Arthroplasty Today 20 (2023) 101116

Contents lists available at ScienceDirect

Arthroplasty Today

journal homepage: http://www.arthroplastytoday.org/

Systematic Review

Improving Resource Utilization for Arthroplasty Care by Leveraging Machine Learning and Optimization: A Systematic Review

Bahar Entezari, BMSc [a, b, *], Robert Koucheki, HBSc [c, d], Aazad Abbas, HBSc [c, e], Jay Toor, MD, MBA [f], Jesse I. Wolfstadt, MD, MSc, FRCSC [a, f], Bheeshma Ravi, MD, PhD, FRCSC [f, g], Cari Whyne, PhD, FIOR [d, e, f, g], Johnathan R. Lex, MB, ChB [e, f]

[a] Granovsky Gluskin Division of Orthopaedics, Mount Sinai Hospital, Toronto, Ontario, Canada
[b] Queen's University School of Medicine, Kingston, Ontario, Canada
[c] Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
[d] Institute of Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada
[e] Orthopaedic Biomechanics Lab, Sunnybrook Research Institute, Toronto, Ontario, Canada
[f] Division of Orthopaedic Surgery, Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
[g] Division of Orthopaedic Surgery, Holland Bone and Joint Program, Sunnybrook Health Science Centre, Toronto, Ontario, Canada

A R T I C L E   I N F O

Article history:
Received 28 December 2022
Accepted 28 January 2023
Available online xxx

Keywords:
Artificial intelligence
Predictive modeling
Surgical scheduling
Optimization
Total knee arthroplasty
Total hip arthroplasty

A B S T R A C T

Background: There is a growing demand for total joint arthroplasty (TJA) surgery. The applications of machine learning (ML), mathematical optimization, and computer simulation have the potential to improve efficiency of TJA care delivery through outcome prediction and surgical scheduling optimization, easing the burden on health-care systems. The purpose of this study was to evaluate strategies using advances in analytics and computational modeling that may improve planning and the overall efficiency of TJA care.
Methods: A systematic review including MEDLINE, Embase, and IEEE Xplore databases was completed from inception to October 3, 2022, for identification of studies generating ML models for TJA length of stay, duration of surgery, and hospital readmission prediction. A scoping review of optimization strategies in elective surgical scheduling was also conducted.
Results: Twenty studies were included for evaluating ML predictions and 17 in the scoping review of scheduling optimization. Among studies generating linear or logistic control models alongside ML models, only 1 found a control model to outperform its ML counterpart. Furthermore, neural networks performed superior to or at the same level as conventional ML models in all but 1 study. Implementation of mathematical and simulation strategies improved the optimization efficiency when compared to traditional scheduling methods at the operational level.
Conclusions: High-performing predictive ML-based models have been developed for TJA, as have mathematical strategies for elective surgical scheduling optimization. By leveraging artificial intelligence for outcome prediction and surgical optimization, there exist greater opportunities for improved resource utilization and cost-savings in TJA than when using traditional modeling and scheduling methods.

Introduction

Total joint arthroplasty (TJA) procedures, including total knee arthroplasty (TKA) and total hip arthroplasty (THA), are the most commonly performed surgical procedures in North America [1–3]. Given their success in restoring function and improving quality of life, combined with an aging population and increasing demand,

* Corresponding author. Mount Sinai Hospital, 15 Arch Street, Kingston, Ontario, Canada K7L 3N6. Tel.: +1 647 866 8729.
E-mail address: bentezari@qmed.ca

https://doi.org/10.1016/j.artd.2023.101116

the rate at which TJA procedures are performed will continue to rise [4,5]. By 2030, the number of TKAs and THAs performed annually in the United States are projected to reach over 1.26 million and 635,000, respectively [6]. With an average cost of $16,000 to $60,000 USD, the financial burden of these procedures will place significant strain on health-care institutions [7,8].

In order to keep up with the growing demand for TJA procedures, significant investments must be made in strategies to improve the efficiency and cost-effectiveness of care [9]. The growing volume of patient data combined with the utilization of new technologies such as artificial intelligence (AI) provides opportunities to improve the delivery of care in orthopaedic surgery [9–12]. In particular, machine learning (ML), a subset of AI, has caught the attention of orthopaedic surgeons and health-care institutions due to its potential for generating accurate patient-specific predictive models by recognizing linear and nonlinear relationships from large data sources [13–16]. Deep learning (DL) models, typically in the form of neural networks, are a subset of ML models theoretically capable of generating highly accurate predictions when trained with appropriate and sufficient data (Fig. 1a). Using conventional ML and DL models, perioperative outcomes associated with increased costs can be anticipated, and measures can be taken to minimize their burden. Such outcomes include length of stay (LOS), duration of surgery (DOS), and unplanned hospital readmissions [17–20]. Following the generation of accurate predictions, in order to implement or realize any predicted gains, optimization of available resources must also be performed. The optimization of surgical schedules is typically considered a nondeterministic polynomial time-hard problem, with various optimization strategies available (Fig. 1b) [17]. Applying the principles of operations research in health care is facilitated by advances and wider-spread availability of high-performance computing and growing amounts of curated digital data.

Numerous studies have recently emerged assessing the ability of ML models to predict perioperative resource-utilization-related outcomes surrounding TJA; however, the success of different algorithms, data sets, and comparison to traditional statistical methods have not been systematically evaluated. Furthermore, the theoretical optimization of elective surgical scheduling across surgical specialties has been heavily investigated in the engineering literature. However, many of these principles have not been translated into the orthopaedic literature to guide surgeons and key stakeholders as they evaluate and implement these methods of efficiency optimization in real-life situations. The purpose of this study was to systematically evaluate the literature investigating the impact of ML and optimization strategies on the planning, scheduling, and overall efficiency of TJA care.

Material and methods

This systematic review was registered in the National Institute for Health Research PROSPERO (International Prospective Register of Systematic Reviews) database (ID #: CRD42022377977). The review process was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [18].

Search strategy

The MEDLINE, Embase, and IEEE Xplore databases were searched using the following combination of search terms: ((machine learning) OR (artificial intelligence) OR (deep learning) OR ML OR AI OR prediction OR (neural network) OR (branch and cut) OR (Monte*Carlo) OR simulation OR heuristic OR stochastic OR (ant colony) OR (meta*heuristics) OR optimization OR (operations research)) AND (scheduling OR planning OR theatre OR theater OR (patient admission scheduling) OR (length of* stay) OR (duration of surgery) OR (surg* time) OR (surgical duration) OR (operative time) OR (operating time) OR (outpatient) OR DOS OR LOS) AND (TKA OR THA OR TKR OR THR OR (total knee arthroplasty) OR (total hip arthroplasty) OR (total knee replacement) OR (total hip replacement) OR (joint replacement) OR (joint arthroplasty)). All studies from inception to October 4, 2022, were retrieved, and duplicate manuscripts removed. Manual searching of the reference lists of included studies was also conducted.

Inclusion and exclusion criteria

Abstract and title screening of search results was performed by 3 independent reviewers (B.E., J.R.L., and R.K.). Full texts were screened for study inclusion according to the inclusion and exclusion criteria by all 3 reviewers. Conflicts were resolved by consensus among all reviewers throughout the screening process. Relevant data were extracted and recorded on a predetermined data-collection form by 2 independent reviewers (B.E. and R.K.). Studies of all languages were included and translated into English as required. Abstracts and reviews were excluded.

