BRITISH JOURNAL OF
HOSPITAL
MEDICINE

REVIEW

Using Artificial Intelligence to predict outcomes of operatively managed neck of femur fractures

Geeth Silva[1]

Robert Ashford[1]

Author details can be found at the end of this article

Correspondence to:
Geeth Silva; geethsilva95@gmail.com

Abstract

Aims/Background Patients with neck of femur fractures present a tremendous public health problem that leads to a high incidence of death and dysfunction. An essential factor is the postoperative length of stay, which heavily impacts hospital costs and the quality of care. As an extension of traditional statistical methods, machine learning (ML) provides the possibility of accurately predicting the length of hospital stay. This review assesses how machine learning can effectively use healthcare data to predict the outcomes of patients with operatively managed neck of femurs.

Methods A narrative literature review on the use of Artificial Intelligence to predict outcomes in the neck of femurs was undertaken to understand the field and critical considerations of its application. The papers and any relevant references were scrutinised using the specific inclusion and exclusion criteria to produce papers that were used in the analysis.

Results Thirteen papers were used in the analysis. The critical themes recognised the different models, the 'backbox' conundrum, predictor identification, validation methodology and the need to improve efficiency and quality of care. Through reviewing the themes in this paper, current issues, and potential avenues of advancing the field are explored.

Conclusions This review has demonstrated that the use of machine learning in Orthopaedic pathways is in its infancy. Further work is needed to leverage this technology effectively to improve outcomes.

Key words: Artificial Intelligence; Literature review; Neck of femur fractures; Orthopaedics

Submitted: 26 January 2024; Revised: 28 February 2024; Accepted: 16 March 2024

Introduction

Patients with neck of femur (NOF) fractures present a tremendous public health problem leading to high morbidity and mortality. Over 70,000 NOFs occur in a year within the U.K., resulting in over £2 billion in healthcare costs alongside 10% mortality at 30 days and up to 30% mortality at one year. Positive outcomes are only achieved through close multidisciplinary teamwork between many specialities. Furthermore, their trajectory, including inpatient stay, is determined by medical co-morbidities, their rehabilitative ability and the care they are given (British Orthopaedic Association, 2012). Outcomes that heavily impact hospital costs and act as a surrogate for the quality of care provided include mortality, postoperative length of stay and return to mobility.

Many predictive models are used to assess the outcomes of NOFs, the most used in the U.K. being the Nottingham Hip Fracture Score (NHFS). This score was retrospectively calculated using, 6202 patients between 1999 and 2009 who underwent hip fracture surgery. Researchers collected one-year and 30-day postoperative mortality data from hospital records and the Office of National Statistics. This data underwent statistical analysis to create a validated prediction tool for 30-day mortality after hip fracture surgery (Wiles et al, 2011). As powerful as these statistical analyses are, machine learning (ML) provides the possibility of more accurately predicting outcomes in NOFs (Bayliss and Jones, 2019).

Artificial Intelligence (AI) describes using technology to mimic intelligent behaviour and critical thinking like humans. Machine learning is a branch of AI that analyses data sets to predict associations without human input. A training data set is used to teach the

How to cite this article:
Silva G, Ashford R. Using Artificial Intelligence to predict outcomes of operatively managed neck of femur fractures. Br J Hosp Med. 2024. https://doi.org/10.12968/hmed.2024.0034

algorithm how to assess variables and predict future patterns (Bayliss and Jones, 2019). These assumptions develop as the dataset increases and the programme becomes more predictively accurate. The algorithm can finally be tested and evaluated for accuracy and validity. ML can be seen on a spectrum of supervised to unsupervised approaches. Supervised approaches, such as linear regression, identifies relationships between labelled input data to discrete output data. Unsupervised approaches, such as neural networks, organise the input data and interactions among certain variables to identify unspecified patterns that can predictively be applied to future unseen data sets (Bayliss and Jones, 2019). A review by the King's Fund noted the tremendous potential of such tools to support staff and patients against the backdrop of increasing pressures on the National Healthcare Services (NHS). Nonetheless significant gaps in the evidence, lack of validated evaluation methods and support during implementation have limited their application (Maguire et al, 2021).

Within Orthopaedics, there has been an explosion in data sources, including the National Joint Registry and the National Hip Fracture Database. These sources present an opportunity for ML to diagnose injury and predict Orthopaedic-specific patient outcomes (Bayliss and Jones, 2019). Identifying high-risk patients can potentially improve resource allocation, therefore improving efficiency and quality of care against the ever-increasing demand on the NHS (Inouye et al, 1999).

