## RESEARCH



# A radiomics and deep learning nomogram developed and validated for predicting no-collapse survival in patients with osteonecrosis after multiple drilling

Fan Liu<sup>1</sup>, De-bao Zhang<sup>1</sup>, Shi-huan Cheng<sup>2</sup> and Gui-shan Gu<sup>1\*</sup>

## Abstract

**Purpose** Identifying patients who may benefit from multiple drilling are crucial. Hence, the purpose of the study is to utilize radiomics and deep learning for predicting no-collapse survival in patients with femoral head osteonecrosis.

**Methods** Patients who underwent multiple drilling were enrolled. Radiomics and deep learning features were extracted from pelvic radiographs and selected by LASSO-COX regression, radiomics and DL signature were then built. The clinical variables were selected through univariate and multivariate Cox regression analysis, and the clinical, radiomics, DL and DLRC model were constructed. Model performance was evaluated using the concordance index (C-index), area under the receiver operating characteristic curve (AUC), net reclassification index (NRI), integrated discrimination improvement (IDI), calibration curves, and Decision Curve Analysis (DCA).

**Results** A total of 144 patients (212 hips) were included in the study. ARCO classification, bone marrow edema, and combined necrotic angle were identified as independent risk factors for collapse. The DLRC model exhibited superior discrimination ability with higher C-index of 0.78 (95%CI: 0.73–0.84) and AUC values (0.83 and 0.87) than other models. The DLRC model demonstrated superior predictive performance with a higher C-index of 0.78 (95% CI: 0.73–0.84) and area under the curve (AUC) values of 0.83 for 3-year survival and 0.87 for 5-year survival, outperforming other models. The DLRC model also exhibited favorable calibration and clinical utility, with Kaplan–Meier survival curves revealing significant differences in survival rates between high-risk and low-risk cohorts.

**Conclusion** This study introduces a novel approach that integrates radiomics and deep learning techniques and demonstrates superior predictive performance for no-collapse survival after multiple drilling. It offers enhanced discrimination ability, favorable calibration, and strong clinical utility, making it a valuable tool for stratifying patients into high-risk and low-risk groups. The model has the potential to provide personalized risk assessment, guiding treatment decisions and improving outcomes in patients with osteonecrosis of the femoral head.

Keywords Osteonecrosis of the femoral head, Core decompression, Multiple drilling, Radiomics, Deep learning

\*Correspondence: Gui-shan Gu gugs@jlu.edu.cn Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

## Background

Nontraumatic osteonecrosis of the femoral head (NONFH) arises from various factors that impact the blood supply to the femoral head, leading to the necrosis of bone cells and marrow tissue [1, 2]. In China, there are approximately 8.12 million ONFH patients, with an annual incidence of new cases ranging from 75,000 to 150,000, primarily affecting individuals aged between 30 and 50 years [3, 4]. If left untreated, approximately 80% of NONFH cases may progress to collapse, necessitating total hip arthroplasty (THA) [5]. However, outcomes following THA in younger NONTH patients are often suboptimal, frequently requiring subsequent revision surgeries. Consequently, the primary objective is to preserve the femoral head and prevent collapse [6]. Although the optimal approach to preserving the femoral head remains controversial, core decompression (CD) currently ranks among the most performed hip preservation surgeries in both the United States and China [7-12]. Studies have indicated an increased risk of further collapse and fractures following conventional CD procedures. However, the utilization of multiple drilling (MD) has been associated with a reduction in these risks, offering a simpler surgical procedure while demonstrating comparable success rates [13].