Artificial intelligence prediction

This section of the review was performed systematically. Studies were included if they used any type of ML model to predict 1 of 3 outcomes following TJA; LOS, DOS, or postoperative readmission. Studies evaluating patients undergoing primary or revision TKA, THA, partial knee arthroplasty, or hip resurfacing were included. Studies were not excluded based on varying definitions of DOS, namely whether it was total patient time in the room, surgical time, or included room turnover. No studies were excluded based on data set size or type.

Mathematical optimization

Due to the size, complexity, and depth of the existing literature, many of the concepts and articles were beyond the scope of this review. Therefore, this section of the review was performed in the form of a narrative review. Studies were included if they used computational modeling or mathematical optimization in an attempt to improve the efficiency of elective surgical scheduling. These articles were recorded regardless of surgery specialty as operations research for health care is traditionally specialty-agnostic and aims to optimize the schedule of the entire institution; however, these strategies can be applied to any elective surgery scheduling problem.

Data extraction

Artificial intelligence prediction

General study data collected included first author name, publication year and country, data source, and procedure. Model data were collected for control and ML models, where control models were considered those using average times, multivariable linear or logistic regression. ML models were considered as any algorithms having undergone training without explicit manual programming. Model data collected included outcome(s) of interest (LOS, DOS, and hospital readmission), algorithm type(s), total population size, training size, validation size and method, testing size, input features, and feature importance. For articles generating multiple models, the most important features from the best-performing model were recorded. Metrics for scoring model performance were collected. For the purpose of model comparison, mean squared error (MSE) was used as the gold standard for regression-type models, followed by root MSE, and mean absolute error; area

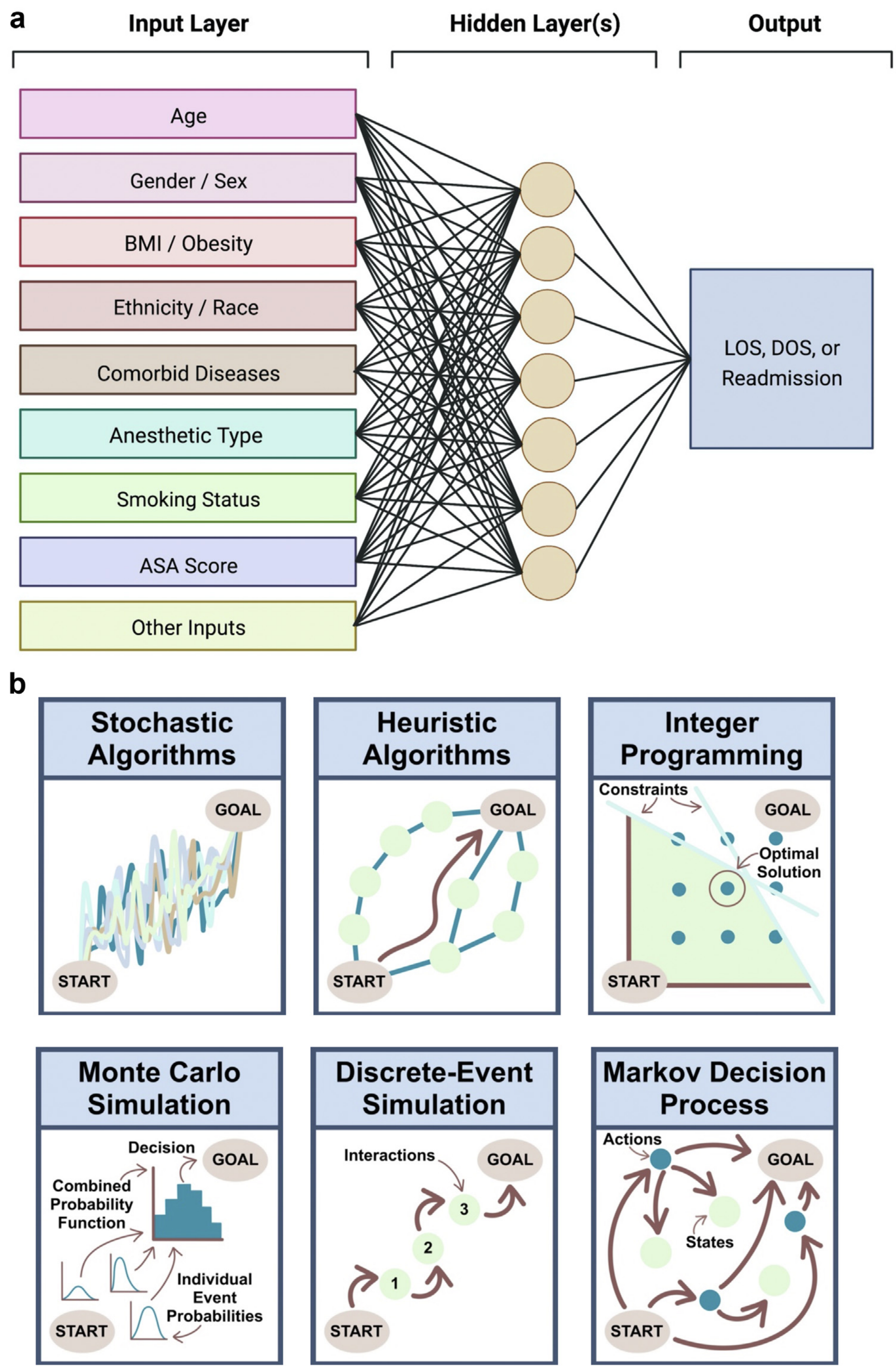

Figure 1. Diagrammatic representation of (a) neural network models, demonstrating the transformation of a combination of input features in hidden layers to yield outcome predictions, and (b) types of optimization strategies. ASA, American Society of Anesthesiologists; BMI, body mass index.

under the curve (AUC) was used as the gold standard for classification-type models, followed by accuracy, and F1 score.

Mathematical optimization

Collected data included the study optimization problem goal, overall strategy, and the main findings. Surgical scheduling is typically broken down into 3 levels with corresponding definitions: (1) strategic level, planning and operating room (OR) allocation to specialties/surgeons over the period of a year or longer; (2) tactical level, cyclic regular seasonal or weekly OR schedules; and (3) operational level, scheduling cases by date, time, and specific resources required. The decision level at which the optimization problem was targeted to solve was recorded. The optimization strategy was categorized into the use of a heuristic algorithm, stochastic algorithm, integer programming, or simulations (including Markov decision process) to find the optimized scheduling solution (Table 1).

Results

Search results and study characteristics

Following removal of duplicates, the search revealed a total of 1753 unique articles via MEDLINE, Embase, and IEEE Xplore databases (Fig. 2). Upon abstract and title screening, 1715 articles were excluded. Eighteen additional articles were excluded following full-text review, leaving 25 eligible articles for inclusion [19–43]. An additional 12 articles were retrieved via manual searching of references and included in the scoping review of mathematical optimization [44–55].

Study characteristics

Across the 20 included studies reporting results of AI models, 16 developed predictive models for LOS, 3 for DOS, and 2 for 90-day hospital readmission (Table 2). Models were built using various data sources, ranging from single institutional data sets to national (ie, the New York State Department of Health's Statewide Planning and Research Cooperative System [SPARCS], National Inpatient Sample [NIS], and Orthopedic Minimal Data Set Episode of Care [OrthoMiDaS OME] databases) and multinational databases (ie, the American College of Surgeons National Surgical Quality Improvement Program [NSQIP] database), with data set sizes ranging from 525 to 424,443 patients (Table 2). Despite the variability in data sets, there was significant similarity in the features used to generate the models. The 10 most frequently utilized input features for model development were age (100%), gender/sex (100%), body mass index/obesity (70%), ethnicity/race (60%), diabetes (55%), hypertension (50%), anaesthetic type (45%), smoking status (45%), cardiovascular disease (45%), and American Society of Anesthesiologists score (45%) (Supplementary Table S1). Fifteen different types of ML algorithms were developed (Supplementary Table S2). These included Bayesian, K-nearest neighbor, support vector machine, stochastic gradient descent, random forest classifier, decision tree, gradient boosted decision tree, XGBoost, AdaBoost, CatBoost, RUSBoost, ridge regression, lasso regression, elastic net regression, and artificial neural network (ANN) models. Ten of the 20 studies (50.0%) compared the developed ML models to control models, including mean regressor, multivariate linear regression, and logistic regression models (Supplementary Table S2).

