Published in final edited form as:
Comput Methods Biomech Biomed Eng Imaging Vis. 2022 ; 10(3): 308–312. doi:10.1080/21681163.2021.2002724.

Using machine learning to predict perfusionists' critical decision-making during cardiac surgery

R. D. Dias[a,b], **M. A. Zenati**[c,d], **G. Rance**[c,d], **Rithy Srey**[c,d], **D. Arney**[e,f], **L. Chen**[g], **R. Paleja**[g], **L. R. Kennedy-Metz**[c,d], **M. Gombolay**[g]

[a]Human Factors and Cognitive Engineering Lab, Stratus Center for Medical Simulation, Brigham and Women's Hospital, Boston, MA, USA;

[b]Department of Emergency Medicine, Harvard Medical School, Boston, Ma, USA;

[c]Medical Robotics and Computer Assisted Surgery Lab, Division of Cardiac Surgery, Va Boston Healthcare System, Boston, Ma, USA;

[d]Department of Surgery, Harvard Medical School, Boston, MA, USA;

[e]Medical Device Plug and Play Interoperability Program, Massachusetts General Hospital, Boston, Ma, USA;

[f]Department of Anesthesia, Harvard Medical School, Boston, Ma, USA;

[g]College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

Abstract

The cardiac surgery operating room is a high-risk and complex environment in which multiple experts work as a team to provide safe and excellent care to patients. During the cardiopulmonary bypass phase of cardiac surgery, critical decisions need to be made and the perfusionists play a crucial role in assessing available information and taking a certain course of action. In this paper, we report the findings of a simulation-based study using machine learning to build predictive models of perfusionists' decision-making during critical situations in the operating room (OR). Performing 30-fold cross-validation across 30 random seeds, our machine learning approach was

---

✉**CONTACT** R. D. Dias, rdias@bwh.havard.edu, Department of Emergency Medicine, Brigham and Women's Hospital, Harvard Medical School, 10 Vining Street, Boston, MA 0211, USA.

Notes on contributors

Roger D. Dias, MD, PhD, MBA is an Assistant Professor of Emergency Medicine at Harvard Medical School and the Director of Research and Innovation at the STRATUS Center at Brigham and Women's Hospital.

Marco A Zenati, MD, MSc, FEBCTS is a Professor of Surgery at Harvard Medical School, Chief of Cardiac Surgery at the Veterans Affairs Boston Healthcare System and Director of the Medical Robotics and Computer Assisted Surgery Lab.

Geoffrey Rance, BS, CCP is a staff perfusionist at Brigham and Women's Hospital in Boston, MA and a member of Brigham and Women's cardiac team at Cape Cod Hospital in Hyannis, MA

Rithy Srey, MS, CCP is the Chief Perfusionist at the West Roxbury VA Medical Center, Boston, MA

David Arney, PhD is the Lead Engineer of the MD PnP Program at Massachusetts General Hospital, an Instructor in Anaesthesia at Harvard Medical School, and a Faculty Associate at the Berkman Klein Center for Internet and Society at Harvard University.

Letian Chen is a Computer Science Ph.D. student at the Georgia Institute of Technology.

Rohan Paleja is a Robotics Ph.D. student at the Georgia Institute of Technology.

Lauren R. Kennedy-Metz is an Instructor of Surgery at Harvard Medical School and a member of the Medical Robotics and Computer Assisted Surgery (MRCAS) Lab at the Va Boston Healthcare System.

Matthew C. Gombolay, PhD is an assistant professor of interactive computing at the Georgia Institute of Technology and the Director of the Cognitive Optimization and Relational (CORE) Robotics Laboratory.

Disclosure statement

No potential conflict of interest was reported by the author(s).

able to achieve an accuracy of 78.2% (95% confidence interval: 77.8% to 78.6%) in predicting perfusionists' actions, having access to only 148 simulations. The findings from this study may inform future development of computerised clinical decision support tools to be embedded into the OR, improving patient safety and surgical outcomes.