However, due to variations in research methodologies, differing outcome definitions, and variations in followup, the effectiveness of MD remains debatable [13–15]. While some studies suggest that clinical symptoms can be improved after MD [13, 14], others, such as the study conducted by Liu et al., suggest that MD may not mitigate the rate of THA, and patients with predisposing factors such as extensive lesions and bone marrow edema may face an elevated risk of collapse [15]. Hence, predicting the risk of femoral head collapse and survival time following MD, as well as identifying patients who may benefit from it, are crucial endeavors. While several studies have identified risk factors associated with postoperative collapse after MD, these factors remain unvalidated [16]. Two studies have developed scoring systems or prediction models, yet one of them lacks further validation, while the other underwent validation within the same cohort [17, 18]. Radiomics and deep learning (DL) are novel methods that are more commonly applied in the diagnosis and prognosis research within oncology, and they have shown promising results [19]. However, their application is less common in the field of orthopedics. Therefore, the primary objective of this study is to employ radiomics and deep learning to develop and validate a novel model for predicting postoperative nocollapse survival in non-traumatic femoral head osteonecrosis who underwent multiple drilling.

## Methods

## Study design

This observational retrospective study was approved by the Ethics Committee of the First Hospital of Jilin University, and patient informed consent was waived in accordance with the Helsinki Declaration. Patients with early stage nontraumatic osteonecrosis of the femoral head who underwent multiple drilling core decompression in our hospital from January 2015 to January 2022 and followed for at least 6 months were included. The flow chart of the study methodology is depicted in Fig. 1a.

The inclusion criteria were as follows: (1). Diagnosis of nontraumatic osteonecrosis of the femoral head. (2). Age between 18 and 75 years. (3). Patients with ARCO 3A who refused THA. (4). No prior treatment for osteonecrosis of the femoral head. (5). A minimum follow-up period of 6 months. (6). Availability of a complete preoperative X-ray in DICOM (Digital Imaging and Communications in Medicine) format, taken one month prior to surgery. The exclusion criteria were as follows: (1). Traumatic osteonecrosis of the femoral head. (2). Classified as ARCO IIIB or above, indicating collapse of the femoral head. (3). Presence of contraindications to zoledronic acid, such as severe renal insufficiency. (4). Complications including fracture, infections, and deep vein thrombosis (DVT).

The staging adhered to the 2019 ARCO staging system based on a comprehensive radiological evaluation conducted before surgery, including an anterior-posterior pelvic radiograph, a frog-leg lateral radiograph, and magnetic resonance imaging (MRI) of the hip [20]. Lesion location was categorized into four types (Type A, B, C1, and C2) based on the midcoronal T1-weighted MRI images, following the classification described by Sugano [21]. Given that sagittal slices were not included in the MRI images of several studies, the arc of the necrosis surface on the femoral head was measured on midcoronal and mid-axis T1-weighted MRI images, documented as A and B, respectively. We then calculated the combined necrotic angle by summing the values of A and B and determined the necrosis degree index using the formula (A/180) \* (B/180) \* 100, as outlined by Cherian [22]. Additionally, bone marrow edema was documented, with assessment criteria based on the T2-weighted MRI images. Clinical data, comprising age, gender, body mass index (BMI), etiologies (corticosteroid-induced, alcoholrelated, and idiopathic) and the duration of symptoms were extracted from the medical records.

## Surgical procedures and bisphosphonate medication

The patients were placed in a supine position, with the affected hip slightly elevated and the lower extremity rotated inward by 15 degrees. Three to four Kirschner



in patients after multiple drilling. DLRC: Deep Learning-Radiomics-Clinical

wires, each with a diameter of 3.0 mm, were carefully inserted into the femoral head and neck. The wires were advanced until they reached the necrotic area, typically located 0.5–1.0 cm below the articular surface. After complete decompression, the Kirschner wires were removed. Patients were allowed to ambulate with crutches, and weight-bearing restrictions were imposed for six months. They were advised to avoid excessive weight-bearing, jumping, and strenuous activities for one year. Additionally, some patients received a 5 mg intravenous infusion of zoledronic acid one month after surgery, as recommended by the medical team. They were also prescribed daily supplements of calcium (500–1000 mg) and vitamin D3 (400–800 units). Patients who received a weekly dose of 70 mg of alendronate sodium for a three-month duration were also included in the study, as this regimen provides an equivalent dose to the intravenous injection of 5 mg of zoledronic acid [23].

