

# Revolutionizing total hip arthroplasty: The role of artificial intelligence and machine learning

Umile Giuseppe Longo<sup>1,2</sup> | Sergio De Salvatore<sup>2,3</sup> | Alice Piccolomini<sup>2</sup> |  
Nathan Samuel Ullman<sup>2</sup> | Giuseppe Salvatore<sup>1,2</sup> | Margaux D'Hooghe<sup>4</sup> |  
Maristella Saccomanno<sup>5,6</sup> | Kristian Samuelsson<sup>7,8,9</sup>  | Rocco Papalia<sup>1,2</sup> |  
Ayoosh Pareek<sup>10</sup>

<sup>1</sup>Fondazione Policlinico Universitario Campus Bio-Medico, Roma, Italy

<sup>2</sup>Research Unit of Orthopaedic and Trauma Surgery, Department of Medicine and Surgery, Università Campus Bio-Medico di Roma, Roma, Italy

<sup>3</sup>Orthopedic Unit, Department of Surgery, Bambino Gesù Children's Hospital, Palidoro, Rome, Italy

<sup>4</sup>Department of Medicine, University of Navarra, Pamplona, Spain

<sup>5</sup>Department of Medical and Surgical Specialties, Radiological Sciences, and Public Health, University of Brescia, Brescia, Italy

<sup>6</sup>Department of Bone and Joint Surgery, Speciali Civili, Brescia, Italy

<sup>7</sup>Sahlgrenska Sports Medicine Center, Gothenburg, Sweden

<sup>8</sup>Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

<sup>9</sup>Department of Orthopaedics, Sahlgrenska University Hospital, Mölndal, Sweden

<sup>10</sup>Hospital for Special Surgery, New York, New York, USA

## Correspondence

Kristian Samuelsson, Sahlgrenska Sports Medicine Center, Gothenburg, Sweden.  
Email: [kristian@samuelsson.cc](mailto:kristian@samuelsson.cc)

## Abstract

**Purpose:** There has been substantial growth in the literature describing the effectiveness of artificial intelligence (AI) and machine learning (ML) applications in total hip arthroplasty (THA); these models have shown the potential to predict post-operative outcomes using algorithmic analysis of acquired data and can ultimately optimize clinical decision-making while reducing time, cost and complexity. The aim of this review is to analyze the most updated articles on AI/ML applications in THA as well as present the potential of these tools in optimizing patient care and THA outcomes.

**Methods:** A comprehensive search was completed through August 2024, according to the PRISMA guidelines. Publications were searched using the Scopus, Medline, EMBASE, CENTRAL and CINAHL databases. Pertinent findings and patterns in AI/ML methods utilization, as well as their applications, were quantitatively summarized and described using frequencies, averages and proportions. This study used a modified eight-item Methodological Index for Non-Randomized Studies (MINORS) checklist for quality assessment.

**Results:** Nineteen articles were eligible for this study. The selected studies were published between 2016 and 2024. Out of the various ML algorithms, four models have proven to be particularly significant and were used in almost 20% of the studies, including elastic net penalized logistic regression, artificial neural network, convolutional neural network (CNN) and multiple linear regression. The highest area under the curve (=1) was reported in the preoperative planning outcome variable and utilized CNN. All 20 studies demonstrated a high level of quality and low risk of bias, with a modified MINORS score of at least 7/8 (88%).

**Abbreviations:** AI, artificial intelligence; ANN, artificial neural network; ASA, American Society of Anesthesiologists; AUC, area under the curve; BMI, body mass index; CNN, convolutional neural network; COPD, chronic obstructive pulmonary disease; CSO, clinically significant outcome; ENPLR, elastic net penalized logistic regression; HJC, hip joint centre; HOA, hip osteoarthritis; HOOS, Hip disability and osteoarthritis outcome score; HRQoL, Health-related quality of life; JSN, Joint space narrowing; KL, Kellgren and Lawrence; LOS, length of stay; MARS, multivariate adaptive regression splines; MeSH, medical subject headings; METs, metabolic equivalent of task; MINORS, Methodological Index for Non-Randomized Studies; ML, machine Learning; MLR, multiple linear regression; OECD, Organization for Economic Co-operation and Development; OST, osteophyte score; PRHS, patient-reported health state; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-analyses; PROM, patient-reported outcome measure; rHOA, radiographic hip osteoarthritis; SGB, stochastic gradient boosting; SVM, support vector machine; THA, total hip arthroplasty; XGBoost, extreme gradient boosting.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Journal of Experimental Orthopaedics* published by John Wiley & Sons Ltd on behalf of European Society of Sports Traumatology, Knee Surgery and Arthroscopy.

**Conclusions:** Developments in AI/ML prediction models in THA are rapidly increasing. There is clear potential for these tools to assist in all stages of surgical care as well as in challenges at the broader hospital administrative level and patient-specific level.

**Level of Evidence:** Level III.

#### KEYWORDS

artificial intelligence, deep learning, machine learning, orthopaedic surgery, total hip arthroplasty

## INTRODUCTION

Life expectancy has increased over the past few decades, and total hip arthroplasty (THA) utilization rates have grown accordingly. A 16% increase in THA has been observed in patients aged 65 years or older; this same population has also grown by 12%. Between 2015 and 2050, some European countries, as well as Australia, are projected to experience a 95%–120% increase in THA. The future demand for THA will increase globally, creating a need for modern solutions to healthcare obstacles [31–33, 35, 36, 45].

Artificial intelligence (AI) broadly refers to the ability of computers and machines to simulate human intelligence, mainly decision-making and problem-solving abilities, by using algorithms to process large amounts of data [3]. Machine learning (ML), a branch of AI, is a more autonomous process split into two phases: a teaching phase, where the machine is provided with data to find patterns, and an inference phase, where the machine can infer outcomes based on the previous patterns found [3].

Applying AI's predictive power to THA can ultimately optimize clinical decision-making while reducing time, cost and complexity [49]. The integration of AI and ML technology into THA can sharpen the field of orthopaedics by enabling more accurate clinical prediction and decision-making during pre- and post-operative management [33, 44, 46].

AI/ML tools can be applied at an early stage for patient selection and planning, as well as predicting readmission and reoperation risk. During revision arthroplasty preoperative planning, neural networks, a subset of ML, can be used to identify failed implants [6]. At the post-operative stage, these tools can predict post-operative complications, patient-reported outcomes, pain and even prolonged opioid use [12, 34].

As a result, there has been substantial growth in the literature describing the effectiveness of AI and ML applications in THA. These models have shown the potential to predict post-operative outcomes using algorithmic analysis of acquired data [46]. Lalehzarian et al. reported that decision-making models could optimize patient satisfaction based on many factors,

including age, gender and presence of comorbidities, with the advantage of assessing raw inputs and identifying the most relevant features for analysis [25]. Overall, the utility of AI and ML in orthopaedic surgery continues to expand thanks to the aforementioned benefits as well as the ability to identify patients who are more predisposed to post-operative complications, thus allowing more targeted therapeutics for better quality outcomes in patients. Given these promising results, it is critical to build an accurate understanding of the implementation of the current models and further develop their applicability in the future [24].

Considering the developments of AI/ML and the increasing demand for THA, it is important to have an up-to-date framework of the literature on this topic. Therefore, the purpose of this review is to analyze the most updated articles on AI/ML applications in THA and present the potential of these tools in optimizing patient care and THA outcomes.

## METHODS

### Study selection

The formulation of the research question was done using a PIOS-approach: Population (P); Intervention (I); Outcome (O) and Study design (S). This systematic review collected and analyzed data on patients who underwent THA (P). The AI/ML methods (I) that processed the patient data to form models which predict preoperative, perioperative and post-operative outcomes, complications and related costs (O), were reviewed. The following study designs were included (S): retrospective cohort study (RC), prospective cohort study (PC), diagnostic study (DS) and prognostic study (PS).