Length of stay

Six studies developed predictive models for THA LOS, 9 for TKA LOS, and 1 for revision TKA LOS [19–21,24–28,31–33,35–37,56]. One study generated a combined model for TKA and THA LOS [38]. The average number of input features was 16.7 (range: 8 to 29) (Table 3). Among those studies reporting AUC values, model performance ranged from 0.64 to 0.87. The best-performing model for THA LOS prediction was the Bayesian model generated by Ramkumar et al. using 8 input features, with an AUC of 0.87 [22]. For TKA LOS, the ANN developed by Ramkumar et al. using 15 input features yielded the highest reported AUC of 0.83 [20]. The ANN developed by Klemt et al. using 26 input features for the prediction of revision TKA LOS had an AUC of 0.87, outperforming the support vector machine and elastic net regression models developed with the same data [31]. Within no single study did a control model outperform its equivalent ML model in the prediction of LOS.

Among studies that generated ANNs and conventional ML models, only 1 found a conventional ML model to outperform its neural network counterpart: The random forest classifier model generated by Han et al. using 27 input features for the prediction of LOS showed significantly superior performance to their ANN (Table 3) [26]. Using a decision curve, this model was also found to yield superior clinical usefulness when compared to the ANN model [45]. All other studies demonstrated ANN models performed superior to or at the same level as their conventional ML counterparts (including k-nearest neighbor, support vector machine, stochastic gradient descent, random forest classifier, XGBoost, etc.) [25,29,33,35].

Duration of surgery

Three studies developed ML models for the prediction of DOS, all of which were for TKA [30,33,34]. The CatBoost model of Motesharei et al. demonstrated an R2 value of 0.76, outperforming their control multivariate linear regression model with an R2 value of 0.71, as well as the random forest classifier and gradient-boosted

Table 1
Categories of optimization strategies and their associated benefits and limitations.

Optimization strategy	Definition	Benefits	Limitations
Heuristic algorithms	Specific rules-based functions to provide an approximate solution	Less computational resources and time to solve	Solution may not be optimal or complete
Stochastic algorithms	Optimization that uses random inputs, random objective functions, or random constraints	Accommodates for imprecise measurements in data, estimates the average model performance	Random elements prevent problem from finding optimal solution
Integer programming	Mathematical optimization to solve problems, can be linear or mixed (if some variables are not integers). Overarching term including exact algorithm solutions and heuristic and stochastic algorithms	Used to find optimal optimization problem solutions	Searching for optimal solutions may require high computational power, and random inputs after solution implementation may differ from optimal solution
Monte Carlo simulation	Use of random sampling with a range of assigned probabilities to model or find best outcome from systems with uncertainty (such as scheduling)	To solve mathematical problems too complicated to solve analytically	Requires many samples for good approximation, high computational power requirement
Discrete-event simulation	Used to model systems over time. Systems are broken down into a sequence of steps where entities compete for limited resources and can develop queues	Evaluate the impact of changes in practice prior to implementation	Complex decision-making processes and real-world constraints can be hard to model, require significant amount of data
Markov decision process	Mathematical framework for modeling decision-making, combining random inputs with chosen inputs	Evaluate the impact of changes in practice prior to implementation, less computational power required than other simulations	Complex decision-making processes and real-world constraints can be hard to model, require significant amount of data

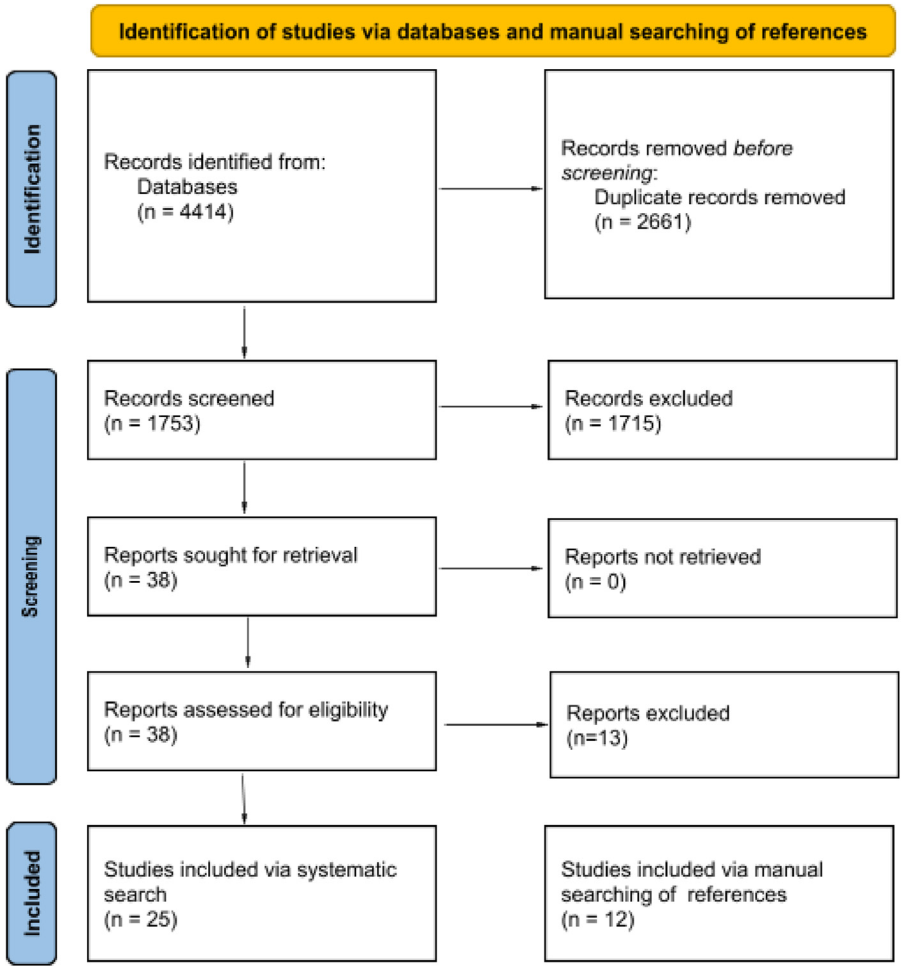

Figure 2. PRISMA flow diagram of the search strategy.

Table 2
Characteristics of the included predictive modeling studies.