Keywords

Decision-making; machine learning; cardiac surgery; perfusionists; decision support

Introduction

In the U.S. alone, it is estimated that more than 500,000 cardiac surgery operations are performed annually, with most of these procedures involving cardiopulmonary bypass (CPB) through a perfusion system managed by a perfusionist. (Epstein 2011; Alkhouli et al. 2020) Over the past 50 years, despite considerable improvements in patient safety, the incidence of preventable adverse events continues to be high in cardiac surgery, compared to other surgical specialities. (Cooley and Frazier 2000; Melly et al. 2018) Among the many factors that impact patient outcomes, intraoperative performance during CPB presents a strong association with both short- and long-term morbidity and mortality among patients undergoing cardiac surgery. (Salis et al. 2008; Chalmers et al. 2014)

During CPB, the perfusionist plays a critical role in managing the perfusion system while concurrently coordinating other complex tasks with the surgical, anaesthesia, and nursing sub-teams. (Dias et al. 2021) In addition to having sufficient knowledge and technical skills, perfusionists are required to simultaneously process a vast amount of information received from multiple sources (i.e. auditory and visual). They also need to perform tasks under stressful conditions, and efficiently communicate with the cardiac team using closed-loop techniques. (Wiegmann et al. 2009; Dias et al. 2018) Although safety devices and systems (e.g. line filters, bubble alarms, and automatic shut-offs) have been incorporated into modern perfusion systems, several studies have continued to report large variation in perfusionist practices, (Tuble et al. 2009; Ali et al. 2018) as well as persistent incidences of perfusionist-related errors, impacting patient safety and surgical outcomes in the cardiac operating room (OR). (Mejak et al. 2000; Charrière et al. 2007)

Throughout a cardiac surgery procedure, perfusionists face several challenging situations that demand critical decision-making. In a previous study, mapping process models for the intraoperative phase of cardiac surgery, (Dias et al. 2021) have found that during CPB, there are 4 steps and 21 substeps where the perfusionists play a central role while interacting with the perfusion system. This study has identified 12 decision points, 9 critical communications, 17 pitfalls, and 10 problem-solving/prevention strategies that cardiac teams use during the CPB phase. In the present study, we aimed to model expert perfusionists' decision-making process using a simulated dataset and predict perfusionists' actions during a critical clinical situation using machine learning techniques. The findings from this study may inform the future development of computerised decision-support systems in the OR with the potential to improve patient safety and surgical outcomes.

Study design and setting

This study received regulatory approval from Harvard Medical School and VA Boston Institutional Review Boards (IRB #3296). All participants involved signed an informed consent document before the start of the procedures. The study used a computer-generated simulated dataset.

Participants and procedures

We recruited two expert perfusionists (more than 10 years of experience) to participate in a total of 154 simulated decision-making trials (77 trials per perfusionist). Each perfusionist evaluated a different set of trials. All trials involved the situation during CPB in which the patient's delivery of oxygen (DO_2) levels were below a clinically relevant pre-determined threshold. A patient's DO_2 level, measured through new perfusion systems, is constantly monitored among many metrics by the perfusionist during the CPB phase. Based on evidence from the Goal-directed Perfusion Trial (GIFT), (Ranucci et al. 2018) when DO_2 levels are below a certain threshold (<280 ml/min/m^2) the perfusionist needs to make a decision in order to improve the delivery of oxygen to the patient's tissues to avoid a rise in lactate and associated morbidity. This is considered a critical decision since the DO_2 levels are directly associated with post-operative outcomes. (Ranucci et al. 2018)

Each decision trial consisted of selecting one action out of a set of five: 1) reduce flow; 2) transfuse red cells; 3) hemoconcentrate; 4) standby and alert the surgeon; and 5) standby but do not alert the surgeon. According to discussions with expert perfusionists, each of these actions is considered a suitable response to suboptimal DO_2 levels, depending upon specific conditions. These decisions (output) were based upon a set of five state variables (input): haematocrit, haemoglobin, arterial flow, body surface area, and a triggering event (e.g. an obstruction exists due to the position of the cannulas). The values for the continuous input variables were based on ranges commonly seen in cardiac surgery and generated randomly in Microsoft Excel. We encode the natural language triggering event via one-hot encoding since there are three possible events in total.