### Eligibility criteria

Inclusion criteria comprised original clinical studies, including studies which evaluate AI/ML-based application in clinical decision-making in hip arthroplasty.

Exclusion criteria comprised studies that did not evaluate AI/ML applications in hip arthroplasty, medical imaging analysis studies without explicit reference or application to hip arthroplasty, studies with nonhuman subjects, non-English-language studies, inaccessible articles, conference abstracts, reviews and editorials. No limits were placed on the level of evidence or timing of the study because the majority of the reviewed studies were published within the last 10 years.

## Search

A comprehensive search was performed from the inception of the databases included to August 2024, according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines. Publications were searched using the Scopus, Medline, EMBASE, CENTRAL and CINAHL databases, using the following Medical subject headings (MeSH) terms or keywords: 'hip osteoarthritis', 'replacement', 'prosthesis', 'artificial intelligence', 'algorithm', 'machine learning', 'complications', 'blood loss', 'blood transfusion', 'length of stay', 'hospitalisation', 'costs', 'economic analyses', 'functional outcomes', 'ROM', 'revision', 'revision rate' and 'surgical technique'. Keywords were used both isolated and combined. Additional studies listed in selected systematic review reference lists were searched. In order to optimize the search, boolean operators (OR and AND) were used.

## Data collection process

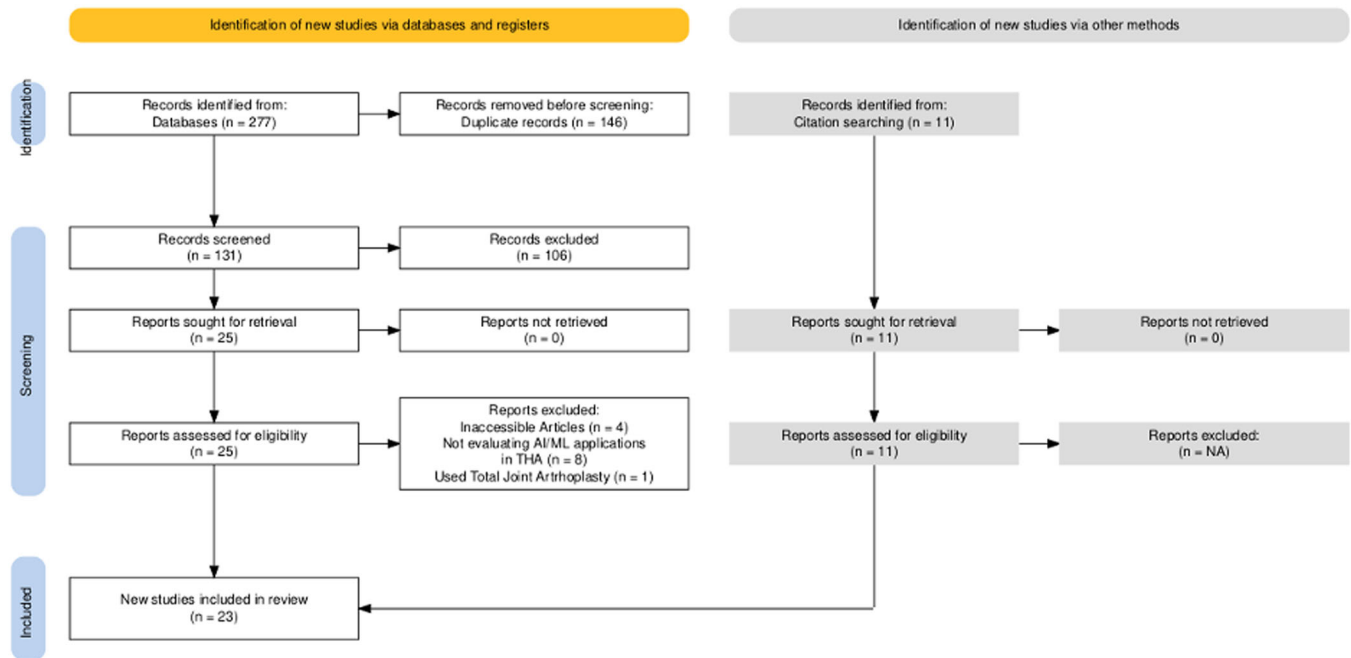
Two independent reviewers (AP and NU) screened each individual article in two stages. Initially, article titles and abstracts were screened, followed by a full-text review of the selected articles. Any disagreements were resolved by consulting with a third reviewer (SDS). Inclusion and exclusion of the reviewed articles are reported below by the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flowchart found in Figure 1.

## Data items

General study characteristics of the selected publications were extracted and organized into a database, which included the primary author, year of publication, study design, level of evidence, study duration, AI/ML methods, data sources, input and output variables, sample size of patients in testing sets, average patient age and percentage female patients.

## Data analysis

Pertinent findings and patterns in AI/ML methods utilization, as well as their applications, were quantitatively summarized and described using frequencies, averages and proportions. Brier score and  $R^2$  values were not included in this review. The reported metrics for



**FIGURE 1** PRISMA diagram for systematic review data collection process. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-analyses.

each individual AI/ML method, and its corresponding application, were assessed. Metrics used to summarize the performance of reported AI/ML methods include area under the curve (AUC) of receiver operating characteristic curves, accuracy (%), sensitivity (%) and specificity (%).

Input and output variables used in the ML models from the reviewed articles were analyzed as well. Input variables, otherwise known as predictors, are measurable factors which are inputted into the algorithms. These variables can include patient demographics, preoperative clinical data, radiographic measurements and laboratory results. Output variables, also known as target variables or outcomes, are the predictions or results generated by the algorithm. These variables can include post-operative outcomes, cost prediction and readmission. This system of inputs and outputs allows complex patterns and interdependencies to be found; leveraging these associations then enables the models to predict outcomes of interest when presented with new input data [3].

## Risk of bias assessment

This study used a modified eight-item Methodological Index for Non-Randomized Studies (MINORS) checklist for quality assessment. The eight items included: disclosure, study aim, input feature, output feature, validation method, data set distribution, performance metric and AI model. A binary scale was used to score each item: 0 (*not reported or unclear*) and 1 (*reported and clear*). Reviewers AP and NU individually assessed and scored each item. These classifications create a more accurate scoring tool for quality assessments, specifically for publications on diagnostic studies and AI healthcare applications, compared to the original MINORS checklist [26, 51].

## RESULTS

### Search results

A flow-chart diagram reported the selection process of the studies according to the PRISMA protocol (Figure 1). A total of 277 studies were identified through predefined search terms, with a total of 131 records maintained after duplicate removal. Of those, 106 were excluded from the study through title and abstract screening because they were not inherent to AI/ML applications or not focused on hip joint arthroplasty. Then, 25 full-text articles were screened. Of these studies, 12 met full inclusion criteria. Through further screening of the bibliography of the included records, 11 additional studies were identified and reported. Ultimately, a total of 23 articles were eligible for this study.

## Study characteristics

The majority of the articles were retrospective cohort studies, accounting for over 68% of the cohort, however prognostic studies (5%), prospective cohort studies (21%) and diagnostic studies (5%) were also included. Level of evidence of the reviewed studies ranged from I to IV. Of the selected studies, the following levels of evidence and study designs were included: 1 Level I prognostic study, 3 Level II prospective studies, 17 Level III retrospective studies and 2 Level IV diagnostic study. The selected studies were published between 2016 and 2024. The average number of patients included in the testing model was 10,738 (standard deviation [SD] = 20.45), with an average patient age of 62.6 years. On average, 61.2% of the patients were females. The study by Ramkumar et al. reported the highest number of patients ( $n = 78,335$ ) [47]. Of the 20 studies that were analyzed in this review, 15 were single-centred studies, while 5 were multicentered. The aforementioned data can be found in Table 1.