Study, year	Country	Procedure	Total patients	Data source	Outcome(s) of interest	Number of input features
Navarro et al., 2018	USA, UK	TKA	141,446	SPARCS	LOS	8
Ramkumar et al., 2019	USA, UK	TKA	170,766	NIS, OrthoMiDaS OME	LOS	15
Ramkumar et al., 2019	USA, UK	THA	79,226	NIS, OrthoMiDaS OME	LOS	15
Ramkumar et al., 2019	USA, UK	THA	122,334	SPARCS	LOS	8
Lee et al., 2019	USA	Both	525	Local (USA)	90-d readmission	33
Gabriel et al., 2019	USA	THA	960	Local (USA)	LOS	9
Wei et al., 2021	USA	TKA	25,115	NSQIP	LOS	11
Han et al., 2021	China	TKA	1298	Local (China)	LOS	27
Kugelman et al., 2021	USA	THA	1409	Local (USA)	LOS	15
Yeo et al., 2022	USA	TKA	10,021	Local (USA)	DOS	15
Klemt et al., 2022	USA	rTKA	2588	Local (USA)	LOS	26
Lopez et al., 2022	USA	Both	424,443	NSQIP	LOS	20
Abbas et al., 2022	Canada	TKA	302,300	NSQIP	LOS, DOS	29
Moteshareí et al., 2022	France	TKA	1061	Local (USA)	DOS	17
Zalikha et al., 2022	USA	TKA	305,577	NIS	LOS	15
Johannesdottir et al., 2022	Denmark	Both	9512	Local (Denmark)	LOS	22
Klemt et al., 2022	USA	TKA	10,021	Local (USA)	90-d readmission	24
Li et al., 2022	China	TKA	1590	Local (Singapore)	LOS	14
Kugelman et al., 2022	USA	TKA	899	Local (USA)	LOS	15
Trunfio et al., 2022	Italy	THA	2515	Local (Italy)	LOS	15

NIS, National Inpatient Sample; NSQIP, National Surgical Quality Improvement Program; OrthoMiDaS OME, Orthopedic Minimal Data Set Episode of Care; rTKA, revision total knee arthroplasty; SPARCS, New York State Department of Health's Statewide Planning and Research Cooperative System.

Table 3
Metrics describing the performance of the best-performing machine learning and control algorithms of the included studies.

Stud, year	Procedure	Outcome	Training size	Validation size	Testing size	Best control algorithm	Control metric, value	Best ML algorithm	ML metric, value
Navarro et al., 2018	TKA	LOS	106,085	-	35,361	-	-	Bayesian	AUC, 0.78
Ramkumar et al., 2019	TKA	LOS	150,074	16,675	4017	-	-	ANN	AUC, 0.83
Ramkumar et al., 2019	THA	LOS	68,810	7645	2771	-	-	ANN	AUC, 0.80
Ramkumar et al., 2019	THA	LOS	91,751	-	30,583	-	-	Bayesian	AUC, 0.87
Lee et al., 2019	Both	90-d readmission	473	52	-	Logistic regression	Accuracy, 0.93	RUSBoost	Accuracy, 0.87
Gabriel et al., 2019	THA	LOS	644	316	-	Logistic regression	AUC, 0.75	Ridge regression	AUC, 0.76
Wei et al., 2021	TKA	LOS	15,069	-	10,046	Logistic regression	AUC, 0.80	ANN	AUC, 0.80
Han et al., 2021	TKA	LOS	1038	260	-	Logistic regression	AUC, 0.70	RFC	AUC, 0.77
Kugelman et al., 2021	THA	LOS	902	225	282	-	-	XGBoost	AUC, 0.82
Yeo et al., 2022	TKA	DOS	6413	1603	2005	-	-	ANN	AUC, 0.82
Klemt et al., 2022	rTKA	LOS	1656	414	518	-	-	ANN	AUC, 0.87
Lopez et al., 2022	TKA	LOS	216,960	-	54,420	-	-	ANN	AUC, 0.80
	THA	LOS	122,442	-	30,611	-	-	ANN	AUC, 0.81
Abbas et al., 2022	TKA	DOS	182,000	57,841	62,459	Linear regression	MSE, 0.99	ANN	MSE, 0.89
	TKA	LOS	182,000	57,841	62,459	Linear regression	MSE, 0.79	ANN	MSE, 0.69
Motesharei et al., 2022	TKA	DOS	708	177	176	Linear regression	R2, 0.71	CatBoost	R2, 0.76
Zalikha et al., 2022	TKA	LOS	195,556	48,906	61,115	-	-	SVM	AUC, 0.68
Johannesdottir et al., 2022	Both	LOS	8561	951	-	Logistic regression	AUC, 0.70	RFC	AUC, 0.71
Klemt et al., 2022	TKA	90-day readmission	6413	1603	2005	Logistic regression	-	ANN	AUC, 0.85
Li et al., 2022	TKA	LOS	-	-	-	Logistic regression	AUC, 0.64	XGBoost	AUC, 0.74
Kugelman et al., 2022	TKA	LOS	575	144	180	-	-	XGBoost	AUC, 0.69
Trunfio et al., 2022	THA	LOS	2012	-	503	Linear regression	RMSE, 3.84	GBDT	RMSE, 3.84

GBDT, gradient-boosted decision tree; rTKA, revision total hip arthroplasty; R2, coefficient of determination; RFC, random forest classifier; RMSE, root mean square error; SVM, support vector machines.

decision tree models generated in this study (Table 3) [34]. Similarly, the ANN developed by Abbas et al. demonstrated a MSE of 0.89, outperforming their control multivariate linear regression model with an MSE of 0.99, as well as every other ML model (Bayesian, K-nearest neighbor, stochastic gradient descent, random forest classifier, decision tree, XGBoost, AdaBoost, and elastic net regression) [33]. Finally, with an AUC value of 0.82, the ANN model of Yeo et al. yielded the best performance when compared to their k-nearest neighbor and random forest classifier models [30].

Hospital readmission

Two studies developed ML models for the prediction of hospital readmissions [23,29]. In addition to a control logistic regression model, Klemt et al. developed K-nearest neighbor, support vector machine, elastic net regression, and ANN models for the prediction of 90-day readmissions following TKA using 24 input variables (Supplementary Table S2) [29]. The ANN was the best-performing model, with an associated AUC of 0.85 (Table 3). Using 33 input variables, Lee et al. developed a RUSBoost model for the prediction of 90-day readmissions following both TKA and THA [23]. This model demonstrated an accuracy of 87%, compared to an accuracy of 93% produced by the control logistic regression model. The recall rate of this control model, however, was significantly lower than that of the ML model, thus resulting in the ML model yielding an overall more reliable output.

Optimization

There has been a wide variety of optimization goals for the scheduling problems ranging from maximizing overall OR utilization to optimizing all processes of OR, recovery, and ward bed utilization (Table 4). Few studies have been conducted specifically targeting schedule optimization at the tactical level [39,43,46,49]. Adan et al. and Cardoen et al. utilized mixed integer programming to generate an optimized assignment of weekly OR schedules to maximize hospital bed capacity [39,46]. Adan et al. also used stochastic LOS times from a distribution of surgical groupings to improve the model by accounting for random variation [46].

Compared to using heuristic, or rules-based, optimization strategies alone, accounting for uncertain DOS times with the combination of stochastics rather than using an average time was found to significantly improve model performance [42,44]. However, other strategies such as tight clustering of surgeries with similar DOS have also been proven to be effective when combined with heuristic schedule optimization [40,45,48]. Patient and provider characteristics can also be considered in optimization problems. Silva et al. optimized OR utilization accounting for anaesthetist skill, while Wang et al. optimized scheduling while accounting for patient priority level [52,53].

Using discrete-event simulation, Lehtonen et al. identified that using a schedule with higher granularity improved OR utilization [40]. Similarly, Baesler et al. used discrete-event simulation to also account for preoperative and postoperative times in addition to DOS to optimize surgical scheduling [51]. Other simulation strategies such as Monte Carlo simulation and Markov decision process have been used to simulate the impact of different optimization strategies and determine the ideal number of required hospital resources from historical data prior to the implementation of change (Table 4).