Methods

A variety of literature exists on the topic of the use of AI in NOF management. A Narrative Literature Review (NLR) is more appropriate than a systematic review as it allows for a broader scope (Green et al, 2006). However, a systematic approach is deemed suitable for the NLR to prevent selection bias resulting in misguided conclusions (Rumrill and Fitzgerald, 2001). This NLR involves using a search string and inclusion-exclusion criteria.

Medline and Embase databases are used for the analysis. These databases ensure that the papers gathered are high quality, validated within the scientific community, and appropriate for analysis. The search string included phrases such as: 'femur neck, neck of femur, NOF, femoral neck fractures, proximal femur' to encompass NOFs alongside 'Artificial Intelligence, neural networks, computer neural networks, machine learning and Deep Learning to encompass Artificial Intelligence' (Appendixes 1 and 2). Inclusion and exclusion criteria were used to filter through the most relevant papers generated from the search string (Figure 1) (Rumrill et al, 2001).

The papers were reviewed, and any relevant references were scrutinised using the same inclusion and exclusion criteria and included in the review (Figure 2).

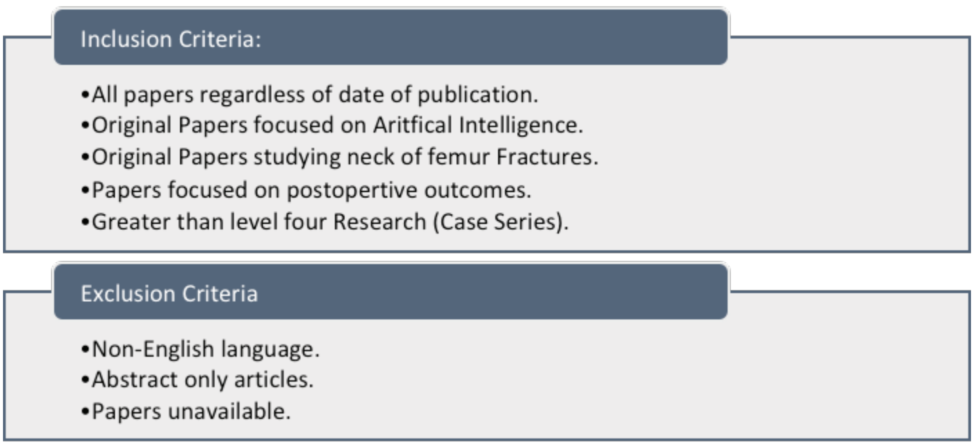

Figure 1. Inclusion and exclusion criteria (Rumrill et al, 2001).