Data modelling and machine learning

In this section, we present a framework for learning, via expert demonstration, a perfusionist behaviour policy that correctly determines a perfusionist action as a function of the state. Based on prior work, (Gombolay et al. 2016; Paleja et al. 2020) in learning via expert demonstration, we leverage a machine learning model that performs counterfactual reasoning through pairwise comparisons to learn to predict which action an expert would take in a novel situation.

Our pairwise classification model infers a ranking over output classes (i.e. perfusionist actions) via pairwise comparisons with all other output classes (i.e. alternative actions). We then compute the highest-ranking class, representing the model's prediction of the preferred action. As we are only given the actions chosen by the demonstrating perfusionist and not the relative importance between outputs, we must first reformat training examples to construct pairwise comparisons between the decision chosen by the perfusionist and those not chosen.

We can create positive counterfactuals and negative counterfactuals via Equation 1 and 2, respectively. Here, the term counterfactual denotes that the model is comparing two output classes and reasoning about why an output class was chosen over another. Equation 1 creates a positive counterfactual example for each perfusionist decision. The counterfactual example consists of the feature vector describing the current state, $\hat{x}$ (haematocrit, haemoglobin, arterial flow, body surface area, and a trigger event), and descriptors encoding output-specific information relating to each of the possible outputs (x_a and $x_{a'}$). Utilising the difference of output-specific features between the decision chosen, a, and a decision that was not chosen, a', the state $\hat{x}$, and a positive label, $y_{a,a'} = 1$, we define positive examples as displayed in Equation 1. We can similarly define negative examples by utilising the difference of output-specific features between the decision not chosen, a', and a decision that was chosen, a, the state $\hat{x}$, and a negative label, $y_{a,a'} = 0$, denoting that a' is not preferred over decision a. We note that the dataset does not provide output-specific features, x_a. Accordingly, we utilise a one-hot encoding of length five to represent output-specific information, where the output chosen would be given a value of 1 in the corresponding index of the one-hot encoding. Given a single observation, we can compute four positive examples, and four negative examples (as there are five total classes). We note that this model is technically a hybrid pointwise-pairwise model, as we include a pointwise term, $\hat{x}$, that captures the state of the world, and pairwise terms, $x_a - x_{a'}$ or $x_{a'} - x_a$, to reason about the differences in actions. Preserving this pointwise information is critical, as it provides the contextual information to perform the pairwise comparison.

$$z_{a,\,a'} \coloneqq [\hat{x}, x_a - x_{a'}], y_{a,\,a'} = 1\ for \forall a' \in A \backslash a \tag{1}$$

$$z_{a',\,a} \coloneqq [\hat{x}, x_{a'} - x_a], y_{a',\,a} =\ 0 for \forall a' \in A \backslash a \tag{2}$$

Given the generated positive and negative examples, we train a Random Forest (Breiman, 2001) (RF) classifier $f_{RF}(a, a') \in \{0, 1\}$ to predict whether it would be better to perform action a in comparison to a' in the current state $\hat{x}$. With this pairwise classifier, we can determine the action associated with the highest ranking via Equation 3. In Equation 3, we compare each action to another within the current state and determine a vector that represents the aggregate preference of each decision choice. We take the argument max of this vector to determine the demonstrator's output decision.

$$y* = \arg\max_{a} \sum_{a' \in A} f_{RF}(a, a') \tag{3}$$

Results

We collected a preliminary dataset consisting of simulated trials completed by two expert perfusionists, yielding a total of 154 perfusionist decisions under the condition of a low patient's DO_2 level, which is a real-life critical situation that requires prompt action by the perfusionist. Performing 30-fold cross-validation across 30 random seeds, our machine learning approach was able to achieve an accuracy of 78.2% (95% confidence interval:

77.8% to 78.6%) in predicting perfusionists' actions (Figure 1), having access to only 148 simulations.