Discussion

Given the rapid projected rise in costs associated with TJA, there exists an imminent need for the development and implementation of novel strategies aimed at improving hospital efficiency and optimizing resource utilization. Generally, the findings of this study support the use of ML for prediction modelling and surgical optimization in TJA. In the prediction of LOS, the ML models evaluated in this study performed superior to or at the same level as matched control models. With only 3 studies generating ML models for the prediction of DOS and 2 for hospital readmissions, further research is required to assess the performance of ML models for the prediction of these outcomes in particular. Preliminary models,

Table 4
Characteristics and summary findings of the included optimization studies.

Study	Decision level	Schedule optimization strategy	Optimization goal	Main findings
Denton et al., 2007	Operational level	Stochastic, heuristics	Minimize cost	Heuristic to sequence surgeons in order of increasing DOS variance and use of stochastic modelling to hedge against uncertain DOS times improves OR utilization.
Hans et al., 2008	Tactical level Operational level	Heuristics, Monte Carlo simulation	Minimize overtime	Clustering surgeries with a similar DOS and variability leads to reduced overtime and slack compared to base surgical plans generated by specialists.
Adan et al., 2009	Tactical level	MIP, stochastic	Minimize OR, ICU and ward bed overutilization and underutilization	Using MIP, can generate improved master surgical schedules by considering a stochastic LOS.
Lamiri et al., 2009	Operational level	Monte Carlo simulation, MIP, multiple heuristics	Minimize cost and overtime	Compared multiple optimization techniques. Combination of Monte Carlo simulation and MIP performed best and with least data.
Fei et al., 2009	Tactical level	Heuristics	Maximize OR utilization, minimize cost	Using a column-generation-based heuristic, cases are assigned to optimized ORs for the week, using an open scheduling strategy.
Cardoen et al., 2009	Tactical level Operational level	MIP	Maximize bed utilization	To determine the amount of OR time assigned to surgeons for outpatient surgery.
Marques et al., 2012	Operational level	Integer linear programming	Maximize OR utilization	Improvement in total OR utilization with reduction in length of surgical wait lists.
Lehtonen et al., 2013	Operational level	Discrete-event simulation	Maximize OR utilization	Improved DOS categorization and higher levels of schedule granularity (30 min vs 60 min) improve utilization.
M'Hallah et al., 2014	Operational level	Discrete-event simulation	Maximize OR utilization, minimize overtime	Cases grouped by mean DOS and OR utilization simulated. Recommends transfer of the last case in a busy room to a free one, group patient waitlists, and reduce workload by 10% or cancel last cases if planned overtime in schedule.
Van Huele et al., 2014	Tactical level Operational level	MIP	Minimize overtime	Evaluated the effect of certain surgeon constraints (surgeon availability, number of OR days/week, and consecutive surgeon hours and days) on performance of elective OR schedule.
Astaraky et al., 2015	Operational level	Heuristics, stochastic, Markov decision process	Minimize patient wait, overtime, and ward capacity	Improved surgical planning using combined model with stochastics over heuristics alone. Provides different schedules depending on hospital resource availability.
Baesler et al., 2015	Operational level	Heuristics, discrete-event simulation	Minimize total OR time	Account for surgery-grouping-specific preoperative, postoperative, setup, and recovery times. Combined heuristics with simulation to search for an optimal schedule.
Silva et al., 2015	Operational level	Heuristics, integer linear programming	Maximize OR utilization	Assign surgeries to maximize the OR utilization while matching surgeries to anaesthetist skills.
Wang et al., 2015	Operational level	Heuristics	Optimize number of ORs and PACU beds	Schedule patient surgeries based on priority where fixed resources are limited. But optimizes ORs and surgery allocation if flexibile.
Guido et al., 2017	Tactical level	Heuristics	Maximize number of surgeries	Assigns available OR time to surgeons while considering hospital objectives, surgery characteristics.
Zhang et al., 2019	Operational level	Stochastic, Markov decision process	Minimize cost	Combined Markov decision process and stochastic optimization lowered cost, shortened wait time, and improved OR and recovery bed utilization compared to stochastic optimization alone.
Bai et al., 2022	Operational level	Heuristic	Minimize OR idle time	Model can reduce total OR time while meeting resource (personnel and hospital) constraints. Connected stages of preop, OR, and recovery optimization.

ICU, intensive care unit; MIP, mixed integer programming; PACU, postoperative anesthesia care unit.

however, show promising outputs and encourage further investigation. As for surgical scheduling, a majority of optimization research for surgical scheduling has been targeted at the operational level. This is likely due to the complex and multifactorial nature of the problem, presenting as an ideal target for optimization compared to the distribution of yearly or weekly OR time among specialties. Various optimization strategies have been utilized to improve the efficiency of surgical scheduling, all of which improved the outcome of interest compared to traditional manual scheduling practices.

Potential clinical uses of these algorithms include the automation of surgical scheduling and improved utilization of hospital resources. This may lead to a reduction in cost and increased patient throughput. Targeted education, resources, and monitoring can be provided to those identified to be at high risk of readmission following TJA to reduce readmission events. Custom-bundled hospital compensation based on patient-specific predicted resource requirements may also be developed using more accurate models.

There exist countless variations of ML models that may be generated for the prediction of TJA LOS, DOS, and hospital readmissions. To avoid research waste, algorithms standing out above others must be identified, refined, validated, and ultimately implemented in clinical practice [57]. As a part of the fine-tuning process, input features may be adjusted to improve model performance, with more heavily weighted features in top-performing models being retained in future model iterations. In addition to patient factors, institutional factors, including surgeon, teaching status, geographic region, and location, as well as features derived from medical imaging may also be considered where sufficiently sized data sets are available. When considering the use of DL models vs ML models, the unique advantages and disadvantages of both must be considered. It is recognized that compared to conventional ML models, neural networks have the potential to yield superior results due to their ability to perform more sophisticated transformations of data [14,58]. For this to hold true, ANNs must be built using expansive data sets, requiring considerable investments in time and computational resources [59,60]. Another potential barrier to the utilization of ANNs is that they are less interpretable than typical ML models [61]. This is important as methods to evaluate these models must be developed to ensure bias (eg, based on certain patient characteristics) does not affect access to care if these models are to eventually be used to inform surgical scheduling.

All attempts to optimize scheduling rely on using an average DOS or LOS. Assumptions of DOS have a significant impact on OR underutilization and overtime [62,63]. Despite strategies to account for this by using stochastic methods of randomly sampling from a historical distribution, this is still a major limitation to the current attempts at optimization. A potentially effective strategy to improve scheduling may be to combine patient-specific ML predictions of DOS and downstream resource requirements, by predicting LOS, with optimization research.

A wide range of optimization goals were encountered in the literature depending on the generated mathematical problem. Most models were generated to optimize total OR utilization and minimize overtime and/or idle time, which may be the best metric when trying to maximize the utility of a finite resource. However, one ideal metric cannot be defined, as the solution to each OR optimization problem depends on the specific goals of the institution. Some hospitals may have incentive to maximize the throughput of cases, reducing the size of surgical waitlists. Without appropriate constraints, this could bias the optimization models to select cases with a shorter predicted DOS and impact care for patients with more complex needs. Most research in OR optimization has focused on a daily planning horizon. This focused the problem on direct patient care and not at decisions made involving stakeholders from multiple different specialties. However, optimizing OR utilization at a higher level, based on overall hospital resources and demand, may be important and reveal greater room for improvement.