Out of the various ML algorithms, four models have proven to be particularly significant and were used in almost 20% of the studies: elastic net penalized logistic regression (ENPLR), artificial neural network (ANN), convolutional neural network (CNN) and multiple linear regression (MLR).

## OUTCOME VARIABLES

The following sections are the outcome variables that are reported in Table 2.

Under each outcome variable, there is included the specific ML methods used to predict the outcome variables, prediction model performance metric values and how many studies developed prediction models for the specific outcome variable, all of which are reported in Table 3.

### Risk of THA

Two studies explored the necessity of THA in patients with OA based on preoperative radiographs [7, 8].

### Preoperative planning

Four of the seven studies reporting this outcome relied on CNN, one of which had an accuracy of 1, while the other studies used ENPLR, ANN and stochastic gradient boosting (SGB) reporting an AUC of 0.71, 0.84 and 0.77, respectively [5, 7, 14, 16, 17, 19, 20, 50, 53].

**TABLE 1** Study characteristics.

Authors and Year	Country	Study design	Level of evidence	Follow-up	Sample size	Mean patient age	% Female patients
Alam et al., 2019 [1]	UK	RCS	III	–	10	–	–
Borjali et al., 2020 [5]	USA	RCS	III	1 year	25	61.3	47%
Chen et al., 2023 [7]	Taiwan	DS	IV	3 months	4643	–	–
Crawford et al., 2023 [8]	USA	RCS	III	4 months	158	65	60.8%
Gabriel et al., 2019 [10]	USA	RCS	III	2 years	240	–	50.50%
Gielis et al., 2019 [11]	Netherlands	RCS	III	3 years	1044	55.9 (±5.1)	87.30%
Hirvasniemi et al., 2019 [14]	Netherlands	PCS	II	10 years	197	55.7 (±5.2)	83.80%
Hyer et al., 2020 [15]	USA	RCS	III	2 years	19,522	–	–
Jang et al., 2022 [16]	USA	RCS	III	10 years	4796	31	58%
Jang et al., 2022 [17]	USA	DS	IV	11 years	793	61.4	57.60%
Jodeiri et al., 2020 [18]	Japan	PCS	II	–	95	–	–
Jones et al., 2019 [19]	USA	RCS	III	1 year	43,407	–	–
Kang et al., 2020 [20]	South Korea	RCS	III	–	1202	–	–
Karhade et al., 2019 [21]	USA	RCS	III	18 years	1101	66 (IQR 57–74)	49.50%
Klemm et al., 2022 [22]	USA	RCS	III	3 years	183	62 (IQR 54–71)	57.30%
Kunze et al., 2020 [23]	USA	RCS	III	3 years	183	62 (IQR 54–71)	57.30%
Nemes et al., 2016 [43]	Sweden	RCS	III	1 years	646	68.9	58.40%
Ramkumar et al., 2019 [47]	USA	RCS	III	4 years	78,335	75.3 (65–106)	63.60%
Ramkumar et al., 2019 [48]	USA	RCS	III	4 years	30,584	–	–
Sherafati et al., 2020 [49]	USA	RCS	II	4 years	78	63.1 (±8.5)	47%
Sniderman et al., 2020 [51]	Canada	PS	I	3 months	53	66.7 (±9.7) (34–84)	44%
Van de Meulebroucke et al., 2019 [52]	Belgium	RCS	III	1 year	100	–	–
Zhang et al., 2021 [56]	Singapore	PCS	II	12 years	302	62.9 (±12.1)	69.90%

Abbreviations: DS, diagnostic study; PCS, prospective cohort study; PS, prognostic study; RCS, retrospective cohort study.

## Cost prediction

Three of the five studies reporting this outcome variable used ANN, one of which reported an AUC of 0.82; however, SBG and Bayesian were also used, which reported an AUC of 0.87 [1, 15, 47, 48, 53].

## Length of stay (LOS)

Three studies analyzed this outcome, using ANN, Bayesian and MLR, reporting an AUC of 0.82 and 0.87, respectively [10, 47, 48].

## Readmission

Only one study analyzed this variable together with the revision outcome, using logistic regression. The reported AUC value was 0.66 [18].

## Revision

Two studies reporting this outcome variable used ANN, SGB, random forest, ENPLR and logistic regression. The more powerful model was ANN, with an AUC of 0.85 [18, 21].

## Other complications

One study reported this outcome variable, using ANN, with an accuracy of 0.95 [53].

## Functional outcomes

Two studies reporting this outcome variable used SGB, random forest, SVM, neural network and ENPLR, with an average AUC of 0.82. SGB and random forest were the more powerful methods, with

**TABLE 2** Input and output variables (OV).

Authors and Year	Input variables	OV	OV2
Alam et al., 2019 [1]	Primary prostheses cost, revision procedures cost, HRQoL, state-to-state transition probabilities	Cost prediction	–
Borjali et al., 2020 [5]	X-ray	Preoperative planning	–
Chen et al., 2023 [7]	X-ray	Total Hip Arthroplasty prediction	–
Crawford et al., 2023 [8]	Age, sex, BMI, race, CCI, diabetes, COPD, CKD, depression, opioid use, benzodiazepine use, smoking status, degree of arthritis, physical therapy	Total Hip Arthroplasty prediction	–
Gabriel et al., 2019 [10]	Age, preoperative opioid use, mets, sex, preoperative anaemia, COPD, hypertension, obesity, primary anaesthesia type	LOS	–
Gielis et al., 2019 [11]	Demographics, clinical examination, radiographic parameters, incident HOA risk	Clinical outcomes (rhoa)	–
Hirvasniemi et al., 2019 [14]	Age, gender, BMI, baseline KL grade, bone texture	Preoperative planning	Clinical Outcomes (rhoa, JSN, OST)
Hyer et al., 2020 [15]	Age, sex, comorbidity, cost, LOS	Cost prediction	–
Jang et al., 2023 [16]	Demographics, clinical examination, radiographic parameters, socioeconomic factors	Predictive model for 10-year THA risk	Clinical outcomes
Jang et al., 2022 [17]	X-ray	Preoperative planning	Clinical outcomes (HJC)
Jodeiri et al., 2020 [18]	X-ray	Preoperative planning	–
Jones et al., 2019 [19]	Comorbidity, LOS	Readmission	Revision
Kang et al., 2020 [20]	X-ray	Preoperative planning	–
Karhade et al., 2019 [21]	Age, opioid exposure duration, preoperative haemoglobin, preoperative medications	Preoperative planning	Functional Outcomes (Prolonged post-operative opioid prescription)
Klemm et al., 2022 [22]	ASA score, BMI, revision surgery indication, bearing surface, female gender, implant fixation	Revision	–
Kunze et al., 2020 [23]	Demographics, medical history, PRHS, modified Harris Hip Score	Functional outcomes (CSO)	–
Nemes et al., 2016 [43]	Preoperative EQ-5D score, EQ visual analogue scale, visual analogue scale pain, Charnley classification, age, gender, BMI, surgical approach prosthesis type	Clinical outcomes (HRQoL)	–
Ramkumar et al., 2019 [47]	Age, sex, comorbidity, readmission	LOS	Cost prediction
Ramkumar et al., 2019 [48]	Age, sex, comorbidity, readmission	LOS	Cost prediction
Sherafati et al., 2020 [49]	Biomarkers (Necrosis score, Co and Cr metal-ion serum levels, synovial volume and thickness, mscore (metallosis), off-resonance frequency density)	Preoperative planning	–
Sniderman et al., 2020 [51]	Patient demographics, PRHS, cognitive appraisal processes, surgical approach	Functional outcome (HOOS)	–
Van de Meulebroucke et al., 2019 [52]	THA implant characteristics, Intraoperative complications, post-operative complications, interventions	Cost prediction	Preoperative planning
Zhang et al., 2021 [56]	Patient demographics, comorbidities, preoperative prompts	Clinical outcome (patient satisfaction)	–