## Follow-up and assessment

Patients were followed up at 6 months, 12 months, and then annually, and whenever they experienced worsening hip pain or limited activity. Diagnostic imaging, including pelvic radiographs and hip MRIs, was performed at each follow-up. The Harris Hip Score was used to assess the functional recovery of the hip joint before the surgery and at 12 and 24 months postoperatively. The main outcome of interest was no-collapse survival (femoral head depression < 2 mm), determined by pelvic radiographs or MRIs, and a collapse greater than 2 mm was considered a failure [24]. The imaging evaluation was conducted by two evaluators (Zhang and Li), with any disagreements resolved by a third senior surgeon (Gu). Secondary outcome measures included complications related to core decompression, such as intertrochanteric or femoral neck fractures and deep vein thrombosis. Additionally, adverse reactions to intravenous zoledronic acid, which could manifest as fever, flu-like symptoms, and musculoskeletal pain, were also monitored. Patients with missing data were excluded from the study analysis.

## Image preprocessing and extraction of radiomics and DL features

Anterior-posterior pelvic radiographs were obtained from the Picture Archiving and Communication System (PACS) in the Digital Imaging and Communications in Medicine (DICOM) format after anonymization. Subsequently, these radiographs were converted to the Neuroimaging Informatics Technology Initiative (NIfTI) format using the dicom2nii package, a Python-based tool. Following image intensity normalization, experienced orthopedic surgeons Zhang and Li, each possessing over 10 years of expertise, manually delineated regions of interest (ROI) within the femoral head in the ITK-SNAP software (version 3.4, http://www.itksnap.org) [25]. A variety of radiomics features, including first-order statistical features, shape-based features, texture features, and high-order features, were extracted from each patient using the PyRadiomics package [26]. To incorporate deep learning features, the pretrained ResNet18 architecture was included and deep learning (DL) features were extracted from the average pooling layer.

## Feature selection and radiomics and DL signature establishment

The feature selection process for radiomics and DL features involved the following steps: (1). Features with missing values or outliers were eliminated; (2). Data standardization was carried out through Z-score transformation to achieve a mean of 0 and a variance of 1; (3). Features with intraclass and interclass correlation coefficients below 0.75 were excluded; (4). A

Mann–Whitney U test was conducted, and only features with a *P*-value < 0.05 were kept; (5). Pearson's correlation coefficients were calculated among the features, and one of the features was retained when the correlation coefficient exceeded 0.9 between any two features; (6). The Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression model was applied to using tenfold cross-validation. Features with non-zero coefficients were selected based on the minimum lambda value. These selected features, along with their respective weights, were retained to construct a radiomics and DL signature for all patients.

## Model construction and evaluation

Univariate and multivariate Cox regression analyses were conducted on the clinical variables to identify risk factors. Subsequently, a clinical model was constructed using the identified factors. Furthermore, a Radiomics model was developed by integrating the radiomic signature with clinical variables. Simultaneously, a Deep Learning (DL) model was created by incorporating the deep learning signature along with clinical factors. Following these steps, a Deep Learning-Radiomics-Clinical model (DLRC) was formulated, which amalgamated the DL signature, radiomics signature, and clinical variables.

Model performance was assessed using the concordance index (C-index) and receiver operating characteristic (ROC) curve analysis to calculate the area under the ROC curve (AUC). Additionally, the net reclassification index (NRI) and integrated discrimination improvement (IDI) were computed to compare the model performance. Calibration efficiency of the model was evaluated through the generation of calibration curves, while assessment of the clinical utility of the predictive models involved Decision Curve Analysis (DCA).

#### Statistical analysis

Statistical analyses were performed using R software (version 4.3.1, http://www.r-project.org). The normality of continuous variables was evaluated using the Kolmogorov-Smirnov test. Continuous variables following a normal distribution are presented as mean ± standard deviation, while those not conforming to normality are described using the median and interquartile range (Q25-Q75). Group differences were assessed using Student's t-test or the Mann-Whitney U test, depending on the distribution of the data. Categorical data were reported as frequency (percentage) and analyzed using the Chisquare test or Fisher's exact test. When conducting Cox regression analysis, continuous variables were discretized into categorical variables based on their median values. Covariates with a significance level of p < 0.1 in the univariate Cox regression analysis were subsequently included in the multivariate Cox regression analysis using the backward elimination approach. Risk scores were computed for each patient according to the nomogram. Subsequently, using the median score, the samples were stratified into high-risk and low-risk groups. Kaplan– Meier survival curves were constructed and statistical differences were assessed using log-rank tests. The level of statistical significance was set at p < 0.05.

