STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(a) Machine learning applications for the prediction of eLOS in geriatric hip fracture patients: a case control study
		Background: Machine learning (ML) has become prevalent in clinical data processing and predictive models. It has been applied to investigate the length of stay (LOS) for hip fracture patients utilizing the enhanced recovery after surgery (ERAS) concept. This study aims to develop ML models for predicting extended LOS (eLOS) among geriatric patients with hip fractures and identifying the associated risk factors. Methods: A retrospective study was conducted at a single orthopaedic trauma centre, enrolling all patients who underwent hip fracture surgery between January 2018 and December 2022. The study collected various patient characteristics, encompassing demographic data, general health status, injury-related data, laboratory examinations, surgery-related data, and LOS. Features that exhibited significant differences in univariate analysis were integrated into the ML model establishment and subsequently cross-verified. The study compared the performance of the ML models and determined the risk factors for eLOS. Results: The study included 763 patients, with 380 experiencing eLOS. Among the models, the decision tree, random forest, and eXtreme Gradient Boosting models demonstrated the most robust performance. Notably, the artificial neural network model also exhibited impressive results. After cross-validation, the support vector machine and logistic regression models demonstrated superior performance. Predictors for eLOS included delayed surgery, D-dimer level, American Society of Anaesthesiologists (ASA) classification, type of surgery, and sex. Conclusions: Machine learning proved to be highly accurate in predicting eLOS for geriatric patients with hip fractures. The identified key risk factors were delayed surgery, D-dimer level, ASA classification, type of surgery, and sex. This valuable information can aid clinicians in allocating resources more efficiently to meet patient demand effectively.
Introduction		
Background/rationale	2	Hip fractures have become more prevalent as the global geriatric population increases [1]. They are associated with higher incidence, mortality, and disability, significantly impacting the quality of life of affected individuals [2, 3]. Prolonged length of stay (LOS) not only places a financial burden on patients but also elevates the risk of mortality and complications [4]. Enhanced recovery after surgery (ERAS) refers to the integration of perioperative concepts using evidence-based medicine tools to reduce surgical stress and complications, shorten hospital stays, lower financial costs, and hasten postoperative recovery [5-7]. This concept was introduced by Danish surgeon Kehlet in the 1990s and has become a critical aspect of modern surgery in the 21st century [8]. In recent

		years, ERAS has been increasingly applied in traumatic orthopaedics, yielding favourable outcomes [9, 10]. Additionally, Andrew et al. developed a logistic regression model to identify risk factors for extended length of stay (eLOS), offering new insights for optimizing treatment for hip fracture patients under the ERAS concept [11]. However, traditional statistical methods suffer from poor performance and lack of features. Machine learning (ML) is a scientific discipline focused on teaching computers to learn from data, first proposed by Arthur Samuel in 1959 [12]. In recent times, ML has shown superior predictive performance compared to traditional methods and has found extensive application in clinical data processing and predictive modelling [13, 14]. In the context of hip fractures among geriatric individuals, Shtar et al. and Oosterhoff et al. established ML models to predict prognosis and mortality, enhancing clinician decision-making ability [15, 16].
Objectives	3	This study aims to develop ML models for predicting eLOS among geriatric patients with hip fractures, identify associated risk factors, and compare the performance of each model.
Methods		
Study design	4	A retrospective study was conducted at a single orthopaedic trauma centre between January 2018 and December 2022. The study employed specific inclusion and exclusion criteria as follows: Inclusion Criteria: 1. Age older than 60 years at the time of injury, 2. Confirmed diagnosis of hip fracture, 3. Hospitalization at our centre. Exclusion Criteria: 1. Admission with multiple fractures, pathological fracture, or fractures around the prosthesis, 2. Receipt of conservative treatment due to severe comorbidities, 3. Presence of missing data. The enrolled patients had a median hospital stay of 9.5 days. Based on this median length of stay, the patients were retrospectively divided into two groups: noneLOS (length of stay ≤ 9.5 days, n = 383, 50.2%) and eLOS (length of stay > 9.5 days, n = 380, 49.8%).
Setting	5	Data for the study were retrospectively gathered from electronic patient records at the institution. Demographic data encompassed sex, age, body mass index (BMI), general health status categorized by the American Society of Anaesthesiologists (ASA) classification, history of smoking, oral anticoagulant use, and comorbidities [18]. Injury-related data included fracture type, time from injury to admission, and the day of admission. Surgery-related data consisted of the type of surgery, anaesthesia used, ICU transfer, time to surgery, duration of surgery, intraoperative blood loss, and transfusion. Laboratory examinations conducted at admission and after surgery were also collected.