This present review is not without limitations. The quality of the included studies was not formally assessed due to the lack of a standardized risk-of-bias assessment tool for studies generating ML-based predictive models; however, this tool is currently under development [64]. This can be attributed to the relatively recent rise in popularity of ML in predictive analytics within medicine, resulting in a paucity of defined criteria describing best practices in model generation and evaluation. For example, it remains unclear as to what performance metrics to report and the minimum required sample size to train a robust model [65]. Despite there being no formal assessment of the quality of the included studies, many gross shortcomings are observed. First, the details of the model-generation process, most notably, validation size and/or method, were not described for 12 of the presented ML models. This calls into question the veracity of the results described in these studies [66]. Furthermore, only 2 of the 22 generated models externally validated their models on test sets from sources unique to those used during training and validation. Beyond the limitations associated with the quality of the included studies, an additional shortcoming of this present review lies in the lack of quantitative comparison of the predictive abilities of ML models vs control models. This is due to the variable use of metrics describing the predictive abilities of models generated across studies. Many studies reported AUC; however, this is a classification metric based on varying arbitrary cutoffs (eg, LOS of <2 days or ≥ 2 days vs <3 days or ≥ 3 days), which should be avoided for continuous outcomes [67]. Finally, synthesizing optimization research to improve elective surgical scheduling is a challenge as unique solutions may be required to address different goals, institutional constraints, and data sets. As such, optimization results can only be compared to local historical performance or within a limited context.

Conclusions

With TJA costs expected to rise rapidly in association with an increasing demand, there exists an imminent need for the development and implementation of novel strategies aimed at improving hospital efficiency and optimizing resource utilization. With increasing access to big data in addition to technological advances in AI reaching new heights, applications of ML within medicine are becoming increasingly feasible and gaining notable popularity. High-performing ML models have been developed for predictive analytics in TJA, as have mathematical strategies for surgical scheduling optimization. While there remains work to be done in refining these tools, there exists considerable opportunities for improved efficiency in resource utilization surrounding TJA, especially when considering the combined utilization of predictive modelling with optimization strategies.

Conflicts of interest

Dr. J. I. Wolfstadt is in the speakers' bureau of or gave paid presentations for Depuy-Synthes, is a paid consultant for Microport Orthopaedics, is the American Association of Hip and Knee Surgeons Young Arthroplasty Group Vice Chair, a Canadian Arthroplast Society Education Committee member, and is a Canadian Orthopaedic Association Standards Committee member. Dr. J. R. Lex serves on the Resident Advisory Board for PrecisionOS Technologies. The other authors declare no potential conflicts of interest.

For full disclosure statements refer to https://doi.org/10.1016/j.artd.2023.101116.

Consent to publish

All authors have read the final manuscript and consent to publication.

Author contributions

Bahar Entezari: Methodology, Writing—original draft, Writing—review & editing, Visualization. Johnathan R. Lex: Conceptualization, Methodology, Writing—original draft, Writing—review & editing. Robert Koucheki: Methodology, Data curation, Writing—review & editing, Visualization. Aazad Abbas: Methodology, Writing—Review & Editing. Jay Toor: Supervision, Writing—review and editing. Jesse I. Wolfstadt: Supervision, Writing—review and editing. Bheeshma Ravi: Supervising, Writing—review & editing. Cari Whyne: Conceptualization, Supervision, Writing—review & editing.

Data availability

The data collected were from publicly available journal articles.