## Results

## **Baseline characteristics**

A total of 144 patients (212 hips) underwent multiple drilling procedures and were included in this study. The median follow-up period was 29.55 months (19.48-40.83). At the end of the study, 64 hips (43.24%) experienced collapse. The overall no-collapse survival rate was 39.75% (95% CI, 28.22%-55.98%). The postoperative 3-year no-collapse survival rate was 71.59% (95% CI, 64.50%-79.46%), and the 5-year no-collapse survival rate was 44.72% (95% CI, 34.73%-57.57%). Among the patients, 80.19% (170/212) were male, with a median Body Mass Index BMI of 23.71 (21.58-26.04). The median age at surgery was 42.00 years (33.00-49.00), and the median symptom duration before surgery was 2.00 months (1.00-5.00). Bisphosphonate medication was administered to 45.75% (97/212) of the hips. Simultaneous contralateral THA was performed in 10.85% (23/212) of the hips, while bilateral MD was performed in 62.26% (132/212). We identified 58 hips (27.36%) attributed to corticosteroid use, 33 hips (15.57%) associated with alcohol consumption, and 121 hips (57.08%) categorized as idiopathic. Apart from Bone Marrow Edema, Combined Necrotic Angle, and Necrosis Degree Index, no statistically significant differences were observed between the collapse and no-collapse groups in other baseline characteristics (Table 1).

## Selection of clinical factors

In the analysis, twelve clinical variables were subjected to univariate and multivariate Cox regression analyses. The results identified ARCO classification (hazard ratio [HR], 2.12; 95% confidence interval [CI], 1.25—3.58; P=0.037), bone marrow edema (BME) (HR, 2.13; 95% CI, 1.19— 3.81; P=0.010), and combined necrotic angle (HR, 2.41; 95% CI, 1.43—4.07; P<0.001) as independent risk factors for femoral head collapse after MD. It is noteworthy that the administration of bisphosphonate (HR, 1.14; 95% CI, 0.69—1.89; P=0.602) was not a significant risk factor. The criteria for selecting significant variables were based on a P-value threshold of <0.05 in the multivariate Cox regression model. These findings emphasize the importance of ARCO classification, bone marrow edema, and combined necrotic angle in predicting femoral head collapse risk. Detailed results of the univariate and multivariate Cox regression analyses are presented in Table 2.

#### Model construction and evaluation

For each patient, a total of 1562 radiomics features and 512 deep learning (DL) features were extracted. Feature selection was performed using LASSO regression with cross-validation to determine the optimal lambda values. A total of 14 radiomics features with non-zero coefficients were identified at an optimal lambda value of 0.054, while 21 DL features with non-zero coefficients were identified at an optimal lambda value of 0.712 (Fig. 2 a-d). The radiomics and DL signatures were computed as linear combinations of these selected features, weighted by their respective coefficients.

The performance of the radiomics, DL, and DLRC models surpassed that of the clinical model, as demonstrated by both C-index and AUC metrics. The C-indices for the radiomics, DL, and DLRC models were 0.73 (95% CI: 0.70-0.82), 0.76 (95% CI: 0.70-0.82), and 0.78 (95% CI: 0.73-0.84), respectively, compared to the clinical model's C-index of 0.64 (95% CI: 0.56-0.71) (Table 3). For the 3-year AUC values (Table 3, Fig. 3a), the clinical model achieved 0.65, whereas the radiomics, DL, and DLRC models achieved 0.80, 0.80, and 0.83, respectively. Similarly, for the 5-year AUC values (Table 3, Fig. 3b), the clinical, radiomics, DL, and DLRC models yielded 0.77, 0.87, 0.84, and 0.87, respectively. Notably, the DLRC model demonstrated superior predictive performance, achieving the highest C-index and 3-year AUC value among all models.

