitive preoperative predictor of surgical complications and postoperative mortality in a cohort of 267 patients who underwent major surgical procedures in visceral, thoracic, vascular, urological, and gynecological surgery [13].

The trauma research group at the University Hospital Zurich, Switzerland, pursued a new innovative approach by combining the knowledge from large data patient registries with machine learning algorithms, with the intent to elevate the synergy of the predictive capabilities of these individual research tools [14]. The authors combined multiple data platforms from local, regional, and national registries with various artificial intelligence approaches stratified from narrow, to general, to “super” intelligence, aimed at predicting and improving postsurgical outcomes in trauma patients [14]. Novel software technologies, such as the “TRAUMA Pathway Explorer” by IBM Watson Health, are furthermore discussed as available solutions for predicting trauma outcomes based on demographic and physiological patient variables (Fig. 4).

The last article in the special collection describes a deep learning-based computer visual model for the automated detection of surgical tools and instruments in the operating room [15]. The authors developed a dataset of based on 1,000 images from more than 13,000 surgical tools to feed a new artificial

RI Warning Level	RI Implications for Clinical Consideration	RI Implications for Discharge Planning
Rate of Change Alert ΔRI decline >20 points over the past 24 hours.	Re-assess patient and identify the underlying driver of decreasing RI.	
Very High Warning RI: <20 (or automated ΔRI decline warning)	RN: Re-assess patient; identify underlying driver of low RI; notify provider. Consider initiating rapid response. MD: Prioritize rounding, consider goals of care discussion/palliative care consultation, if appropriate.	Clarify goals of care with provider. Evaluate post-acute care options.
High Warning RI: 40-20 (or automated ΔRI decline warning)	RN: Re-assess patient; identify underlying driver of low RI; notify provider if change in RI is acute. MD: Prioritize rounding; consider transfer to higher level of care.	Confirm family and POA contact information. Evaluate advanced directives.
Medium Warning RI: 60-40 (or automated ΔRI decline warning)	Consider increased monitoring re-assessment.	Re-evaluate appropriate level of care.
Low Warning RI: 80-60	Consider transfer to lower level of care or discharge.	Address discharge planning barriers.
No Warning RI: 100-80	Execute on discharge plan.	

Fig. 3 Rothman Index (RI) grading system as a predictive tool for clinical decision-making and patient downgrade/discharge planning, as implemented at Mission Hospital, Asheville, NC. © Stahel PF, et al., 2024. Reproduced with permission by the Creative Commons Attribution 4.0 International License, from: Stahel PF, et al., The Rothman Index predicts unplanned readmissions to intensive care associated with increased mortality and hospital length of stay: a propensity-matched cohort study. *Patient Saf. Surg.* 2024, 18:10

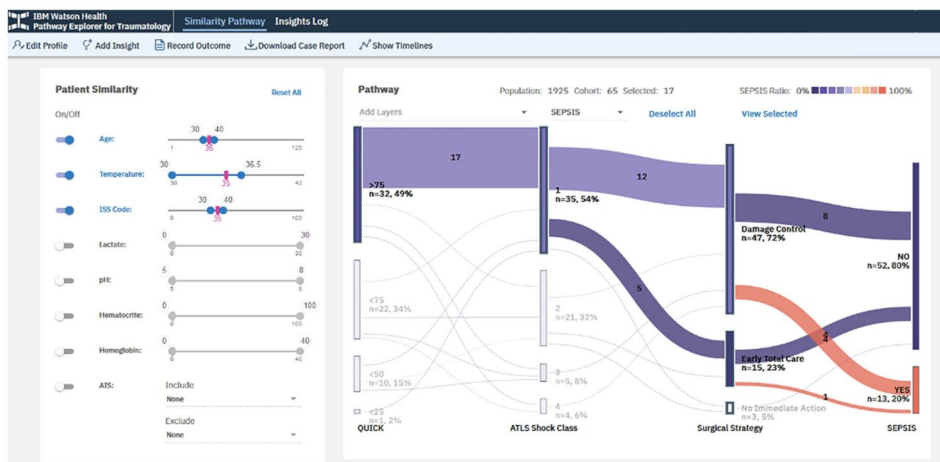


Fig. 4 “TRAUMA Pathway Explorer” by IBM Watson Health as an example of a novel software analytics tool designed to predict posttraumatic mortality and patient outcomes. © Pape HC, et al., 2024. Reproduced with permission by the Creative Commons Attribution 4.0 International License, from: Pape HC, et al., The role of big data management, data registries, and machine learning algorithms for optimizing safe definitive surgery in trauma: a review. *Patient Saf. Surg.* 2024, 18:22

intelligence model for optimization of the post-surgical instrument count accuracy and prevention of unintentionally retained foreign objects after surgery [15]. The deep learning-based computer vision model demonstrated a high accuracy up to 100% for distinguishing surgical tools from background, and a high sensitivity up to 98.2% for detecting and identifying overlapping instruments (Fig. 5). The authors concluded that their innovative deep learning-based visualization model

may provide a future artificial intelligence “safeguard” to prevent surgical “never events” from occurring and harming patients, e.g. by the unintentional retention of foreign objects, such as surgical instruments or lap sponges [16].