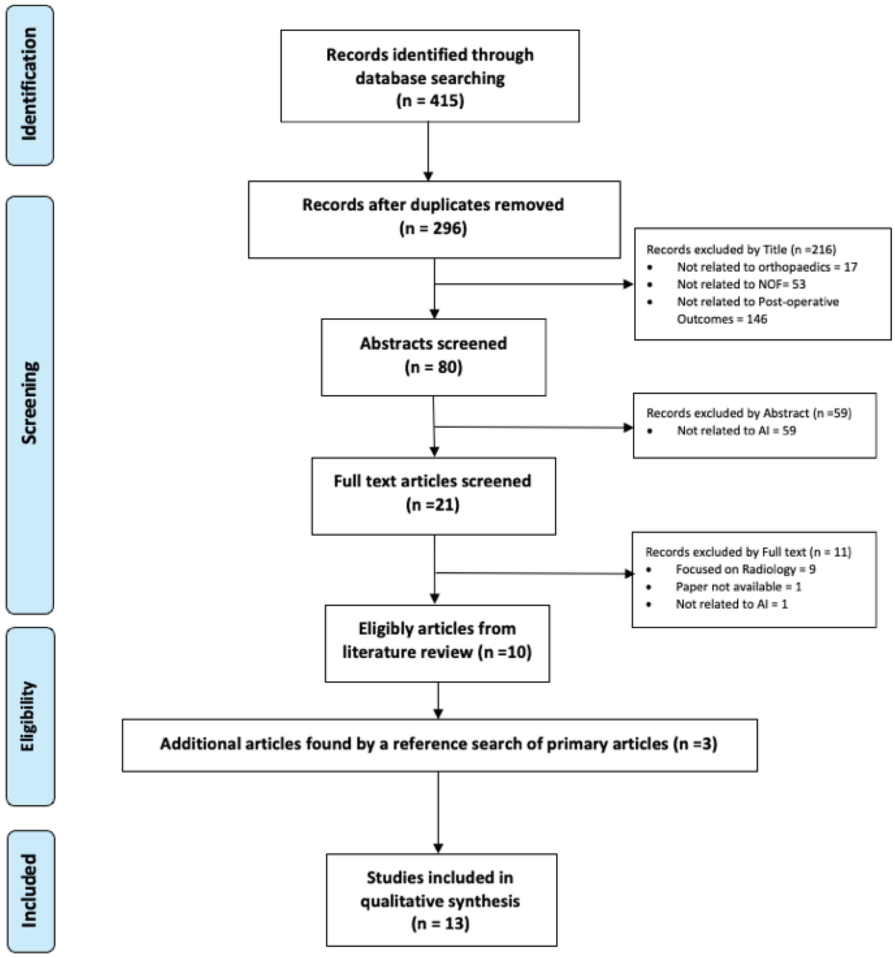

Figure 2. PRISMA diagram outlining the literature search.

Results

The final papers included in the review are displayed in Table 1. Common themes were identified and collated into a structured fashion to understand how healthcare data is used to manage NOFs.

Discussion: thematic analysis of the literature

Machine learning models

ML's advantages in Orthopaedics include identifying trends using an evolving algorithm to handle complex datasets and automating data analysis in real time. However, these functions require a great deal of processing power and time. Moreover, the results depend on the underlying data quality, which clinicians must trust and interpret (Bayliss and Jones, 2019). The review highlighted various ML models used to predict outcomes in NOF patients ranging from supervised to unsupervised. Unsupervised algorithms such as the Artificial Neural Network (ANN) by DeBaun et al (2021) worked exceptionally well compared to other more supervised models such as naive Bayes and logistic regression, thus increasing the validity of the outcome. In other contexts, such as predicting delirium after NOF, a supervised Pen0alised Logistic Regression algorithm superseded others with a Receiver Operating Characteristic (ROC) of 0.79, demonstrating that no one model fits all (Oosterhoff et al, 2021).

When selecting a model, clinicians need to be aware of its flaws. Jaret et al (2019) noted in their limitations that their 'Naive Bayes' models function under the assumption that all inputs are independent. Thus, confounding or modifying variables may not have been identified in the model's development. Although effectively predicting length of stay and cost, the result must be interpreted carefully, and an explanation of why the model arrived at such a conclusion is warranted (Karnuta et al, 2019). On the other hand, algorithms with a boosting function have been shown to be promising in multiple studies (Hillina et al, 1981; Karnuta et al, 2019; Wang et al, 2021; Liu et al, 2022). Liu et al (2022) noted that the effectiveness of gradient boosting is due to increasing the emphasis on observations poorly modelled by a set of existing variables through repeated training rounds.

Multiple ML models can be used when using two data sources as imaging and clinical predictors. Zhu et al (2020) used Convolutional Neural Networks (CNNs) for image recognition of postoperative X-rays, as they had been widely used for the Orthopaedic diagnosis of wrists and ankles fractures, alongside multivariable logistic regression analysis to predict Osteonecrosis of Femoral Head (ONFH) After Internal Fixation. This hybrid approach was mirrored by Zheng et al (2022) when predicting the prognosis of NOFs six Months after Total Hip Arthroplasty. This work showed how multiple methods could be

Table 1. Final papers

Number	Title	Authors	Year
1	Artificial Neural Networks Predict 30-Day Mortality After Hip Fracture: Insights From Machine Learning (DeBaun et al, 2021)	Malcolm R. DeBaun et al	2021
2	Bundled Care for Hip Fractures: A Machine-Learning Approach to an Untenable Patient-Specific Payment Model (Jaret et al, 2019)	Jaret M. Karnuta et al	2019
3	Does the SORG Orthopaedic Research Group Hip Fracture Delirium Algorithm Perform Well on an Independent Intercontinental Cohort of Patients With Hip Fractures Who Are 60 Years or Older? (Oosterhoff et al, 2022)	Jacobien Oosterhoff et al	2022
4	The application of machine learning algorithms in predicting the length of stay following femoral neck fracture (Zhong et al, 2021)	Hao Zhong et al	2021
5	Patients With Femoral Neck Fractures Are at Risk for Conversion to Arthroplasty After Internal Fixation: A Machine - learning Algorithm (Kuit, 2022)	Anouk van de Kuit et al	2022
6	A novel machine-learning algorithm for predicting mortality risk after hip fracture surgery (Li et al, 2021)	Yi Li et al	2021
7	Prediction of acute kidney injury in patients with femoral neck fracture utilising machine learning (Liu et al, 2022)	Jun Liu et al	2022
8	Development and internal validation of a clinical prediction model using machine learning algorithms for 90 day and 2 year mortality in femoral neck fracture patients aged 65 years or above (Hillina et al, 1981)	Hillina J et al	1981
9	Prediction of Postoperative Delirium in Geriatric Hip Fracture Patients: A Clinical Prediction Model Using Machine Learning Algorithms (Oosterhoff et al, 2021)	Jacobien Oosterhoff et al	2021
10	Prediction Model of Osteonecrosis of the Femoral Head After Femoral Neck Fracture: Machine Learning–Based Development and Validation Study (Wang et al, 2021)	Huan Wang	2021
11	Using Naive Bayes Classifier to predict osteonecrosis of the femoral head with cannulated screw fixation (Cui et al, 2018)	Shuangshuang Cui	2018
12	Prediction Models for Prognosis of Femoral Neck–Fracture Patients 6 Months after Total Hip Arthroplasty (Zheng et al, 2022)	Xiaofeng Zhen	2022
13	Deep Learning Improves Osteonecrosis Prediction of Femoral Head After Internal Fixation Using Hybrid Patient and Radiograph Variables (Zhu et al, 2020)	Wanbo Zhu	2020