Authors and Year	Country	Study design	Level of evidence	Follow-up	Sample size	Mean patient age	% Female patients
Alam et al., 2019 [1]	UK	RCS	III	–	10	–	–
Borjali et al., 2020 [5]	USA	RCS	III	1 year	25	61.3	47%
Chen et al., 2023 [7]	Taiwan	DS	IV	3 months	4643	–	–
Crawford et al., 2023 [8]	USA	RCS	III	4 months	158	65	60.8%

TABLE 2 (Continued)

Authors and Year	Country	Study design	Level of evidence	Follow-up	Sample size	Mean patient age	% Female patients
Gabriel et al., 2019 [10]	USA	RCS	III	2 years	240	–	50.50%
Gielis et al., 2019 [11]	Netherlands	RCS	III	3 years	1044	55.9 (±5.1)	87.30%
Hirvasniemi et al., 2019 [14]	Netherlands	PCS	II	10 years	197	55.7 (±5.2)	83.80%
Hyer et al., 2020 [15]	USA	RCS	III	2 years	19,522	–	–
Jang et al., 2023 [16]	USA	RCS	III	10 years	4796	31	58%
Jang et al., 2022 [17]	USA	DS	IV	11 years	793	61.4	57.60%
Jodeiri et al., 2020 [18]	Japan	PCS	II	–	95	–	–
Jones et al., 2019 [19]	USA	RCS	III	1 year	43,407	–	–
Kang et al., 2020 [20]	South Korea	RCS	III	–	1202	–	–
Karhade et al., 2019 [21]	USA	RCS	III	18 years	1101	66 (IQR 57–74)	49.50%
Klemm et al., 2022 [22]	USA	RCS	III	3 years	183	62 (IQR 54–71)	57.30%
Kunze et al., 2020 [23]	USA	RCS	III	3 years	183	62 (IQR 54–71)	57.30%
Nemes et al., 2016 [43]	Sweden	RCS	III	1 year	646	68.9	58.40%
Ramkumar et al., 2019 [47]	USA	RCS	III	4 years	78,335	75.3 (65–106)	63.60%
Ramkumar et al., 2019 [48]	USA	RCS	III	4 years	30,584	–	–
Sherafati et al., 2020 [49]	USA	RCS	II	4 years	78	63.1 (±8.5)	47%
Sniderman et al., 2020 [51]	Canada	PS	I	3 months	53	66.7 (±9.7) (34–84)	44%
Van de Meulebroucke et al., 2019 [52]	Belgium	RCS	III	1 years	100	–	–
Zhang et al., 2021 [56]	Singapore	PCS	II	12 years	302	62.9 (±12.1)	69.90%

Abbreviations: ASA, American Society of Anesthesiologists; BMI, body mass index; COPD, chronic obstructive pulmonary disease; CSO, clinically significant outcome; HJC, hip joint centre; HOA, hip osteoarthritis; HOOS, Hip disability and Osteoarthritis Outcome Score; HRQoL, health-related quality of life; IQR, interquartile range; JSN, joint space narrowing; KL, Kellgren and Lawrence; LOS, length of stay; MET, metabolic equivalent task; OST, osteophyte score; PRHS, patient-reported outcome measure; THA, total hip arthroplasty.

AUC averages of 0.86. The third study used LASSO, reporting an  $R^2$  value of 0.29 [20, 22, 52].

## Clinical outcome

Of the five studies reporting clinical outcomes, two used ENPLR with an average AUC of 0.79. One study used MARS and linear regression, with  $R^2$  values of 0.158 and 0.157, respectively, while the stacked model had a more powerful  $R^2$  value of 0.17. The fourth study used four methods; of these, LASSO and SVM with respective AUC values of 0.76 and 0.74 had the highest predictive values. The final study used CNN with an accuracy of 91% [11, 14, 16, 42, 57].

## MINORS RESULTS

A total of 23 studies were included in our review, all were assessed to check for quality and risk of bias according to the modified MINORS method. With the exception of one,

which did not clearly describe its validation method or data set distribution resulting in potentially biased outcomes, all 23 studies demonstrated a high level of quality and low risk of bias, with a modified MINORS score of at least 7/8 (88%), as reported in Figure 2.

## DISCUSSION

This systematic review analyzed the current state of AI/ML prediction models in THA, extracting the relevant data to demonstrate the models' potential in improving clinical decision-making, preoperative planning, related costs, complications, and ultimately patient outcomes. The primary aim of this review is to evaluate the effectiveness of AI/ML applications in THA compared to traditional methods in predicting outcomes, improving clinical decision-making, and optimizing healthcare delivery. In doing so, we also describe the current landscape of AI/ML applications in the field, identifying both strengths and limitations. A key finding was the outstanding discriminative ability of CNN prediction

TABLE 3 AI/ML characteristics.

Authors and Year	AI/ML methods	Data sources	AUC	Accuracy	Sensitivity	Specificity
Alam et al., 2019 [1]	ANN	Multicenter	–	–	–	–
Borjali et al., 2020 [5]	CNN	Single-centre study	–	1	–	–
Chen et al., 2023 [7]	CNN	Single-centre study	0.994	0.977	0.920	0.992
Crawford et al., 2023 [8]	SGB	Multicenter study	0.83	–	–	–
Gabriel et al., 2019 [10]	MLR	Single-centre study	0.735	–	64.5	72.1
Giellis et al., 2019 [11]	ENPLR	CHECK cohort	0.86	–	–	–
Hirvasniemi et al., 2019 [14]	ENPLR	CHECK cohort	0.71	–	–	–
Hyer et al., 2020 [15]	decision tree	Medicare	–	83.40%	–	–
Jang et al., 2023 [16]	GA2M	Osteoarthritis Initiative	0.81	–	–	–
Jang et al., 2022 [17]	CNN	Osteoarthritis Initiative (Multicenter)	–	91%	–	–
Jodeiri et al., 2020 [18]	CNN	Single-centre study	–	–	–	–
Jones et al., 2019 [19]	Logistic regression	Medicare	0.66	–	–	–
Kang et al., 2020 [20]	CNN	Multicenter	0.99	–	–	–
Karhade et al., 2019 [21]	SGB	Multicentre	0.77	–	–	–
	random forest		0.75	–	–	–
	SVM		0.59	–	–	–
	ANN		0.77	–	–	–
	ENPLR		0.77	–	–	–
Klemt et al., 2022 [22]	ANN	Single-centre study	0.85	–	–	–
	SGB		0.81			
	random forest		0.8			
	ENPLR		0.81			
Kunze et al., 2020 [23]	SGB	Single-centre study	0.94	–	–	–
	random forest		0.97	–	–	–
	SVM		0.87	–	–	–
	ANN		0.92	–	–	–

TABLE 3 (Continued)