We note that achieving an accuracy of 78.2% requires encoding a variable (trial id) to capture which perfusionist (from a set of two) made each decision. As an ablation study, we compare the model performance with and without the trial ID across 30 random seeds. The model without the trial ID achieves an accuracy of 71.3% (95% confidence interval: 71.0% to 71.7%). Compared to the model with trial ID, we see an average 6.91% improvement. We further compare the two models' performance across 30 random seeds via the Mann-Whitney U test since the normality assumption for independent t-test does not hold (Shapiro–Wilk test $p < .05$). The Mann-Whitney U test shows significant improvement when we have the trial ID as a feature, $U = 0.0$, $p < .001$.

We also compare the efficacy of the pairwise model with the pointwise model.(Valizadegan et al. 2009) Pointwise models output a probability of taking a certain action given an output-specific feature, x_a, and the current state, $\hat{x}$. Positive and negative pointwise examples can be generated through Equations 4 and 5, respectively.

$$z_a := [\hat{x}, x_a], y_a = 1 for \forall a' \in A \backslash a \tag{4}$$

$$z_{a'} := [\hat{x}, x_{a'}], y_{a'} = \ 0 for \forall a' \in A \backslash a \tag{5}$$

Like the pairwise model, we learn a Random Forest classification model $f_{RF}^{Pointwise}(a) \in \{0, 1\}$ to predict whether we should perform action a in the current state. We can then determine the action associated with the highest score via Equation 6.

$$y* = \arg\max_a f_{RF}^{Pointwise}(a) \tag{6}$$

The pointwise model obtains an accuracy of 67.7% (95% confidence interval: 66.9% to 68.5%). The Mann-Whitney U test shows a significant improvement of 10.6% on accuracy with our pairwise model compared with the pointwise model (normality assumption does not hold for t-test: $p < .05$; Mann-Whitney U test $U = 0.0$, $p < .001$). For our pairwise model, we report an average sensitivity of .78 ± .20 and specificity of .92 ± .04 over five output actions (weighted by the number of times each output appeared within the dataset) averaged over 30 folds and 30 seeds. We provide a comparison across models in Figure 2, displaying the efficacy and high-performance of our model.

Discussion

In this study, we used machine learning techniques to infer expert perfusionists' decision-making during critical intraoperative situations using a relatively small simulated dataset. Despite these limitations, our machine learning model achieved substantial accuracy in predicting expert perfusionists' actions.

Increasing research in the field of surgical data science has been generated seeking to incorporate artificial intelligence (AI) and machine learning applications into surgical practice. (Nagy et al. 2017) Although most of these studies are focused on surgical workflow segmentation and image guidance navigation, recent studies are proposing to use AI and machine learning to augment human cognition and improve decision-making, situational awareness, and communication in the OR. (Dias et al. 2020; Zenati et al. 2020; Condello et al. 2021) for example, have recently reviewed the literature and provided a roadmap on management algorithms and AI systems for cardiopulmonary bypass.

In this present study, we modelled a simulated dataset with limited decision options based on expert perfusionists' input, and future studies should incorporate substantially more decisions, as well as time-series data, both of which will improve the power of our models. The inter-operator disagreement presented in our curated dataset of simulated decisions provides support for the need to develop and validate a data-driven approach to inferring optimal standardised care.