		Age was stratified into 60 – 85 and > 85 age groups; ASA classification was grouped as I-II and III-IV; admission day was grouped into Monday to Thursday and Friday to Sunday; injury time was stratified into ≤ 24 and >24 hours; and delayed surgery was defined as an operation performed more than 48 hours after admission. Laboratory examinations were stratified according to normal values.
Participants	6	(*a*) A retrospective study was conducted at a single orthopaedic trauma centre between January 2018 and December 2022. The study employed specific inclusion and exclusion criteria as follows: Inclusion Criteria: 1. Age older than 60 years at the time of injury, 2. Confirmed diagnosis of hip fracture, 3. Hospitalization at our centre. Exclusion Criteria: 1. Admission with multiple fractures, pathological fracture, or fractures around the prosthesis, 2. Receipt of conservative treatment due to severe comorbidities, 3. Presence of missing data.
		(*b*) NA
Variables	7	The enrolled patients had a median hospital stay of 9.5 days. Based on this median length of stay, the patients were retrospectively divided into two groups: noneLOS (length of stay ≤ 9.5 days, n = 383, 50.2%) and eLOS (length of stay > 9.5 days, n = 380, 49.8%). Age was stratified into 60 – 85 and > 85 age groups; ASA classification was grouped as I-II and III-IV; admission day was grouped into Monday to Thursday and Friday to Sunday; injury time was stratified into ≤ 24 and >24 hours; and delayed surgery was defined as an operation performed more than 48 hours after admission. Laboratory examinations were stratified according to normal values.
Data sources/ measurement	8*	Data for the study were retrospectively gathered from electronic patient records at the institution. Demographic data encompassed sex, age, body mass index (BMI), general health status categorized by the American Society of Anaesthesiologists (ASA) classification, history of smoking, oral anticoagulant use, and comorbidities [18]. Injury-related data included fracture type, time from injury to admission, and the day of admission. Surgery-related data consisted of the type of surgery, anaesthesia used, ICU transfer, time to surgery, duration of surgery, intraoperative blood loss, and transfusion. Laboratory examinations conducted at admission and after surgery were also collected.
Bias	9	NA
Study size	10	The number of cases admitted to the Trauma Centre of Zhongda Hospital during the study period determined the sample size.
Quantitative variables	11	Age was stratified into 60 – 85 and > 85 age groups; ASA classification was grouped as I-II and III-IV; admission day was grouped into Monday to Thursday and Friday to Sunday; injury time was stratified into ≤ 24 and >24 hours; and delayed surgery was defined as an operation performed more than 48 hours after admission. Laboratory examinations were stratified according to normal values.
Statistical methods	12	(*a*) In the study, normally distributed data are presented as the means and standard deviations (SDs). Nonnormally distributed variables were expressed as medians along with their interquartile ranges (IQRs). Categorical variables were represented as