Authors and Year	AI/ML methods	Data sources	AUC	Accuracy	Sensitivity	Specificity
Nemes et al., 2016 [43]	ENPLR	Swedish Hip Arthroplasty Register	0.87	—	—	—
	MARS	—	—	—	—	—
	Linear regression	—	—	—	—	—
Ramkumar et al., 2019 [47]	ANN	National Inpatient Sample	0.82 (LOS)	0.75 (LOS)	—	—
Ramkumar et al., 2019 [48]	Bayesian	Patient database	0.87 (LOS)	—	—	—
Sherafati et al., 2020 [49]	ANN	Single-centre study	0.84	—	—	—
Sniderman et al., 2020 [51]	LASSO	Single-centre study	—	—	—	—
Van de Meulebroucke et al., 2019 [52]	ANN	Single-centre study	—	0.95	—	—
Zhang et al., 2021 [56]	Random forest	Single-centre study	0.68	—	53.00%	76.50%
	XGBoost	—	0.66	—	50.50%	76.50%
	SVM	—	0.74	—	57.90%	76.50%
	logistic LASSO	—	0.76	—	65.30%	82.40%
Authors and Year	Input variables	Output variables (OV)	OV2			
Alam et al., 2019 [1]	Primary prostheses cost, revision procedures cost, HROQL, state-to-state transition probabilities	Cost prediction	—			
Borjaji et al., 2020 [5]	X-ray	Preoperative planning	—			
Chen et al., 2023 [7]	X-ray	Total Hip Arthroplasty prediction	—			
Crawford et al., 2023 [8]	Age, sex, BMI, race, CCI, diabetes, COPD, CKD, depression, opioid use, benzodiazepine use, smoking status, degree of arthritis, physical therapy	Total Hip Arthroplasty prediction	—			
Gabriel et al., 2019 [10]	Age, preoperative opioid use, mets, sex, preoperative anaemia, COPD, hypertension, obesity, primary anaesthesia type	LOS	—			
Gielis et al., 2019 [11]	Demographics, clinical examination, radiographic parameters, incident HOA risk	Clinical outcomes (rhoa)	—			
Hirvasniemi et al., 2019 [14]	Age, gender, BMI, baseline KL grade, bone texture	Preoperative planning	Clinical outcomes (rhoa, JSN, OST)			
Hyer et al., 2020 [15]	Age, sex, comorbidity, cost, LOS	Cost prediction	—			
Jang et al., 2023 [16]	Demographics, clinical examination, radiographic parameters, socioeconomic factors	Predictive model for 10-year THA risk	Clinical outcomes			
Jang et al., 2022 [17]	X-ray	Preoperative planning	Clinical outcomes (HJC)			

(Continues)

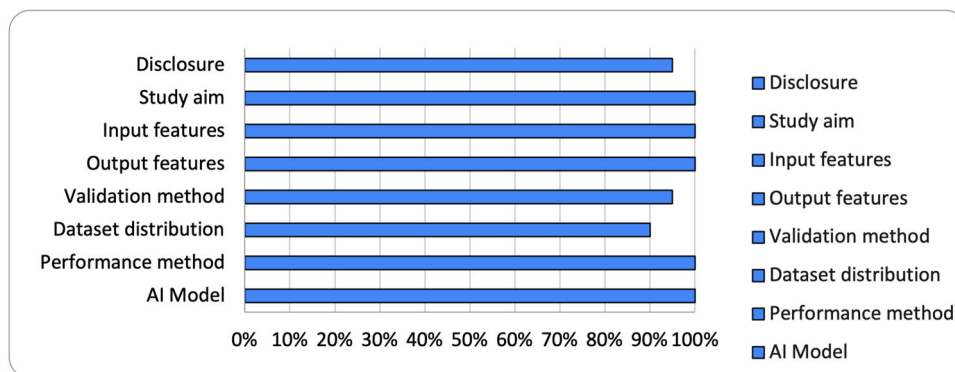
TABLE 3 (Continued)

Authors and Year	Input variables	Output variables (OV)	OV2				
Jodeiri et al., 2020 [18]	X-ray	Preoperative planning	–				
Jones et al., 2019 [19]	Comorbidity, LOS	Readmission	Revision				
Kang et al., 2020 [20]	X-ray	Preoperative planning	–				
Karhade et al., 2019 [21]	Age, opioid exposure duration, preoperative haemoglobin, preoperative medications	Preoperative planning	Functional Outcomes (Prolonged post-operative opioid prescription)				
Klemt et al., 2022 [22]	ASA score, BMI, revision surgery indication, bearing surface, female gender, implant fixation	Revision	–				
Kunze et al., 2020 [23]	Demographics, medical history, PRHS, modified Harris Hip Score	Functional outcomes (CSO)	–				
Nemes et al., 2016 [43]	Preoperative EQ-5D score, EQ visual analogue scale, visual analogue scale pain, Charnley classification, age, gender, BMI, surgical approach prosthesis type	Clinical outcomes (HRQoL)	–				
Ramkumar et al., 2019 [47]	Age, sex, comorbidity, readmission	LOS	Cost prediction				
Ramkumar et al., 2019 [48]	Age, sex, comorbidity, readmission	LOS	Cost prediction				
Sherafati et al., 2020 [49]	Biomarkers (Necrosis score, Co and Cr metal-ion serum levels, synovial volume and thickness, mscore (metallosis), off-resonance frequency density)	Preoperative planning	–				
Sniderman et al., 2020 [51]	Patient demographics, PRHS, cognitive appraisal processes, surgical approach	Functional outcome (HOOS)	–				
Van de Meulebroucke et al., 2019 [52]	THA implant characteristics, intraoperative complications, post-operative complications, interventions	Cost prediction	Preoperative planning				
Zhang et al., 2021 [56]	Patient demographics, comorbidities, preoperative prompts	Clinical outcome (patient satisfaction)	–				
Authors and Year	Country	Study design	Level of evidence	Follow-up	Sample size	Mean patient age	% Female patients
Alam et al., 2019 [1]	UK	RCS	III	–	10	–	–
Borjali et al., 2020 [5]	USA	RCS	III	1 year	25	61.3	47%
Chen et al., 2023 [7]	Taiwan	DS	IV	3 months	4643	–	–
Crawford et al., 2023 [8]	USA	RCS	III	4 months	158	65	60.8%
Gabriel et al., 2019 [10]	USA	RCS	III	2 years	240	–	50.50%
Gielis et al., 2019 [11]	Netherlands	RCS	III	3 years	1044	55.9 (±5.1)	87.30%

TABLE 3 (Continued)

Authors and Year	Country	Study design	Level of evidence	Follow-up	Sample size	Mean patient age	% Female patients
Hirvasniemi et al., 2019 [14]	Netherlands	PCS	II	10 years	197	55.7 (±5.2)	83.80%
Hyer et al., 2020 [15]	USA	RCS	III	2 years	19,522	–	–
Jang et al., 2023 [16]	USA	RCS	III	10 years	4796	31	58%
Jang et al., 2022 [17]	USA	DS	IV	11 years	793	61.4	57.60%
Jodeiri et al., 2020 [18]	Japan	PCS	II	–	95	–	–
Jones et al., 2019 [19]	USA	RCS	III	1 year	43,407	–	–
Kang et al., 2020 [20]	South Korea	RCS	III	–	1202	–	–
Karhade et al., 2019 [21]	USA	RCS	III	18 years	1101	66 (IQR 57–74)	49.50%
Klemt et al., 2022 [22]	USA	RCS	III	3 years	183	62 (IQR 54–71)	57.30%
Kunze et al., 2020 [23]	USA	RCS	III	3 years	183	62 (IQR 54–71)	57.30%
Nemes et al., 2016 [43]	Sweden	RCS	III	1 year	6,46	68.9	58.40%
Ramkumar et al., 2019 [47]	USA	RCS	III	4 years	78,335	75.3 (65–106)	63.60%
Ramkumar et al., 2019 [48]	USA	RCS	III	4 years	30,584	–	–
Sherafati et al., 2020 [49]	USA	RCS	II	4 years	78	63.1 (±8.5)	47%
Sniderman et al., 2020 [51]	Canada	PS	I	3 months	53	66.7 (±9.7) (34–84)	44%
Van de Meulebroucke et al., 2019 [52]	Belgium	RCS	III	1 years	100	–	–
Zhang et al., 2021 [56]	Singapore	PCS	II	12 years	302	62.9 (±12.1)	69.90%

Abbreviations: ANN, artificial neural network; AUC, area under the curve; BMI, body mass index; CNN, convolutional neural network; ENPLR, elastic net penalized logistic regression; GA2M, Generalized Additive Model with Pairwise Interactions; IQR, interquartile range; MARS, multivariate adaptive regression splines; MLR, multiple linear regression; SGB, stochastic gradient boosting; SVM, support vector machine; XGBoost, extreme gradient boosting.