The findings from our study are a first step towards validating useful clinical decision support tools to be embedded into the OR. These tools may help standardisation of clinical practices across different centres and even within the same centre across different levels of experience (juniors vs senior perfusionists). Additionally, the same approach can be used to develop and validate cognitive aids for other OR professionals, such as surgeons, anaesthesiologists, and nurses, as well as other healthcare settings (e.g. emergency room, critical care unit). An example of a possible application in the OR is a machine learning-based recommendation system that evaluates the current situation in real-time and provides recommendations to perfusionists on certain courses of action. This system would use a human-in-the-loop approach in which the human (perfusionists) would be the final decision-maker.

There are important limitations in this study that need to be addressed. First, the decisions were made on a computer which does replicate relevant factors that certainly impact decision-making, such as cognitive overload, environmental noise, miscommunication, and workflow disruptions. Second, only two perfusionists were included and a more representative sample size would probably generate more variability in actions. Finally, the randomness of the simulated input variables provided to perfusionists does not capture physiological patterns and interdependencies between these variables, which may not be realistic in naturalistic decision-making. The validity evidence provided in this study is limited, and future research should establish additional validity, especially related to decision-making in real-life cardiac procedures.

Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Funding

This work was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH) under Award Number R01HL126896 (PI: Zenati). This work was also partially funded by the Office of Naval Research under grant GR10006659 (PI: Gombolay).

References

Ali JM, Miles LF, Abu-Omar Y, Galhardo C, Falter F. 2018. Global cardioplegia practices: results from the global cardiopulmonary bypass survey. J Extra Corpor Technol. 50(2):83–93. [PubMed: 29921986]

Alkhouli M, Alqahtani F, Kalra A, Gafoor S, Alhajji M, Alreshidan M, Holmes DR, Lerman A. 2020. Trends in characteristics and outcomes of patients undergoing coronary revascularization in the United States, 2003–2016. JAMA Netw Open. 3(2):e1921326. DOI:10.1001/jamanetworkopen.2019.21326. [PubMed: 32058558]

Breiman L 2001. Random forests. Mach Learn. 45(1):5–32. DOI:10.1023/A:1010933404324.

Chalmers J, Pullan M, Mediratta N, Poullis M. 2014. A need for speed? Bypass time and outcomes after isolated aortic valve replacement surgery. Interact Cardiovasc Thorac Surg. 19(1):21–26. DOI:10.1093/icvts/ivu102. [PubMed: 24722513]

Charrière J-M, Pélissié J, Verd C, Léger P, Pouard P, de Riberolles C, Menestret P, Hittinger MC, and Longrois D. 2007. Survey: retrospective survey of monitoring/safety devices and incidents of cardiopulmonary bypass for cardiac surgery in France. J Extra Corpor Technol. 39(3):142–157. discussion 158–159. [PubMed: 17972449]

Condello I, Santarpino G, Nasso G, Moscarelli M, Fiore F, Speziale G. 2021. Management algorithms and artificial intelligence systems for cardiopulmonary bypass [published online ahead of print, 2021 Jul 12]. Perfusion. 2676591211030762. DOI:10.1177/02676591211030762

Cooley DA, Frazier OH. 2000. The past 50 years of cardiovascular surgery. Circulation. 102(Supplement 4):IV–87. DOI:10.1161/01.cir.102.suppl_4.iv-87.

Dias RD, Conboy HM, Gabany JM, Clarke LA, Osterweil LJ, Avrunin GS, Arney D, Goldman JM, Riccardi G, Yule SJ, and Zenati MA. 2018. Development of an interactive dashboard to analyze cognitive workload of surgical teams during complex procedural care. IEEE Int Interdiscip Conf Cogn Methods Situat Aware Decis Support. 2018:77–82. [PubMed: 30547096]

Dias RD, Shah JA, Zenati Ma. 2020 Oct. Artificial intelligence in cardiothoracic surgery. Minerva Cardioangiol. 68(5):532–538. Epub 2020 Sep 29. DOI:10.23736/S0026-4725.20.05235-4. [PubMed: 32989966]