counts and percentages. To analyse the overall data, continuous variables were subjected to Student's t test or the Mann–Whitney U test, depending on their distribution. Categorical variables were analysed using the chi-square test, as appropriate. Variables showing significant differences in the univariate analysis were selected and included in the establishment of the machine learning (ML) model.
The predictive eLOS ML model was established according to the selected features, including basic algorithms for logistic regression (LR), decision tree (DT), random forest (RF), support vector machine (SVM), naïve Bayes (NB), K-nearest neighbour (KNN), eXtreme Gradient Boosting (XGBoost) and artificial neural network (ANN) models. Each ML model was integrated to ascertain feature importance. Then, the original data were split into a training set and a test set (training: test = 7:3), and 10-fold cross-validation was carried out. The confusion matrix, accuracy score, precision score, recall score, F1 score, receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) were used to evaluate the performance of the ML model of the original data and cross-validation. All statistical analyses were conducted using Python (version 3.8.2, Python Software Foundation, https://www.python.org) and the sklearn package (version 0.24.1). A 2-sided P value < 0.05 was considered significant. The flow diagram of the research process is shown in Fig. 1.

(*b*) NA

(*c*) NA

(*d*) NA

(*e*) NA

Continued on next page

Results		
Participants	13*	(a) Overall, 763 patients were enrolled in the final analysis; patients were divided into noneLOS (n = 383) and eLOS (n = 380) groups based on length of stay.
		(b) NA
		(c) NA
Descriptive data	14*	(a) The characteristics of the two groups are compared in Table 1.
		(b) NA
		(c) NA
Outcome data	15*	NA
		Overall, 763 patients were enrolled in the final analysis; patients were divided into noneLOS (n = 383) and eLOS (n = 380) groups based on length of stay.
		NA
Main results	16	(*a*) Univariate analysis showed that there were significant differences in sex, fracture type, ASA classification, admission day, injury to admission time, hypertension, diabetes, cerebral infarction, deep venous thrombosis (DVT), delayed surgery, THA, reduction and fixation, AST and D-dimer levels and aortic velocity (AV) at admission between the two groups ($P < 0.05$).
		(*b*) NA
		(*c*) NA
Other analyses	17	We used 8 ML models to evaluate the predictors of LOS in the original data. Figure 2 shows the ROC curve, and Table 2 shows the performance indicators of each model. The tree models, including the DT (accuracy = 0.924, AUC = 0.988), RF (accuracy = 0.924, AUC = 0.985) and XGBoost (accuracy = 0.912, AUC = 0.976) models, showed stronger performance among the models. In addition, the performance of the ANN (accuracy = 0.886, AUC = 0.963) model was impressive.
Discussion		
Key results	18	This study aimed to develop machine learning (ML) models for predicting extended length of stay (eLOS) in geriatric patients with hip fractures and to identify associated risk factors. Additionally, we assessed and compared the performance of each ML model. The DT and ANN models demonstrated the best performance with the original data. After cross-validation, the SVM and LR models also performed well.
Limitations	19	However, there are several limitations to consider in our study. First, it was a single-centre study, and the length of hospital stay might vary significantly across different healthcare systems. Moreover, the high proportion of patients with ASA III-IV in our hospital indicates a higher prevalence of severe comorbidities and advanced disease compared to those treated in the community, leading to potential selection bias. Second, since this study aimed to establish ML models, the sample size might be relatively small, resulting in some ML models being prone to overfitting. As a result, the findings of this study should be further validated and made applicable to a broader population through multicentre and large-sample studies.
Interpretation	20	In conclusion, predicting extended length of stay (eLOS) aligns well with the Enhanced Recovery After Surgery (ERAS) concept. In this study, we successfully developed machine learning (ML) models to predict eLOS and identify associated risk factors among hip fracture patients. Notably, delayed surgery, D-dimer level, ASA classification, type of surgery, and sex were found to be significantly related to eLOS. Comparing the performance of ML models against traditional statistical methods revealed the superior accuracy of ML models. The significance of this study lies in the application of ML, which empowers clinicians to

		make early and informed clinical decisions, mitigate potential risks, and effectively allocate medical resources. By leveraging ML technology, healthcare professionals can optimize patient care and enhance patient outcomes in the context of the ERAS concept.
Generalisability	21	NA
Other information		
Funding	22	There was no funding or support for this study.

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.