**FIGURE 2** Modified methodological index for nonrandomized studies (MINORS).

models in identifying failed implant design in preoperative planning of revision THA. Previous research has shown that 10% of failed implant designs could not be identified preoperatively, as well as 2% intraoperatively, and projected that by 2030, cumulative surgeon time to identify failed implants would rise to almost 130,000 h (about 15 years) while opportunity cost of surgeon as well as staff time would double compare to 2020 [38, 55, 56]. These changes can lead to increased operating room procedure time, costs, blood loss and recovery time [55, 56]. As the CNN model's singular input variable was anteroposterior radiographs, this model can potentially be a low cost, timesaving, accurate detection tool ubiquitous in revision THA clinical settings.

From a larger perspective, AI/ML models performed outstandingly in predicting CSO post-THA and performed excellently in predicting LOS, rHOA and periprosthetic tissue necrosis progression. In predicting CSO post THA and LOS, AI/ML models leveraged modifiable risk factors to directly identify patients at risk for sub-optimal outcomes as well as indirectly provide healthcare providers with actionable risk factors to optimize outcomes at the preoperative stage and facilitate shared decision-making. Preoperative opioid use and body mass index are two modifiable risk factors that have been found to predispose patients to lower patient-reported outcomes post-THA [4, 22]. By using random forest, a supervised ML algorithm, THA patients at greater risk for poor CSO in post-operative PROM can be identified, allowing surgeons to intervene preoperatively and potentially minimize risks [22, 29]. Due to the increasing THA utilization, finding ways to prevent prolonged LOS is important to contain hospital costs and mitigate readmission risk [40, 54]. Pain and physical therapy are two modifiable causes of prolonged LOS; by using ANN and Bayesian prediction models to identify patients at higher risk for prolonged LOS, clinicians can potentially more effectively allocate hospital resources, and decrease hospital costs, by targeting these two factors in higher-risk patients [10, 39].

## Preoperative applications

AI/ML models prove to be a useful predictive tool; two such models are CNN and ENPLR, which have exhibited great potential in optimizing clinical and surgical decision-making at the preoperative stage. These two models can serve as cost-effective solutions to escalating healthcare expenses, while simultaneously paving the way for more personalized care and potentially reduced OR (operating room) time [56]. CNN-based implant recognition programs have proven to be a novel automatic approach to the collection and identification of failed implant designs. Drawing on CNN predictive power can make manual identification of implant devices less time-consuming and less error-prone, thus reducing OR time and overall healthcare costs while increasing revision surgery success [5, 19]. CNN models can be implemented to analyze preoperative data to calculate patients' risk for post-operative complications based on the patients' medical history and needs, further providing a virtual guide to optimize implant placement to achieve better alignment and stability [16, 17]. ENPLR-based algorithms may be useful in predicting the personalized risk of incident rHOA and THA based on radiographic findings [11, 14, 28]. Overall, the precision and effectiveness of preoperative care in THA can be significantly improved by AI and ML models, leading to better patient outcomes and reduced healthcare expenses. Chen and Crawford explored the possibility of predicting the risk of THA in a cohort of patients using ML. AI could help researchers to early identify patients who can benefit from THA replacement, minimizing the risk of complex surgeries due to excessive latency of the treatment.

## Post-operative applications

AI/ML models may prove advantageous at the post-operative stage as well, by utilizing their predictive capabilities clinicians may better prevent or manage

post-operative complications, thus limiting adverse events, as well as forecast surgical outcomes and readmission risk. Misuse and overdosing of prescription opioids are a significant burden on public health systems globally in regard to both mortality and costs; therefore, efficacious preventive measures are needed [2, 27, 43]. ENPLR algorithms have demonstrated the ability to identify patients who are at higher risk for prolonged post-operative opioid prescriptions. The early identification and intervention in high-risk cases could help prevent or alleviate the unfavourable outcomes of opioid dependence [20]. In cases of high projected risk of potentially severe post-operative complications and outcomes, these tools may provide a useful guide to better customize therapy for patients. MARS models may allow for predicted HRQoL, which can be incorporated into the decision-making between providers and patients, enabling a review of optimal treatment options and expected benefits following THA [42]. Finally, ANN and Bayesian models can be useful in identifying patients who are more likely to need prolonged LOS, essentially providing a functional tool to achieve strategic planning for hospital bed availability [10]. In short, AI/ML tools have the potential to improve post-operative care in THA by reducing costs while increasing targeted and personalized therapy based on patients' unique characteristics, thus resulting in enhanced patient satisfaction and overall healthcare [18, 30, 42, 48, 50, 52, 53].

ML methods exhibit a strong potential in optimizing patient outcomes by contributing to each stage of THA, but many models still require improvements before they can become standardized and widely adopted in external institutions, as well as different geographic regions [13, 23]. Among the studies reviewed, the most common and salient constraints in model performance and generalizability were limited data set size, limited data granularity and lack of external validation. Approximately 50% of the reviewed articles used single-centre data sources; as AI models are shaped by the data inputted, the limited representation does not account for regional differences or more varied patient characteristics and thus constrains the model's usability. Access to larger data sets across distinct populations and regions is paramount to preventing overfitting or drawing conclusions that are only applicable to identically distributed data sets [9, 41]. Data granularity relates to the extent of detail in the data; by increasing it, the output variables can become more precise and thus more effective in clinical predictions and decision-making. Larger studies can potentially curtail the limited data set issue while making available

potentially important predictors, such as relevant clinical, socioeconomic, and administrative factors [10, 17, 18, 20, 47]. Finally, collaboration among researchers and their institutions will be required to optimize data sharing and model reliability [13, 16, 20, 22, 47, 48, 50, 57].

## LIMITATIONS

This review does have some potential limitations. Additionally, most developed models did not undergo the relevant testing and validation that is needed to evaluate their clinical efficacy; therefore, conclusions are limited to the potential of these models. It should also be mentioned that one study reported a CNN model with an accuracy of 1, which may indicate, despite the prior highlighted strengths, some weaknesses such as overfitting, although it was reported that steps were taken to reduce this, or a small data set [15]. Lopez et al. reviewed the accuracy of AI/ML applications in THA as well as total knee arthroplasty; the present review is a more precise and exact study focusing solely on THA [37]. The reviewed ML models were formulated for highly specific clinical tasks in the realm of THA; thus, this review presented the exact AI/ML methods used as well as their unique input and output variables, which are the foundations of these models and their clinical applicability. Furthermore, a quality assessment was performed using a modified version of the MINORS checklist. This modified version was developed by Langerhuizen et al. as a response to the lack of a suitable tool for bias assessment in diagnostic studies [26, 51]. Risk of bias assessments must be incorporated in future reviews to help solidify the legitimacy and integrity of the reviewed models. This systematic review did not perform a meta-analysis and, therefore, did not report statistical significance. This is due to a high level of heterogeneity among the articles included. The studies included a broad pool of different populations, interventions, algorithms and outcomes, rendering a meta-analysis inapplicable.

Different levels of evidence were included in the paper, providing high heterogeneity between articles in terms of scientific validity. Finally, one major limitation of current AI/ML applications in THA is the lack of external validation across diverse populations and healthcare settings. Furthermore, the reliance on short follow-up periods (<12 months in most studies) restricts the ability to assess long-term clinical and economic outcomes. Future research should prioritize standardized validation protocols and multi-year follow-ups to enhance the robustness and clinical utility of these models.