In summary, the featured articles in the new special collection provide a broad perspective on opportunities for improving patient safety in surgery via modern technologies based on machine learning and artificial

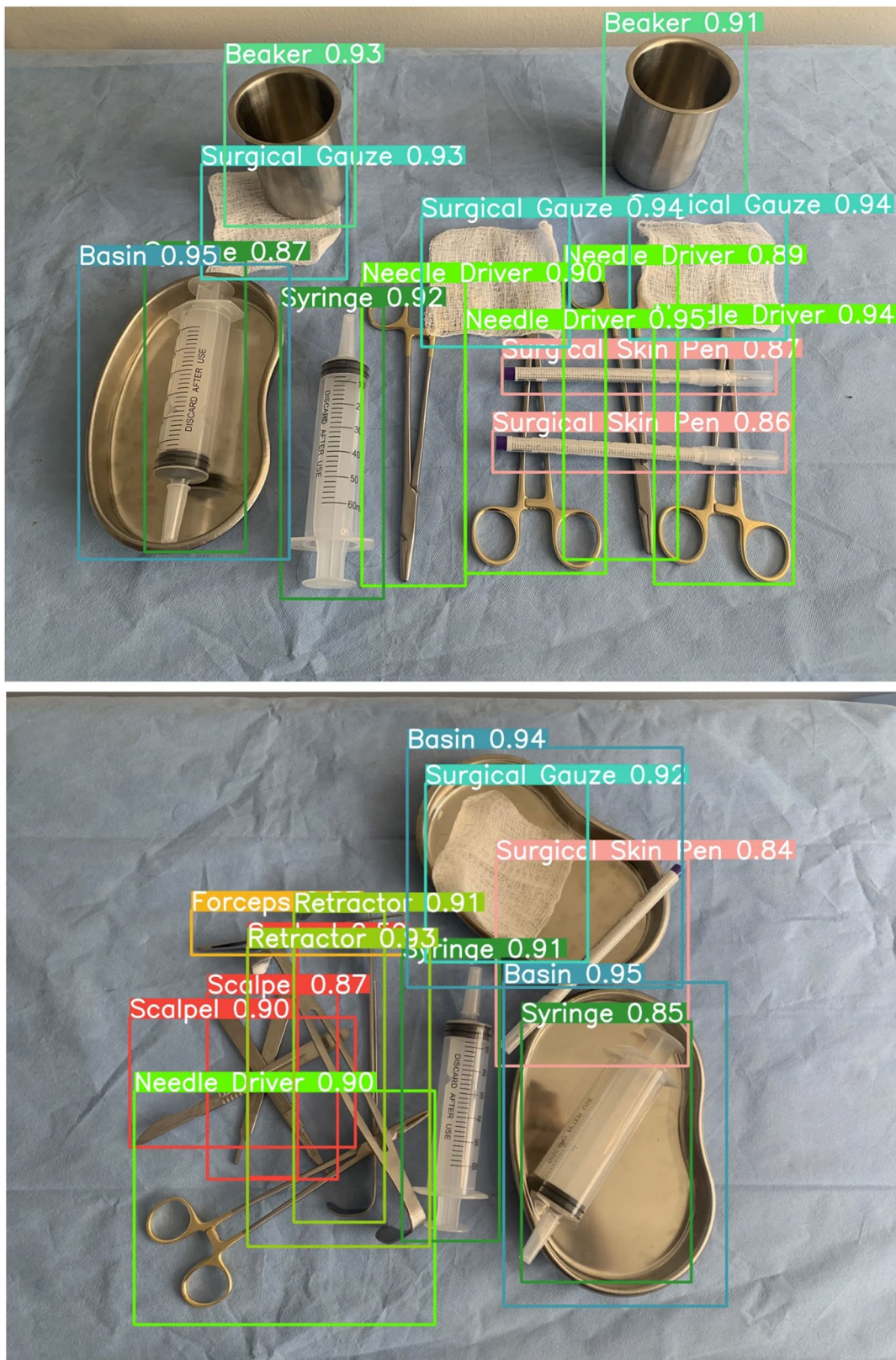


Fig. 5 Deep learning-based artificial intelligence computer vision model for automated detection of surgical tools and instruments for improved post-surgery count accuracy and prevention of unintentionally retained foreign bodies. © Deol ES, et al., 2024. Reproduced with permission by the Creative Commons Attribution 4.0 International License, from: Deol ES, et al., Artificial intelligence model for automated surgical instrument detection and counting: an experimental proof-of-concept study. *Patient Saf. Surg.* 2024, 18:24

intelligence. Ultimately, these novel tools can support, but will not replace, 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.

Acknowledgements

The authors would like to thank Sarah Whitelock (Springer Nature) for her continued support of the journal's operations and mission.