used to predict outcomes in NOFs. It is, therefore, paramount to pick the suitable model for the right question with an understanding of its limitations. Moreover, any model chosen needs to be justified using rigorous validation.

Validation methodology

ML algorithms need to be validated based on generalisability, accuracy, calibration, and relevance, as in healthcare, the consequences of their failure can risk lives (**Figure 3**) (Hillina et al, 1981; Karnuta et al, 2019; Liu et al, 2022). Oosterhoff had the most rigorous validity assessment across his three papers (Hillina et al, 1981; Karnuta et al, 2019; Liu et al, 2022). Although effective, these methods may not be interpretable due to clinicians' limited understanding of ML. Therefore, the method chosen by Cui et al (2018) uses sensitivity, specificity, and positive and negative predictive values, which may be better communicated to a medical audience.

Oosterhoff sought external validation by running the algorithm used in the US for postoperative delirium in NOFs again in an Australian and New Zealand dataset. This change in context challenged the algorithm's assumptions, including definitions and treatment paradigms. Notably, delirium was defined as occurring within seven days of surgery in the validation (Australia/New Zealand) cohort compared with 30 days in the developmental (US) cohort. Furthermore, for almost half of the patient group, preoperative delirium data were missing, showing a drop in data quality. These factors meant, in the validation study, the algorithm performed no better than other existing and validated instruments for assessing postoperative delirium risk (Oosterhoff et al, 2022).

Another method of assessing validity is comparing the ML tool to the gold standard validated prognostic instrument (Oosterhoff et al, 2022) or against clinicians (Zhu et al, 2020) when no validated tool exists. In evaluation, multiple clinicians with various experiences reflective of those making the decisions should be used. Overall validation is critical, and communicating the validity of an ML tool will help gain trust and support for its use in the future.

The blackbox

The ‘blackbox’ refers to the fact that the most powerful machine learning techniques purchase diagnostic or predictive accuracy at the expense of accessibility. Without explaining

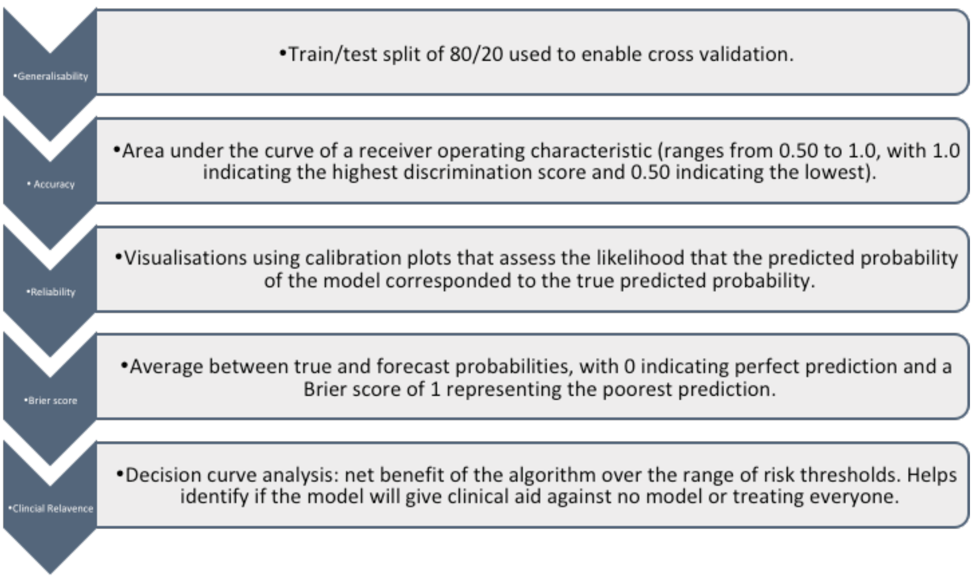

Figure 3. Validation methods (Hillina et al, 1981; Karnuta et al, 2019; Liu et al, 2022).