Dias RD, Zenati MA, Conboy HM, Clarke LA, Osterweil LJ, Avrunin GS, and Yule SJ. 2021 Aug 1. Dissecting cardiac surgery: a video-based recall protocol to elucidate team cognitive processes in the operating room. Ann Surg. 274(2):e181–e186. [PubMed: 31348036]

Epstein AJ. 2011. Coronary revascularization trends in the United States, 2001–2008. JAMA. 305(17):1769. DOI:10.1001/jama.2011.551. [PubMed: 21540420]

Gombolay M, Jensen R, Stigile J, Son S-H, and Shah J. Apprenticeship scheduling: learning to schedule from human experts. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI). New York, New York; 2016.

Mejak BL, Stammers A, Rauch E, Vang S, Viessman T. 2000. A retrospective study on perfusion incidents and safety devices. Perfusion. 15(1):51–61. DOI:10.1177/026765910001500108. [PubMed: 10676868]

Melly L, Torregrossa G, Lee T, Jansens J-L, Puskas JD. 2018. Fifty years of coronary artery bypass grafting. J Thorac Dis. 10(3):1960–1967. DOI:10.21037/jtd.2018.02.43. [PubMed: 29707352]

Nagy DA, Rudas IJ, and Haidegger T. Surgical data science, an emerging field of medicine. 2017 IEEE 30th Neum ann Colloquium (NC). Budapest, Hungary; 2017.

Paleja RR, Silva A, Chen L, and Gombolay M. Interpretable and personalized apprenticeship scheduling: learning interpretable scheduling policies from heterogeneous user demonstrations. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS). Virtual; 2020.

Ranucci M, Johnson I, Willcox T, Baker RA, Boer C, Baumann A, Justison Ga, de Somer F, Exton P, Agarwal S, et al. 2018. Goal-directed perfusion to reduce acute kidney injury: a randomized trial. J Thorac Cardiovasc Surg. 156(5):1918–1927.e2. DOI:10.1016/j.jtcvs.2018.04.045. [PubMed: 29778331]

Salis S, Mazzanti VV, Merli G, Salvi L, Tedesco CC, Veglia F, Sisillo E. 2008. Cardiopulmonary bypass duration is an independent predictor of morbidity and mortality after cardiac surgery.