Author contributions

P.F.S. designed the editorial and wrote the first draft. K.H. and R.N. contributed to revisions of the content and approved the final version of the manuscript for publication.

Funding

There were no external funding sources for this editorial.

Data availability

Please contact the authors for data requests.

Declarations

Ethics approval, consent for participation and publication

Not applicable (Editorial).

Consent for publication

Not applicable (Editorial).

FDA clearance

Not applicable (Editorial).

Competing interests

The authors are employed by HCA Healthcare. The views expressed in this editorial exclusively represent the authors' personal perspective and do not necessarily represent official views of HCA Healthcare or any of its affiliated entities. P.F.S. is the journal's Editor-in-Chief (<https://psjournal.biomedcentral.com/about/editorial-board>). The authors declare no other conflicts of interest related to this editorial.

Published online: 20 December 2024

References

1. Amirthanayagam TD, Emery RJ. Primum non nocere: risk assessment for new surgical procedures. *J Shoulder Elb Surg.* 2014;23(10):1417–8.
2. Stahel PF, Blood. *Sweat & Tears: Becoming a Better Surgeon.* TFM Publishing, Shrewsbury, UK, 1st edition, 2016, pp. 296.
3. Stahel PF, Mehler PS. Medical emergency teams and rapid response triggers—the ongoing quest for the 'perfect' patient safety system. *Crit Care.* 2009;13(5):420.
4. Hanchanale V, Rao AR, Motiwala H, Karim OM. Wrong site surgery! How can we stop it? *Urol Ann.* 2014;6(1):57–62.
5. Stahel PF, Sabel AL, Victoroff MS, Varnell J, Lembitz A, Boyle DJ, Clarke TJ, Smith WR, Mehler PS. Wrong-site and wrong-patient procedures in the universal protocol era: analysis of a prospective database of physician self-reported occurrences. *Arch Surg.* 2010;145(10):978–84.
6. Stahel PF, Cobianchi L, Dal Mas F, Paterson-Brown S, Sakakushev BE, Nguyen C, Fraga GP, Yule S, Damaskos D, Healey AJ, Biffi W, Ansaloni L, Catena F. The role of teamwork and non-technical skills for improving emergency surgical outcomes: an international perspective. *Patient Saf Surg.* 2022;16:8.
7. Radaelli D, Di Maria S, Jakovski Z, Alempijevic D, Al-Habash I, Concato M, Bolcato M, D'Errico S. Advancing patient safety: the future of artificial intelligence in mitigating healthcare-associated infections: a systematic review. *Healthcare (Basel).* 2024;12(19):1996.
8. Kudo MS, Gomes de Souza VM, Neubarth Estivallet CL, Alves de Amorim H, Kim FJ, Ramos Moreira Leite K, Moraes MC. The value of artificial intelligence for detection and grading of prostate cancer in human prostatectomy specimens: a validation study. *Patient Saf Surg.* 2022;16:36.
9. Arad D, Rosenfeld A, Magnezi R. Factors contributing to preventing operating room never events: a machine learning analysis. *Patient Saf Surg.* 2023;17:6.
10. Stahel PF, Belk KW, McInnis SJ, Holland K, Nanz R, Beals J, Gosnell J, Ogundele O, Mastriani KS. The Rothman Index predicts unplanned readmissions to intensive care associated with increased mortality and hospital length of stay: a propensity-matched cohort study. *Patient Saf Surg.* 2024;18:10.
11. Arjmandnia F, Alimohammadi E. The value of machine learning technology and artificial intelligence to enhance patient safety in spine surgery: a review. *Patient Saf Surg.* 2024;18:11.
12. Alimohammadi E, Fatahi E, Abdi A, Reza Bagheri S. Assessing the predictive capability of machine learning models in determining clinical outcomes for patients with cervical spondylotic myelopathy treated with laminectomy and posterior spinal fusion. *Patient Saf Surg.* 2024;18:21.
13. Sliwinski S, Faqar-Uz-Zaman SF, Heil J, Mohr L, Detemble C, Dreilich J, Zmuc D, Bechstein WO, Becker S, Chun F, Derwich W, Schreiner W, Solbach C, Fleckenstein J, Filmann N, Schnitzbauer AA. Predictive value of a novel digital risk calculator to determine early patient outcomes after major surgery: a proof-of-concept pilot study. *Patient Saf Surg.* 2024;18:13.
14. Pape HC, Starr AJ, Gueorguiev B, Wanner GA. The role of big data management, data registries, and machine learning algorithms for optimizing safe definitive surgery in trauma: a review. *Patient Saf Surg.* 2024;18:22.
15. Deol ES, Henning G, Basourakos S, Vasdev RMS, Sharma V, Kavoussi NL, Karnes RJ, Leibovich BC, Boorjian SA, Khanna A. Artificial intelligence model for automated surgical instrument detection and counting: an experimental proof-of-concept study. *Patient Saf Surg.* 2024;18:24.
16. Lembitz A, Clarke TJ. Clarifying "never events" and introducing "always events". *Patient Saf Surg.* 2009;3:26.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.