the output, a lack of interpretability will reduce trust in the outcome (London, 2019). Therefore, the intended benefits may not translate into reality, especially when changing practise, as discovered by Jaret et al (2019) when using ML to predict costings the care of NOFs (Karnuta et al, 2019).

To counteract this issue, Li et al (2021) used variable importance measures, which identified how factors affected mortality rates over time. This model allowed clinicians to recognise that 'in-hospital variables contribute most significantly to short-term mortality risk, whereas admission variables contribute most significantly to long-term mortality risk'. Another effective method used by Wang et al (2021) and Liu et al (2022) is Shapley additive explanations (SHAPs). This post hoc interpretability technique explains the prediction of the outcome (osteoarthritis/acute kidney injury) by displaying the degree of factor contribution, thus empowering clinicians. Oosterhoff effectively used a case-based discussion and tables to visualise the predictive factors when communicating outcomes. Living in a visual society surrounded by software increases the importance of effective data visualisation to engage audiences and patients in their care (Oosterhoff et al, 2021; the Machine Learning Consortium and FAITH Investigators, 2022). In evaluation, ML models can only derive correlation and not causation. Therefore, further work is needed to investigate and explain such predictors to patients and stakeholders.

Predictors and outcomes

Predictors

Factors were chosen based on a mixture of clinical judgement and limitations of the dataset provided. However, statistical methods proved to be vital in streamlining the algorithm. Oosterhoff et al (2021) used random forest plots (R.F.) with recursive selection to identify variables potentially associated with postoperative delirium. Likewise, Zhong et al (2021) used stepwise multiple linear regression analysis to analyse factors associated with the length of stay before applying them to the three ML models chosen for evaluation.

When choosing predictors, the context of the model requires consideration. For example, if the model guides decision-making before surgery to predict the length of stay (LOS), the clinicians will only have preoperative information (Zhong et al, 2021). The model must therefore reflect this by only using preoperative factors. Conversely, clinicians would have intraoperative and postoperative data to predict Osteonecrosis after surgical fixation. Therefore, using these factors in the ML model would be valid (the Machine Learning Consortium and FAITH Investigators, 2022).

The type of data also needs to be considered. In studies using imaging such as X-Ray and C.T., radiological data provides a valuable source for predicting outcomes. Radiomics used by Zheng et al (2022) comprehensively mined image features to capture disease characteristics that are difficult to identify by vision alone and were effectively used in the model to predict prognosis after arthroplasty. However, this study used C.T. imaging, and their use within the U.K. is limited for NOFs, showing the importance of reflecting local context.

Outcomes

Outcomes chosen are based on literature, how they are a surrogate marker of clinical care and their impact on service. The outcomes explored in these studies were mainly clinical (Table 2).

The surrogate markers for these outcomes were limited by the data accessible. For example, serum creatinine was used instead of urine output to depict acute kidney injury (AKI), as the latter was poorly recorded (Liu et al, 2022). Likewise, Osteonecrosis was diagnosed radiologically as clinical interpretation and histology were unavailable (Zhu et al, 2020). Moreover, Oosterhoff showed that the definition of their outcome (Delirium) differed with the U.S. defining it as within seven days of surgery compared with 30 days in New Zealand/Australia. They, therefore, assumed all postoperative delirium events were captured within seven days (Oosterhoff et al, 2022). These issues limit the ML model in

Table 2. Outcomes of studies	
Outcome	**Papers**
Delirium	2
Mortality	2
LOS	3
Osteonecrosis	3
Function	1
AKI	1
Conversion to Arthroplasty	1
Costing Structure	1

LOS, length of stay; AKI, acute kidney injury.

reflecting reality. Therefore, it highlights the need to consider clear diagnostic criteria, surrogate markers, and context when identifying and evaluating the chosen outcome.

The IHI triple aim

The IHI Triple Aim is a framework developed by the Institute for Healthcare Improvement that describes an approach to optimising health system performance through three systems (Berwick et al, 2008; Miranda, 2023). The power of ML in achieving these aims is depicted throughout the literature.