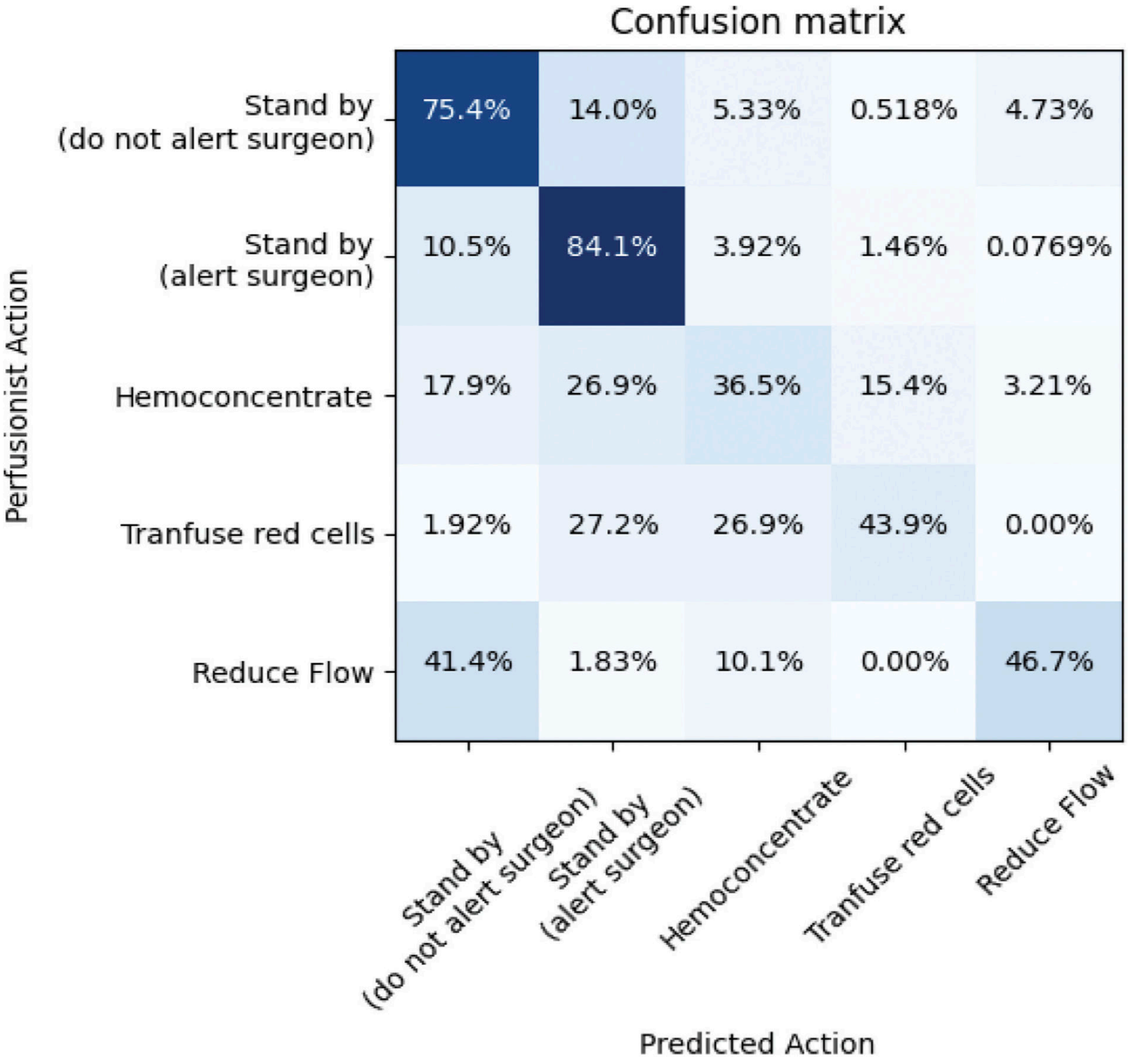
J Cardiothorac Vasc Anesth. 22 (6):814–822. DOI:10.1053/j.jvca.2008.08.004. [PubMed: 18948034]

Tuble SC, Willcox TW, Baker RA. 2009. Australian and New Zealand perfusion survey: management and procedure. J Extra Corpor Technol. 41 (2):64–72. [PubMed: 19681302]

Valizadegan H, Jin R, Zhang R, and Mao J,. Learning to rank by optimizing NDCG measure. In Proceedings of the Conference on Neural Information Processing Systems (NeurIPS). Vancouver, Canada; 2009.

Wiegmann D, Suther T, Neal J, Parker SH, Sundt TM. 2009. A human factors analysis of cardiopulmonary bypass machines. J Extra Corpor Technol. 41(2):57–63. [PubMed: 19681301]

Zenati MA, Kennedy-Metz L, Dias RD. 2020. Cognitive engineering to improve patient safety and outcomes in cardiothoracic surgery. Semin Thorac Cardiovasc Surg. 32(1):1–7. DOI:10.1053/j.semtcvs.2019.10.011. [PubMed: 31629782]



	Stand by (do not alert surgeon)	Stand by (alert surgeon)	Hemoconcentrate	Tranfuse red cells	Reduce Flow
Stand by (do not alert surgeon)	75.4%	14.0%	5.33%	0.518%	4.73%
Stand by (alert surgeon)	10.5%	84.1%	3.92%	1.46%	0.0769%
Hemoconcentrate	17.9%	26.9%	36.5%	15.4%	3.21%
Tranfuse red cells	1.92%	27.2%	26.9%	43.9%	0.00%
Reduce Flow	41.4%	1.83%	10.1%	0.00%	46.7%

Figure 1.