## CONCLUSIONS

AI/ML has the potential to aid surgical care at all stages and broadly optimize healthcare, from hospital administration to individual patient care. Models demonstrated impressive predictive abilities, but standardized implementation remains to be seen. In order to move beyond the theoretical potential of these models, clinical efficacy and standardization must be proven and dealt with. It is essential for future developments to perform external validation as well as assessments such as calibration and decision curve analysis to verify clinical utility. Researchers must ensure that standardized, comprehensive, and high-quality data is used, as ML models are only as good, and unbiased, as the data fed into them. Mainstream prediction models have the luxury of dealing with lesser implications; as the medical field deals with human lives, the accuracy and integrity strived for in surgical interventions should be matched in the orthopaedics ML developments.

## AUTHOR CONTRIBUTIONS

*Manuscript preparation, study design, database interpretation and manuscript revision:* Umile Giuseppe Longo, Sergio De Salvatore and Alice Piccolomini. *Manuscript preparation, database interpretation and statistical analysis:* Nathan Samuel Ullman, Alice Piccolomini and Rocco Papalia. *Manuscript preparation, figures and tables preparation, study design:* Alice Piccolomini, Sergio De Salvatore and Giuseppe Salvatore. *Manuscript preparation and database interpretation:* Umile Giuseppe Longo, Nathan Samuel Ullman, Sergio De Salvatore and Maristella Saccomanno. *Study design, manuscript revision:* Umile Giuseppe Longo, Margaux D'Hooghe and Kristian Samuelsson. The authors read and approved the final manuscript.

## CONFLICT OF INTEREST STATEMENT

Kristian Samuelsson is a member of the Board of Directors of Getinge AB (publ). The other authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

The data sets used and/or analyzed during the current study available from the corresponding author on reasonable request.

## ETHICS STATEMENT

The ethics statement is not available.

## ORCID

Kristian Samuelsson  <http://orcid.org/0000-0001-5383-3370>

## REFERENCES

1. Alam MF, Briggs A. Artificial neural network metamodel for sensitivity analysis in a total hip replacement health economic model. *Expert Rev Pharmacoecon Outcomes Res.* 2020;20:629–40.
2. Alho H, Dematteis M, Lembo D, Maremmani I, Roncero C, Somaini L. Opioid-related deaths in Europe: strategies for a comprehensive approach to address a major public health concern. *Int J Drug Policy.* 2020;76:102616.
3. Barragán-Montero A, Javaid U, Valdés G, Nguyen D, Desbordes P, Macq B, et al. Artificial intelligence and machine learning for medical imaging: a technology review. *Phys Med.* 2021;83:242–56.
4. Bonner BE, Castillo TN, Fitz DW, Zhao JZ, Klemm C, Kwon YM. Preoperative opioid use negatively affects patient-reported outcomes after primary total hip arthroplasty. *J Am Acad Orthop Surg.* 2019;27:e1016–20.
5. Borjali A, Chen AF, Muratoglu OK, Morid MA, Varadarajan KM. Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural network. *J Orthop Res.* 2020;38:1465–71.
6. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. *Front Bioeng Biotechnol.* 2018;6:75.
7. Chen CC, Wu CT, Chen CPC, Chung CY, Chen SC, Lee MS, et al. Predicting the risk of total hip replacement by using a deep learning algorithm on plain pelvic radiographs: diagnostic study. *JMIR Form Res.* 2023;7:e42788.
8. Crawford AM, Karhade AV, Agaronnik ND, Lightsey HM, Xiong GX, Schwab JH, et al. Development of a machine learning algorithm to identify surgical candidates for hip and knee arthroplasty without in-person evaluation. *Arch Orthop Trauma Surg.* 2023;143:5985–92.
9. Eche T, Schwartz LH, Mokrane FZ, Dercle L. Toward generalizability in the deployment of artificial intelligence in radiology: role of computation stress testing to overcome under-specification. *Radiol Artif Intell.* 2021;3:e210097.
10. Gabriel RA, Sharma BS, Doan CN, Jiang X, Schmidt UH, Vaida F. A predictive model for determining patients not requiring prolonged hospital length of stay after elective primary total hip arthroplasty. *Anesth Analg.* 2019;129:43–50.
11. Gielis WP, Weinans H, Welsing PMJ, van Spil WE, Agricola R, Cootes TF, et al. An automated workflow based on hip shape improves personalized risk prediction for hip osteoarthritis in the CHECK study. *Osteoarthr Cartil.* 2020;28:62–70.
12. Grazal CF, Anderson AB, Booth GJ, Geiger PG, Forsberg JA, Balazs GC. A machine-learning algorithm to predict the likelihood of prolonged opioid use following arthroscopic hip surgery. *Arthroscopy.* 2022;38:839–47.e2.
13. Groot OQ, Bindels BJJ, Ogink PT, Kapoor ND, Twining PK, Collins AK, et al. Availability and reporting quality of external validations of machine-learning prediction models with orthopedic surgical outcomes: a systematic review. *Acta Orthop.* 2021;92:385–93.
14. Hirvasniemi J, Gielis WP, Arbabi S, Agricola R, van Spil WE, Arbabi V, et al. Bone texture analysis for prediction of incident radiographic hip osteoarthritis using machine learning: data from the Cohort Hip and Cohort Knee (CHECK) study. *Osteoarthr Cartil.* 2019;27:906–14.
15. Hyer JM, Paredes AZ, White S, Ejaz A, Pawlik TM. Assessment of utilization efficiency using machine learning techniques: a study of heterogeneity in preoperative healthcare utilization among super-utilizers. *Am J Surg.* 2020;220:714–20.
16. Jang SJ, Kunze KN, Vigdorich JM, Jerabek SA, Mayman DJ, Sculco PK. John Charnley Award: deep learning prediction of hip joint center on standard pelvis radiographs. *J Arthroplasty.* 2022;37:S400–7.e1.
17. Jodeiri A, Zoroofi RA, Hiasa Y, Takao M, Sugano N, Sato Y, et al. Fully automatic estimation of pelvic sagittal inclination