Lower cost of care

ML can aid decision-making and, as shown by Oosterhoff, predict surgical complications such as delirium, thus moving resources to maximise outcomes (Hillina et al, 1981). Predicting LOS, for instance, will help plan budgets and allow departments to highlight patients suitable for discharge. Jaret et al (2019) displayed how payment models can be altered to best reflect the pay provided and create a risk-sharing environment. Moreover, this process can be automated to reduce labour costs and done prospectively to improve efficiency (Karnuta et al, 2019). Efficiency measures, however, need to be deployed cautiously to avoid compromising patient care.

Experience of care

Quality of care complements efficiency as moving resources will result in targeted strategies to improve outcomes (Hillina et al, 1981). ML models need to act as decision-support tools and empower clinicians and patients to make the right decisions for their care. In the short term, this might include the decision to operate based on mortality and LOS (Li et al, 2021). In the long term, it may involve identifying individuals who will need an extended surveillance period based on risks of complications (Zheng et al, 2022). For these benefits to come to fruition, clinicians need to trust the outcomes of ML, which need to be explained to the patient to enable evidence-based decision-making.

Improve health

Population health can also be improved by identifying and exploring the factors relating to outcomes for NOFs and then communicating them to patients. Alongside education, this knowledge can create risk-reduction strategies to prevent complications in the future (Karnuta et al, 2019). Healthcare professionals (HCPs) will also see benefits as the ML models can run in the background through an integrated Electronic Patient Record (EPR), thus allowing quick, accurate decision-making against the pressures of clinical work. Moreover, with ML being used as a potential auxiliary diagnostic tool, the workload for other HCPs, as radiologists, will reduce, thereby improving their overall performance (Liu et al, 2022). In evaluation, ML applications are in their infancy and require quality data sources and infrastructure (DeBaun et al, 2021). Therefore, various macro and micro factors must be considered as they are integrated in the National Healthcare Services (NHS).

Conclusions

This narrative literature review has demonstrated that healthcare data must be leveraged through Artificial Intelligence to predict outcomes for patients with operatively managed NOFs. As displayed, the technology is still in its infancy. However, the potential is vast, and its success will depend on further research to identify the best models with vigorous validation in the local context before deployment. Furthermore, strict governance structures must regulate its adoption alongside training to ensure clinicians best use the technology to achieve the desired efficacy and efficiency gains.

Key points:

- Patients with neck of femur fractures present a tremendous problem that leads to a high incidence of death and dysfunction with an essential factor being postoperative length of stay.
- This narrative literature review has demonstrated that machine learning provides the possibility of accurately predicting the length of hospital stay.
- However, the technology is still in its infancy and its success will depend on further research to identify the best models with vigorous validation in the local context before deployment.
- Strict governance structures must regulate its adoption alongside training to ensure clinicians best use the technology to achieve the desired efficacy and efficiency gains.

Author details
[1]Trauma and Orthopaedics, University Hospitals Leicester, Leicester, UK

Availability of data and materials

All data included in this study are available upon request by contact with the corresponding author.

Author contributions

GS and RA contributed to the conception of the work. GS led the literature review and wrote the article. RA oversaw that the review took place with appropriate methodology and assisted in the writeup of the article. Both authors read and approved the final manuscript. Both authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics approval and consent to participate

Not applicable.

Acknowledgement

With thanks to Megan McCarthy, Evidence Support Manager, The Royal College of Surgeons of England Library and Archives Team, for assisting with the literature review.

Funding

This research received no external funding.

Conflict of interest

The authors declare no conflict of interest.

References

Bayliss L, Jones LD. The role of artificial intelligence and machine learning in predicting orthopaedic outcomes. Role Artif Intellig Mach Learn Predict Orthop Outcom. 2019;101-B(12):1476–1478. https://doi.org/10.1302/0301-620X.101B12.BJJ-2019-0850.R1

Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood). 2008;27(3):759–769. https://doi.org/10.1377/hlthaff.27.3.759

British Orthopaedic Association Standards for Trauma. Patients sustaining a fragility hip fracture. 2012. https://www.boa.ac.uk/static/6750e0bf-4aa3-4680-b1612265704512db/patients%20sustaining%20a%20fragility%20hip%20fracture.pdf

Cui S, Zhao L, Wang Y et al. Using Naive Bayes Classifier to predict osteonecrosis of the femoral head with cannulated screw fixation. Injury. 2018;49(10):1865–1870. https://doi.org/10.1016/j.injury.2018.07.025

DeBaun MR, Chavez G, Fithian A et al. Artificial neural networks predict 30-day mortality after hip fracture: insights from machine learning. J Am Acad Orthop Surg. 2021;29(22):977–983. https://doi.org/10.5435/JAAOS-D-20-00429