References

[1] HCUP Fast Stats. Healthcare cost and utilization project (HCUP). Rockville (MD): Agency for Healthcare Research and Quality; 2021.
[2] Kremers HM, Larson DR, Crowson CS, Kremers WK, Washington RE, Steiner CA, et al. Prevalence of total hip and knee replacement in the United States. J Bone Joint Surg Am 2014;97:1386–97. https://doi.org/10.2106/JBJS.N.01141.
[3] Singh JA, Yu S, Chen L, Cleveland JD, Singh JA, Yu S, et al. Rates of total joint replacement in the United States: future projections to 2020 – 2040 using the national inpatient sample. J Rheumatol 2020;46:1134–40. https://doi.org/10.3899/jrheum.170990.
[4] Inacio MCS, Paxton EW, Graves SE, Namba RS, Nemes S. Projected increase in total knee arthroplasty in the United States – an alternative projection model. Osteoarthritis Cartilage 2017;25:1797–803. https://doi.org/10.1016/j.joca.2017.07.022.
[5] Klug A, Gramlich Y, Rudert M, Drees P, Hoffmann R, Weißenberger M, et al. The projected volume of primary and revision total knee arthroplasty will place an immense burden on future health care systems over the next 30 years. Knee Surg Sports Traumatol Arthrosc 2021;29:3287–98. https://doi.org/10.1007/s00167-020-06154-7.
[6] Sloan M, Premkumar A, Sheth NP. Projected volume of primary total joint arthroplasty in the U.S., 2014 to 2030. J Bone Joint Surg Am 2018;100:1455–60. https://doi.org/10.2106/JBJS.17.01617.
[7] Weeks WB, Schoellkopf WJ, Ballard DJ, Kaplan GS, James B, Weinstein JN. Episode-of-Care characteristics and costs for hip and knee replacement surgery in hospitals belonging to the high value healthcare collaborative compared with similar hospitals in the same health care markets. Med Care 2017;55:583–9. https://doi.org/10.1097/MLR.0000000000000710.
[8] BlueCross, BlueShield. A study of cost variations for knee and hip replacement surgeries in the U.S. 2015. https://www.bcbs.com/the-health-of-america/reports/study-of-cost-variations-knee-and-hip-replacement-surgeries-the-us [accessed 20.11.22].
[9] Batailler C, Shatrov J, Sappey-Marinier E, Servien E, Parratte S, Lustig S. Artificial intelligence in knee arthroplasty: current concept of the available clinical applications. Arthroplasty 2022;4:17. https://doi.org/10.1186/s42836-022-00119-6.
[10] Yang CC. Explainable artificial intelligence for predictive modeling in healthcare. J Healthc Inform Res 2022;6:228–39. https://doi.org/10.1007/s41666-022-00114-1.
[11] Chen JH, Asch SM. Machine learning and prediction in medicine — beyond the peak of inflated expectations. N Engl J Med 2017;376:2507–9. https://doi.org/10.1056/NEJMp1702071.
[12] Gangal A, Kumar P, Kumari S, Saini A. Prediction models for healthcare using machine learning. Hershey, PA: IGI Global; 2021. p. 70–91.
[13] Erickson BJ. Basic artificial intelligence techniques: machine learning and deep learning. Radiol Clin North Am 2021;59:933–40. https://doi.org/10.1016/j.rcl.2021.06.004.
[14] Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Peter Campbell J. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol 2020;9:1–12. https://doi.org/10.1167/tvst.9.2.14.
[15] Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? J Arthroplasty 2018;33:2358–61. https://doi.org/10.1016/j.arth.2018.02.067.
[16] Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol 2022;23:40–55. https://doi.org/10.1038/s41580-021-00407-0.
[17] van Essen JT, Hans EW, Hurink JL, Oversberg A. Minimizing the waiting time for emergency surgery. Oper Res Heal Care 2012;1:34–44. https://doi.org/10.1016/j.orhc.2012.05.002.
[18] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. https://doi.org/10.1136/bmj.n71.
[19] Navarro SM, Wang EY, Haeberle HS, Mont MA, Krebs VE, Patterson BM, et al. Machine learning and primary total knee arthroplasty: patient forecasting for a patient-specific payment model. J Arthroplasty 2018;33:3617–23. https://doi.org/10.1016/j.arth.2018.08.028.
[20] Ramkumar PN, Karnuta JM, Navarro SM, Haeberle HS, Scuderi GR, Mont MA, et al. Deep learning preoperatively predicts value metrics for primary total knee arthroplasty: development and validation of an artificial neural network model. J Arthroplasty 2019;34:2220–2227.e1. https://doi.org/10.1016/j.arth.2019.05.034.
[21] Ramkumar PN, Karnuta JM, Navarro SM, Haeberle HS, Iorio R, Mont MA, et al. Preoperative prediction of value metrics and a patient-specific payment model for primary total hip arthroplasty: development and validation of a deep learning model. J Arthroplasty 2019;34:2228–2234.e1. https://doi.org/10.1016/j.arth.2019.04.055.
[22] Ramkumar PN, Navarro SM, Haeberle HS, Karnuta JM, Mont MA, Iannotti JP, et al. Development and validation of a machine learning algorithm after primary total hip arthroplasty: applications to length of stay and payment models. J Arthroplasty 2019;34:632–7. https://doi.org/10.1016/j.arth.2018.12.030.
[23] Lee HK, Jin R, Feng Y, Bain PA, Goffinet J, Baker C, et al. An analytical framework for TJR readmission prediction and cost-effective intervention. IEEE J Biomed Health Inform 2019;23:1760–72. https://doi.org/10.1109/JBHI.2018.2859581.
[24] Gabriel RA, Doan CN, Jiang X, Vaida F. A predictive model for determining patients not requiring prolonged hospital length of stay after elective primary total hip. Arthroplasty 2019;129:43–50. https://doi.org/10.1213/ANE.0000000000003798.
[25] Wei C, Quan T, Wang KY, Gu A, Fassihi SC, Kahlenberg CA, et al. Artificial neural network prediction of same-day discharge following primary total knee arthroplasty based on preoperative and intraoperative variables. Bone Joint J 2021;103-B:1358–66. https://doi.org/10.1302/0301-620X.103B8.BJJ-2020-1013.R2.
[26] Han C, Liu J, Wu Y, Chong Y, Chai X, Weng X. To predict the length of hospital stay after total knee arthroplasty in an orthopedic center in China: the use of machine learning algorithms. Front Surg 2021;8:1–11. https://doi.org/10.3389/fsurg.2021.606038.
[27] Kugelman D, Huang S, Teo G, Doran M, Singh V, Buchalter D, et al. A novel machine learning predictive tool assessing outpatient or inpatient designation for medicare patients undergoing total knee arthroplasty. Arthroplast Today 2022;13:120–4. https://doi.org/10.1016/j.artd.2021.12.004.
[28] Kugelman DN, Teo G, Huang S, Doran MG, Singh V, Long WJ. A novel machine learning predictive tool assessing outpatient or inpatient designation for medicare patients undergoing total hip arthroplasty. Arthroplast Today 2021;8:194–9. https://doi.org/10.1016/j.artd.2021.03.001.
[29] Klemt C, Tirumala V, Habibi Y, Buddhiraju A, Chen TLW, Kwon YM. The utilization of artificial neural networks for the prediction of 90-day unplanned readmissions following total knee arthroplasty. Arch Orthop Trauma Surg 2022. https://doi.org/10.1007/s00402-022-04566-3 [Epub ahead of print].
[30] Yeo I, Klemt C, Melnic CM, Pattavina MH, De Oliveira BMC, Kwon YM. Predicting surgical operative time in primary total knee arthroplasty utilizing machine learning models. Arch Orthop Trauma Surg 2022. https://doi.org/10.1007/s00402-022-04588-x [Epub ahead of print].
[31] Klemt C, Tirumala V, Barghi A, Cohen-Levy WB, Robinson MG, Kwon YM. Artificial intelligence algorithms accurately predict prolonged length of stay following revision total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 2022;30:2556–64. https://doi.org/10.1007/s00167-022-06894-8.
[32] Lopez CD, Ding J, Trofa DP, Cooper HJ, Geller JA, Hickernell TR. Machine learning model developed to aid in patient selection for outpatient total joint arthroplasty. Arthroplast Today 2022;13:13–23. https://doi.org/10.1016/j.artd.2021.11.001.
[33] Abbas A, Mosseri J, Lex JR, Toor J, Ravi B, Khalil EB, et al. Machine learning using preoperative patient factors can predict duration of surgery and length of stay for total knee arthroplasty. Int J Med Inform 2022;158:104670. https://doi.org/10.1016/j.ijmedinf.2021.104670.
[34] Motesharei A, Batailler C, De Massari D, Vincent G, Chen AF, Lustig S. Predicting robotic-assisted total knee arthroplasty operating time benefits of machine-learning and 3D patient-specific data. Bone Jt Open 2022;3:383–9. https://doi.org/10.1302/2633-1462.35.BJO-2022-0014.R1.
[35] Zalikha AK, El-Othmani MM, Shah RP. Predictive capacity of four machine learning models for in-hospital postoperative outcomes following total knee arthroplasty. J Orthop 2022;31:22–8. https://doi.org/10.1016/j.jor.2022.03.006.
[36] Trunfio TA, Borrelli A, Improta G. Is it possible to predict the length of stay of patients undergoing hip-replacement surgery? Int J Environ Res Public Health 2022;19:1–16. https://doi.org/10.3390/ijerph19106219.
[37] Li H, Jiao J, Zhang S, Tang H, Qu X, Yue B. Construction and comparison of predictive models for length of stay after total knee arthroplasty: regression model and machine learning analysis based on 1,826 cases in a single Singapore center. J Knee Surg 2022;35:7–14. https://doi.org/10.1055/s-0040-1710573.
[38] Johannesdottir KB, Kehlet H, Petersen PB, Aasvang EK, Sørensen HBD, Jørgensen CC. Machine learning classifiers do not improve prediction of hospitalization > 2 days after fast-track hip and knee arthroplasty compared with a classical statistical risk model. Acta Orthop 2022;93:117–23. https://doi.org/10.2340/17453674.2021.843.
[39] Cardoen B, Demeulemeester E, Beliën J. Sequencing surgical cases in a day-care environment: an exact branch-and-price approach. Comput Oper Res 2009;36:2660–9. https://doi.org/10.1016/j.cor.2008.11.012.
[40] Lehtonen JM, Torkki P, Peltokorpi A, Moilanen T. Increasing operating room productivity by duration categories and a newsvendor model. Int J Health Care Qual Assur 2013;26:80–92. https://doi.org/10.1108/09526861311297307.
[41] Van Huele C, Vanhoucke M. Analysis of the integration of the physician rostering problem and the surgery scheduling problem topical collection on systems-level quality improvement. J Med Syst 2014;38:43. https://doi.org/10.1007/s10916-014-0043-z.
[42] Astaraky D, Patrick J. A simulation based approximate dynamic programming approach to multi-class, multi-resource surgical scheduling. Eur J Oper Res 2015;245:309–19. https://doi.org/10.1016/j.ejor.2015.02.032.
[43] Guido R, Conforti D. A hybrid genetic approach for solving an integrated multi-objective operating room planning and scheduling problem. Comput Oper Res 2017;87:270–82. https://doi.org/10.1016/j.cor.2016.11.009.
[44] Denton B, Viapiano J, Vogl A. Optimization of surgery sequencing and scheduling decisions under uncertainty. Health Care Manag Sci 2007;10:13–24. https://doi.org/10.1007/s10729-006-9005-4.