Confusion matrix for our method predicting the action of perfusionists.

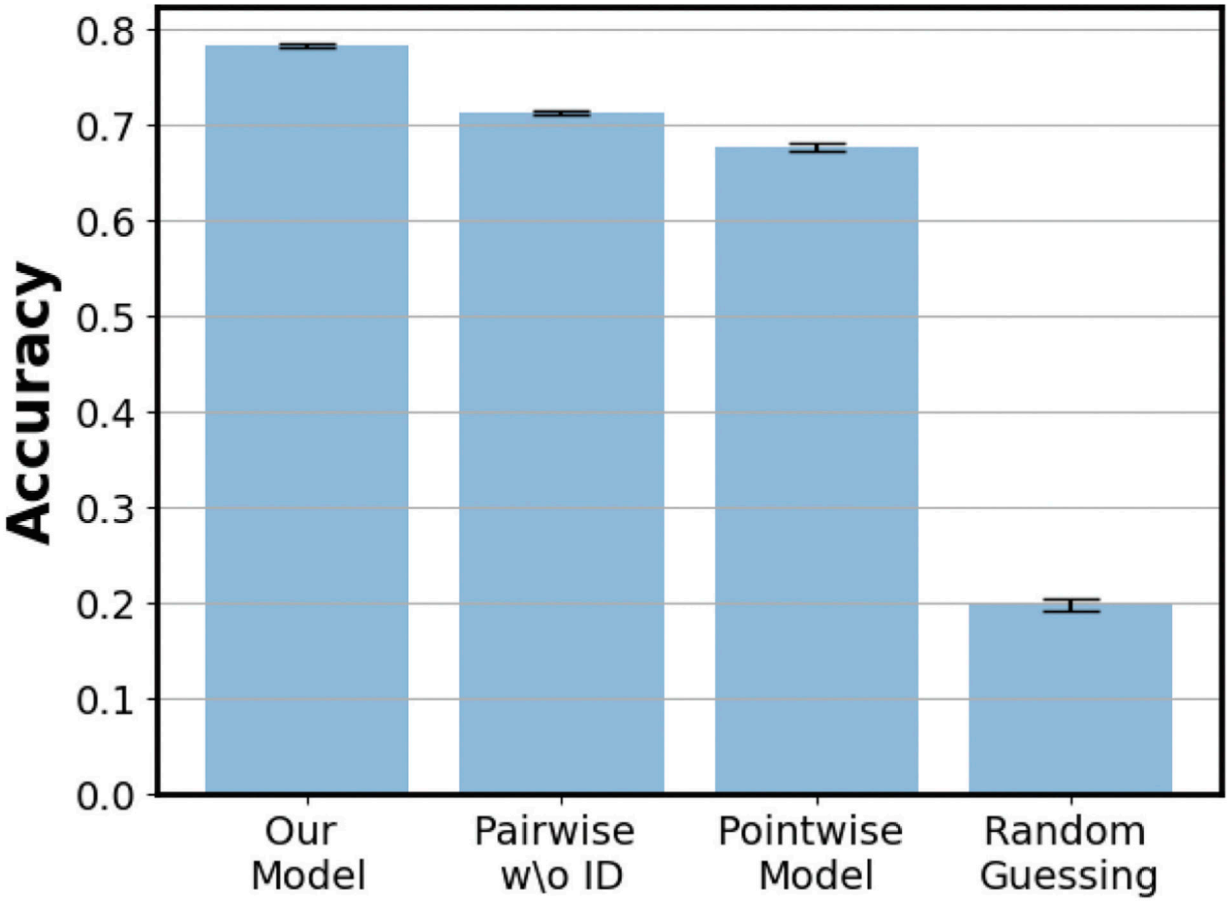


Figure 2.

Prediction accuracies (30-fold cross validation) across 30 random seeds for our model, a pairwise model without ID, a pointwise model, and random guessing.