Green BN, Johnson CD, Adams A. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. J Chiropr Med. 2006;5(3):101–117. https://doi.org/10.1016/S0899-3467(07)60142-6

Hillina J, Oosterhoff F, Berit A et al. Development and internal validation of a clinical prediction model using machine learning algorithms for 90 day and 2 year mortality in femoral neck fracture patients aged 65 years or above. 1981. https://doi.org/10.1007/s00068-022-01981-4

Inouye SK, Bogardus ST Jr., Charpentier PA et al. A multicomponent intervention to prevent delirium in hospitalized older patients. N Engl J Med. 1999;340(9):669–676. https://doi.org/10.1056/NEJM199903043400901

Karnuta JM, Navarro SM, Haeberle HS et al. Bundled care for hip fractures: a machine-learning approach to an untenable patient-specific payment model. J Orthop Trauma. 2019;33(7):324–330. https://doi.org/10.1097/BOT.0000000000001454

Li Y, Chen M, Lv H et al. A novel machine-learning algorithm for predicting mortality risk after hip fracture surgery. Injury. 2021;52(6):1487–1493. https://doi.org/10.1016/j.injury.2020.12.008

Liu J, Xu L, Zhu E, Han C, Ai Z. Prediction of acute kidney injury in patients with femoral neck fracture utilizing machine learning. Front Surg. 2022;9. https://doi.org/10.3389/fsurg.2022.928750

London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep. 2019;49(1):15–21. https://doi.org/10.1002/hast.973

Maguire D, Honeyman M, Fenney D, Jabbal J. Shaping the future of digital technology in health and social care. The King's Fund. 2021. https://assets.kingsfund.org.uk/f/256914/x/f6444844fd/shaping_future_digital_technology_health_social_care_2021.pdf

Midway SR. Principles of effective data visualization. Patterns. 2020;1(9):100141. https://doi.org/10.1016/j.patter.2020.100141

Miranda K. Healthcare triple aim: how Nursa helps hospitals and clinicians, healthcare triple aim: how Nursa helps hospitals and clinicians. 2023. https://nursa.com/blog/healthcare-triple-aim-applied-in-healthcare-setting (accessed 16 April 2024)

Oosterhoff JHF, Karhade AV, Oberai T et al. Prediction of postoperative delirium in geriatric hip fracture patients: a clinical prediction model using machine learning algorithms. Geriatr Orthop Surg Rehabil. 2021;12:215145932110622. https://doi.org/10.1177/21514593211062277

Oosterhoff JHF, Oberai T, Karhade AV et al. Does the SORG orthopaedic research group hip fracture delirium algorithm perform well on an independent intercontinental cohort of patients with hip fractures who are 60 years or older? Clin Orthop Relat Res. 2022;480(11):2205–2213. https://doi.org/10.1097/CORR.0000000000002246

Rumrill PD, Fitzgerald SM. Using narrative literature reviews to build a scientific knowledge base. Work. 2001;16(2):165–170.

van de Kuit A, Oosterhoff JHF, Dijkstra H et al. Patients with femoral neck fractures are at risk for conversion to arthroplasty after internal fixation: a machine-learning algorithm. Clin Orthop Relat Res. 2022;480(12):2350–2360. https://doi.org/10.1097/CORR.0000000000002283

Wang H, Wu W, Han C et al. Prediction model of osteonecrosis of the femoral head after femoral neck fracture: machine learning-based development and validation study. JMIR Med Inform. 2021;9(11):e30079. https://doi.org/10.2196/30079

Wiles MD, Moran CG, Sahota O, Moppett IK. Nottingham Hip Fracture Score as a predictor of one year mortality in patients undergoing surgical repair of fractured neck of femur. Br J Anaesth. 2011;106(4):501–504. https://doi.org/10.1093/bja/aeq405

Zheng X, Xiao C, Xie Z, Liu L, Chen Y. Prediction models for prognosis of femoral neck–fracture patients 6 months after total hip arthroplasty. IJGM. 2022;15:4339–4356. https://doi.org/10.2147/IJGM.S347425

Zhong H, Wang B, Wang D et al. The application of machine learning algorithms in predicting the length of stay following femoral neck fracture. Int J Med Inform. 2021;155:104572. https://doi.org/10.1016/j.ijmedinf.2021.104572

Zhu W, Zhang X, Fang S, Wang B, Zhu C. Deep learning improves osteonecrosis prediction of femoral head after internal fixation using hybrid patient and radiograph variables. Front Med. 2020;7. https://doi.org/10.3389/fmed.2020.573522