- from anterior-posterior radiography image using deep learning framework. *Comput Methods Programs Biomed.* 2020;184:105282.
18. Jones CD, Falvey J, Hess E, Levy CR, Nuccio E, Barón AE, et al. Predicting hospital readmissions from home healthcare in medicare beneficiaries. *J Am Geriatr Soc.* 2019;67:2505–10.
  19. Kang YJ, Yoo JI, Cha YH, Park CH, Kim JT. Machine learning-based identification of hip arthroplasty designs. *J Orthop Transl.* 2020;21:13–7.
  20. Karhade AV, Schwab JH, Bedair HS. Development of machine learning algorithms for prediction of sustained postoperative opioid prescriptions after total hip arthroplasty. *J Arthroplasty.* 2019;34:2272–7.e1.
  21. Klemm C, Cohen-Levy WB, Robinson MG, Burns JC, Alpaugh K, Yeo I, et al. Can machine learning models predict failure of revision total hip arthroplasty? *Arch Orthop Trauma Surg.* 2023;143:2805–12.
  22. Kunze KN, Karhade AV, Sadauskas AJ, Schwab JH, Levine BR. Development of machine learning algorithms to predict clinically meaningful improvement for the patient-reported health state after total hip arthroplasty. *J Arthroplasty.* 2020;35:2119–23.
  23. Kunze KN, Orr M, Krebs V, Bhandari M, Piuze NS. Potential benefits, unintended consequences, and future roles of artificial intelligence in orthopaedic surgery research: a call to emphasize data quality and indications. *Bone Jt Open.* 2022;3:93–7.
  24. Kurmis AP, Iannunzio JR. Artificial intelligence in orthopedic surgery: evolution, current state and future directions. *Arthroplasty.* 2022;4:9.
  25. Lalehzarian SP, Gowd AK, Liu JN. Machine learning in orthopaedic surgery. *World J Orthop.* 2021;12:685–99.
  26. Langerhuizen DWG, Janssen SJ, Mallee WH, van den Bekerom MPJ, Ring D, Kerkhoffs GMMJ, et al. What are the applications and limitations of artificial intelligence for fracture detection and classification in orthopaedic trauma imaging? A systematic review. *Clin Orthop Relat Res.* 2019;477:2482–91.
  27. Leider HL, Dhaliwal J, Davis EJ, Kulakodlu M, Buikema AR. Healthcare costs and nonadherence among chronic opioid users. *Am J Manag Care.* 2011;17:32–40.
  28. Longo UG, Candela V, Berton A, De Salvatore S, Fioravanti S, Giannone L, et al. Biosensors for detection of biochemical markers relevant to osteoarthritis. *Biosensors.* 2021;11:31.
  29. Longo UG, Ciuffreda M, Candela V, Berton A, Maffulli N, Denaro V. Hip scores: a current concept review. *Br Med Bull.* 2019;131:81–96.
  30. Longo UG, De Salvatore S, Greco A, Marino M, Santamaria G, Piergentili I, et al. Influence of depression and sleep quality on postoperative outcomes after total hip arthroplasty: a prospective study. *J Clin Med.* 2022;11:3845.
  31. Longo UG, De Salvatore S, Piergentili I, Indiveri A, Di Naro C, Santamaria G, et al. Total hip arthroplasty: minimal clinically important difference and patient acceptable symptom state for the forgotten joint score 12. *Int J Environ Res Public Health.* 2021;18:2267.
  32. Longo UG, De Salvatore S, Santamaria G, Indiveri A, Piergentili I, Salvatore G, et al. Total hip replacement: psychometric validation of the Italian Version of Forgotten Joint Score (FJS-12). *J Clin Med.* 2023;12:1525.
  33. Longo UG, Marchetti A, Corradini A, Candela V, Salvatore G, De Marinis MG, et al. Patients' perceptions and experiences during the pre-admission phase for total hip replacement surgery: a qualitative phenomenological study. *J Clin Med.* 2023;12:2754.
  34. Longo UG, Papalia R, De Salvatore S, Ruzzini L, Piergentili I, Salvatore G, et al. Trends in hospitalization for pediatric hip arthroplasty: an epidemiological nationwide study in Italy from 2001 to 2015. *BMC Pediatr.* 2022;22:235.
  35. Longo UG, Papalia R, Mazzola A, Campi S, De Salvatore S, Candela V, et al. Bilateral simultaneous hip and knee replacement: an epidemiological nationwide study from 2001 to 2016. *BMC Surg.* 2024;24:172.
  36. Longo UG, Papalia R, Salvatore G, Tecce SM, Jedrzejczak A, Marozzi M, et al. Epidemiology of revision hip replacement in Italy: a 15-year study. *BMC Surg.* 2022;22:355.
  37. Lopez CD, Gazgalis A, Boddapati V, Shah RP, Cooper HJ, Geller JA. Artificial Learning and machine learning decision guidance applications in total hip and knee arthroplasty: a systematic review. *Arthroplast Today.* 2021;11:103–12.
  38. Maffulli N, Margiotti K, Longo UG, Loppini M, Fazio VM, Denaro V. The genetics of sports injuries and athletic performance. *Muscles Ligaments Tendons J.* 2013;3:173–89.
  39. Meyers SJ, Reuben JD, Cox DD, Watson M. Inpatient cost of primary total joint arthroplasty. *J Arthroplasty.* 1996;11:281–5.
  40. Molloy IB, Martin BI, Moschetti WE, Jevsevar DS. Effects of the length of stay on the cost of total knee and total hip arthroplasty from 2002 to 2013. *J Bone Jt Surg.* 2017;99:402–7.
  41. Mutasa S, Sun S, Ha R. Understanding artificial intelligence based radiology studies: what is overfitting? *Clin Imaging.* 2020;65:96–9.
  42. Nemes S, Rolfson O, Garellick G. Development and validation of a shared decision-making instrument for health-related quality of life one year after total hip replacement based on quality registries data. *J Eval Clin Pract.* 2018;24:13–21.
  43. Neuman MD, Bateman BT, Wunsch H. Inappropriate opioid prescription after surgery. *Lancet.* 2019;393:1547–57.
  44. Oettl FC, Pareek A, Winkler PW, Zsidai B, Pruneski JA, Senorski EH, et al. A practical guide to the implementation of AI in orthopaedic research, Part 6: how to evaluate the performance of AI research? *J Exp Orthop.* 2024;11:e12039.
  45. Pabinger C, Lothaller H, Portner N, Geissler A. Projections of hip arthroplasty in OECD countries up to 2050. *HIP International.* 2018;28:498–506.
  46. Purnomo G, Yeo SJ, Liow MHL. Artificial intelligence in arthroplasty. *Arthroplasty.* 2021;3:37.
  47. Ramkumar PN, Karnuta JM, Navarro SM, Haerberle 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;34:2228–34.e1.
  48. Ramkumar PN, Navarro SM, Haerberle 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.
  49. Rupp MC, Moser LB, Hess S, Angele P, Aurich M, Dyrna F, et al. Orthopaedic surgeons display a positive outlook towards artificial intelligence: a survey among members of the AGA Society for Arthroscopy and Joint Surgery. *J Exp Orthop.* 2024;11:e12080.
  50. Sherafati M, Bauer TW, Potter HG, Koff MF, Koch KM. Multi-variate use of MRI biomarkers to classify histologically confirmed necrosis in symptomatic total hip arthroplasty. *J Orthop Res.* 2020;38:1506–14.
  51. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. *ANZ J Surg.* 2003;73:712–6.
  52. Sniderman J, Stark RB, Schwartz CE, Imam H, Finkelstein JA, Nousiainen MT. Patient factors that matter in predicting hip arthroplasty outcomes: a machine-learning approach. *J Arthroplasty.* 2021;36:2024–32.
  53. Van de Meulebroucke C, Beckers J, Corten K. What can we expect following anterior total hip arthroplasty on a regular operating table? A validation study of an artificial intelligence algorithm to monitor adverse events in a high-volume, non-academic setting. *J Arthroplasty.* 2019;34:2260–6.

54. Williams J, Kester BS, Bosco JA, Slover JD, Iorio R, Schwarzkopf R. The association between hospital length of stay and 90-day readmission risk within a total joint arthroplasty bundled payment initiative. *J Arthroplasty*. 2017;32:714–8.
55. Wilson N, Broatch J, Jehn M, Davis 3rd, C. National projections of time, cost and failure in implantable device identification: consideration of unique device identification use. *Healthcare*. 2015;3:196–201.
56. Wilson NA, Jehn M, York S, Davis 3rd, CM. Revision total hip and knee arthroplasty implant identification: implications for use of Unique Device Identification 2012 AAHKS member survey results. *J Arthroplasty*. 2014;29:251–5.
57. Zhang S, Chen JY, Pang HN, Lo NN, Yeo SJ, Liow MHL. Development and internal validation of machine learning

algorithms to predict patient satisfaction after total hip arthroplasty. *Arthroplasty*. 2021;3:33.

**How to cite this article:** Longo UG, De Salvatore S, Piccolomini A, Ullman NS, Salvatore G, D'Hooghe M, et al. Revolutionizing total hip arthroplasty: the role of artificial intelligence and machine learning. *J Exp Orthop*. 2025;12:e70195. <https://doi.org/10.1002/jeo2.70195>