[45] Hans E, Wullink G, van Houdenhoven M, Kazemier G. Robust surgery loading. Eur J Oper Res 2008;185:1038–50. https://doi.org/10.1016/j.ejor.2006.08.022.
[46] Adan I, Bekkers J, Dellaert N, Vissers J, Yu X. Patient mix optimisation and stochastic resource requirements: a case study in cardiothoracic surgery planning. Health Care Manag Sci 2009;12:129–41. https://doi.org/10.1007/s10729-008-9080-9.
[47] Lamiri M, Grimaud F, Xie X. Optimization methods for a stochastic surgery planning problem. Int J Prod Econ 2009;120:400–10. https://doi.org/10.1016/j.ijpe.2008.11.021.
[48] M'Hallah R, Al-Roomi AH. The planning and scheduling of operating rooms: a simulation approach. Comput Ind Eng 2014;78:235–48. https://doi.org/10.1016/j.cie.2014.07.022.
[49] Fei H, Chu C, Meskens N. Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. Ann Oper Res 2009;166:91–108. https://doi.org/10.1007/s10479-008-0413-3.
[50] Marques I, Captivo ME, Pato MV. An integer programming approach to elective surgery scheduling. OR Spectr 2012;34:407–27. https://doi.org/10.1007/s00291-011-0279-7.
[51] Baesler F, Gatica J, Correa R. Simulation optimisation for operating room scheduling. Int J Simul Model 2015;14:215–26. https://doi.org/10.2507/IJSIMM14(2)3.287.
[52] Silva TAO, De Souza MC, Saldanha RR, Burke EK. Surgical scheduling with simultaneous employment of specialised human resources. Eur J Oper Res 2015;245:719–30. https://doi.org/10.1016/j.ejor.2015.04.008.
[53] Wang Y, Tang J, Pan Z, Yan C. Particle swarm optimization-based planning and scheduling for a laminar-flow operating room with downstream resources. Soft Comput 2015;19:2913–26. https://doi.org/10.1007/s00500-014-1453-z.
[54] Zhang J, Dridi M, El Moudni A. A two-level optimization model for elective surgery scheduling with downstream capacity constraints. Eur J Oper Res 2019;276:602–13. https://doi.org/10.1016/j.ejor.2019.01.036.
[55] Bai X, Zhang W, Luo L, Ma H, Zhu T. Day surgery scheduling and optimization in large public hospitals in China: a three-station job shop scheduling problem. J Healthc Eng 2022;2022:1149657. https://doi.org/10.1155/2022/1149657.
[56] Helm JM, Swiergosz AM, Haeberle HS, Karnuta JM, Schaffer JL, Krebs VE, et al. Machine learning and artificial intelligence: definitions, applications, and future directions. Curr Rev Musculoskelet Med 2020;13:69–76. https://doi.org/10.1007/s12178-020-09600-8.
[57] Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ 2021;375:n2281. https://doi.org/10.1136/bmj.n2281.
[58] Olczak J, Pavlopoulos J, Prijs J, Ijpma FFA, Doornberg JN, Lundström C, et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. Acta Orthop 2021;92:513–25. https://doi.org/10.1080/17453674.2021.1918389.
[59] Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci 2021;2:420. https://doi.org/10.1007/s42979-021-00815-1.
[60] Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. AMIA Jt Summits Transl Sci Proc 2020;2020:191–200.
[61] Meng C, Trinh L, Xu N, Enouen J, Liu Y. Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset. Sci Rep 2022;12:7166. https://doi.org/10.1038/s41598-022-11012-2.
[62] Guda H, Dawande M, Janakiraman G, Jung KS. Optimal policy for a stochastic scheduling problem with applications to surgical scheduling. Prod Oper Manag 2016;25:1194–202. https://doi.org/10.1111/poms.12538.
[63] Ogulata SN, Erol R. A hierarchical multiple criteria mathematical programming approach for scheduling general surgery operations in large hospitals. J Med Syst 2003;27:259–70. https://doi.org/10.1023/A:1022575412017.
[64] Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021;11:e048008. https://doi.org/10.1136/bmjopen-2020-048008.
[65] Steyerberg EW. Overfitting and optimism in prediction models; 2009. p. 83–100. https://doi.org/10.1007/978-0-387-77244-8_5.
[66] Ho SY, Phua K, Wong L, Bin Goh WW. Extensions of the external validation for checking learned model interpretability and generalizability. Patterns 2020;1:100129. https://doi.org/10.1016/j.patter.2020.100129.
[67] Altman DG, Royston P. The cost of dichotomising continuous variables. Br Med J 2006;332:1080. https://doi.org/10.1136/bmj.332.7549.1080.

Supplementary Table S1
Fifteen most frequently used input features included among predictive modeling studies.

Study, year	Input features														
	Age	Gender/sex	BMI/obesity	Ethnicity/race	Diabetes	HTN	Anesthetic	Smoking	CVD	ASA score	CCI	Hb/HCT/anemia	Pulmonary disease/COPD	Neoplastic disease	CKD/dialysis
Navarro, 2018	✓	✓		✓							✓				
Ramkumar, 2019	✓	✓		✓											
Ramkumar, 2019	✓	✓		✓											
Ramkumar, 2019	✓	✓		✓							✓				
Lee, 2019	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓		
Gabriel 2019	✓	✓	✓			✓	✓					✓	✓		
Wei, 2021	✓	✓	✓	✓			✓					✓	✓		
Han, 2021	✓	✓	✓		✓	✓		✓	✓			✓	✓	✓	✓
Kugelman, 2021	✓	✓	✓		✓				✓	✓	✓				
Yeo, 2022	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓
Klemt, 2022	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓
Lopez, 2022	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓
Abbas, 2022	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
Motesharei, 2022	✓	✓													
Zalikha, 2022	✓	✓		✓											
Johannesdottir, 2022	✓	✓	✓			✓		✓	✓			✓	✓	✓	✓
Klemt, 2022	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓
Li, 2022	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓			
Kugelman, 2022	✓	✓	✓		✓				✓	✓	✓				
Trunfio, 2022	✓	✓	✓		✓	✓			✓			✓	✓	✓	
Total Number of Studies	20	20	14	12	11	10	9	9	9	9	8	8	8	8	7

ASA, American Society of Anesthesiologists; BMI, body mass index; CCI, Charlson Comorbidity Index; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; CVD, cardiovascular disease; Hb, hemoglobin; HCT, hematocrit; HTN, hypertension.

Supplementary Table S2
Control and machine learning algorithms generated among predictive modeling studies.

Study, year	Control algorithms			ML algorithms														
	Mean regressor	Linear reg	Logistic reg	Bayesian	KNN	SVM	SDG	RFC	DT	GBDT	XGB	ADB	CB	RUSBt	Ridge reg	Lasso reg	Elastic net reg	ANN
Navarro, 2018				✓														
Ramkumar, 2019																		✓
Ramkumar, 2019																		✓
Ramkumar, 2019				✓														
Lee, 2019			✓						✓					✓				
Gabriel 2019			✓					✓							✓	✓		
Wei, 2021			✓															✓
Han, 2021			✓	✓	✓			✓	✓	✓	✓	✓						✓
Kugelman, 2021						✓		✓			✓					✓		
Yeo, 2022					✓			✓										✓
Klemt, 2022						✓											✓	✓
Lopez, 2022																		✓
Abbas, 2022	✓	✓			✓	✓	✓	✓	✓		✓	✓					✓	✓
Motesharei, 2022		✓						✓		✓			✓					
Zalikha, 2022						✓		✓			✓							✓
Johannesdottir, 2022			✓	✓		✓		✓										
Klemt, 2022			✓		✓	✓											✓	✓
Li, 2022			✓								✓							
Kugelman, 2022						✓		✓			✓					✓		
Trunfio, 2022		✓						✓		✓	✓							

ADB, AdaBoost; BMI, body mass index; CB, CatBoost; DT, decision tree; GBDT, gradient boosted decision tree; KNN, k-nearest neighbor; RFC, random forest classifier; reg, regression; RUSB, RUSBoost SDG, stochastic gradient descent; SVM, support vector machine; XGB, XGBoost.