Appendix

Appendix 1. Literature review search in Medline

Database	MEDLINE	Results per line:	Final Result
1	Femoral Neck Fractures/su [Surgery]	5753	249
2	((hip* or femur* or femoral*) adj3 (neck or proximal or intracapsul*) adj4 fracture*).ti,ab,kw,kf.	13327	
3	(arthroplast* or hemiarthroplast* or surger* or surgeon* or surgical* or operat*).ti,ab,kw,kf.	3131470	
4	2 and 3	7342	
5	1 or 4	10256	
6	Femoral Neck Fractures/	9607	
7	((hip* or femur* or femoral*) adj3 (neck or proximal or intracapsul*) adj4 fracture*).ti,ab,kw,kf.	13327	
8	6 or 7	16079	
9	Arthroplasty, Replacement, Hip/	33139	
10	(arthroplast* or hemiarthroplast* or surger* or surgeon* or surgical* or operat*).ti,ab,kw,kf.	3131470	
11	9 or 10	3136122	
12	8 and 11	8384	
13	5 or 12	10489	
14	exp Artificial Intelligence/ or exp machine learning/ or exp neural networks, computer/ or deep learning/	159826	
15	medical informatics/ or exp medical informatics applications/ or exp medical informatics computing/	491367	
16	((machine* or computer* or artificial* or deep) adj2 (intelligence or learning)).ti,ab,kw,kf.	123304	
17	((machine* or artificial* or deep or predict*) adj3 (model* or analy*)).ti,ab,kw,kf.	242730	
18	AI.ti,ab,kw,kf.	39752	
19	((neural or neuronal) adj1 network*).ti,ab,kw,kf.	94271	
20	algorithm*.ti,ab,kw,kf.	333095	
21	radiomic*.ti,ab,kw,kf.	7019	
22	14 or 15 or 16 or 17 or 18 or 19 or 20 or 21	1156168	
23	13 and 22	279	
24	letter/	1198833	
25	editorial/	625805	
26	news/	215052	
27	exp historical article/	409093	
28	anecdotes as topic/	4747	
29	comment/	985895	
30	case report/	2301706	
31	(letter or comment*).ti.	182320	
32	(abstract or comment or conference or letter).pt.	1680563	
33	24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32	4831102	
34	23 not 33	271	
35	limit 34 to english language	249	

The wildcard symbol (*) that broadens a search by finding words that start with the same letters.

Appendix 2. Literature review search in Embase			
Database	**Embase**	**Results per line:**	**Final Result**
1	femoral neck fracture/su [Surgery]	1879	166
2	((hip* or femur* or femoral*) adj3 (neck or proximal or intracapsul*) adj4 fracture*).ti,ab,kw,kf.	15380	
3	(arthroplast* or hemiarthroplast* or surger* or surgeon* or surgical* or operat*).ti,ab,kw,kf.	4055097	
4	2 and 3	9093	
5	1 or 4	9795	
6	femoral neck fracture/	4158	
7	((hip* or femur* or femoral*) adj3 (neck or proximal or intracapsul*) adj4 fracture*).ti,ab,kw,kf.	15380	
8	6 or 7	16635	
9	exp hip arthroplasty/	34961	
10	(arthroplast* or hemiarthroplast* or surger* or surgeon* or surgical* or operat*).ti,ab,kw,kf.	4055097	
11	9 or 10	4060766	
12	8 and 11	10001	
13	5 or 12	10293	
14	exp Artificial Intelligence/ or deep neural network/ or exp artificial neural network/ or exp machine learning/ or exp deep learning/	370292	
15	exp medical informatics/	22646	
16	((machine* or computer* or artificial* or deep) adj2 (intelligence or learning)).ti,ab,kw,kf.	145945	
17	((machine* or artificial* or deep or predict*) adj3 (model* or analy*)).ti,ab,kw,kf.	314997	
18	AI.ti,ab,kw,kf.	53575	
19	((neural or neuronal) adj1 network*).ti,ab,kw,kf.	114201	
20	algorithm*.ti,ab,kw,kf.	424927	
21	radiomic*.ti,ab,kw,kf.	10001	
22	14 or 15 or 16 or 17 or 18 or 19 or 20 or 21	1055942	
23	13 and 22	226	
24	letter/ or case report/ or case study/	3816545	
25	(letter or comment*).ti.	226974	
26	(abstract or comment or conference or letter or editorial or note).pt.	8273420	
27	24 or 25 or 26	10540984	
28	23 not 27	175	
29	limit 28 to english language	166	

The wildcard symbol (*) that broadens a search by finding words that start with the same letters.