According to the IDI values (Supplemental Table 1), both the radiomics, DL, and DLRC models outperformed the clinical model in predicting no-collapse survival after MD. Moreover, the DLRC model exhibited superior predictive performance for overall no-collapse survival compared to the radiomics model. Based on the NRI values (Supplemental Table 1) the radiomics, DL, and DLRC models exhibited superior predictive capacities for 3-year no-collapse survival compared to the clinical model, with no notable discrepancies among these three models. For the 5-year NRI, no significant differences were detected among the clinical, radiomics, DL, and DLRC models. The calibration curves (Fig. 4) demonstrated favorable agreement between the predicted and actual survival outcomes for the DLRC model at both 3-year and 5-year no-collapse survival after multiple drilling, highlighting the model's robustness in survival prediction.

## **Clinical utility**

The radiomics, DL, and DLRC models demonstrated superior clinical utility compared to the clinical model for 3-year no-collapse survival. Notably, the clinical

Variables	Total (n = 212)	No-collapse (n = 148)	Collapse (n=64)	Statistic	Р
Follow-up, M (Q1, Q3)	29.55 (19.48, 40.83)	28.80 (19.43, 41.30)	29.90 (19.55, 39.95)	Z=-0.27	0.788
Gender, n (%)				$\chi^2 = 0.07$	0.799
Female	42 (19.81)	30 (20.27)	12 (18.75)		
Male	170 (80.19)	118 (79.73)	52 (81.25)		
Age, M (Q <sub>1</sub> , Q <sub>3</sub> )	42.00 (33.00, 49.00)	40.50 (33.00, 50.00)	42.50 (35.75, 47.25)	Z = -0.01	0.992
BMI, M (Q <sub>1</sub> , Q <sub>3</sub> )	23.71 (21.58, 26.04)	23.66 (21.22, 26.05)	23.79 (22.32, 25.86)	Z=-1.12	0.262
Symptom duration, $M (Q_1, Q_3)$	2.00 (1.00, 5.00)	2.00 (1.00, 4.25)	2.00 (1.00, 5.00)	Z=-0.36	0.716
Bisphosphonate, n (%)				$\chi^2 = 0.47$	0.493
No	115 (54.25)	78 (52.70)	37 (57.81)		
Yes	97 (45.75)	70 (47.30)	27 (42.19)		
Laterality, n (%)				$\chi^2 = 0.24$	0.888
Unilateral MD and contralateral THA	23 (10.85)	17 (11.49)	6 (9.38)		
Unilateral	57 (26.89)	39 (26.35)	18 (28.12)		
Bilateral	132 (62.26)	92 (62.16)	40 (62.50)		
Etiology, n (%)				$\chi^2 = 3.26$	0.196
Idiopathic	121 (57.08)	84 (56.76)	37 (57.81)		
Corticosteroid	58 (27.36)	37 (25.00)	21 (32.81)		
Alcohol-related	33 (15.57)	27 (18.24)	6 (9.38)		
Stage, n (%)				$\chi^2 = 5.87$	0.053
ARCO I	49 (23.11)	39 (26.35)	10 (15.62)		
ARCO II	108 (50.94)	77 (52.03)	31 (48.44)		
ARCO III A	55 (25.94)	32 (21.62)	23 (35.94)		
Bone marrow Edema, n (%)				$\chi^2 = 9.50$	0.002
No	83 (39.15)	68 (45.95)	15 (23.44)		
Yes	129 (60.85)	80 (54.05)	49 (76.56)		
Location, n (%)				$\chi^2 = 4.95$	0.084
Туре В	15 (7.08)	13 (8.78)	2 (3.12)		
Type C1	67 (31.60)	51 (34.46)	16 (25.00)		
Type C2	130 (61.32)	84 (56.76)	46 (71.88)		
Combined necrotic angle, Mean $\pm$ SD	$211.53 \pm 44.58$	204.23±44.67	228.40 ± 39.82	t=-3.73	<.001
Necrosis degree index, Mean±SD	114.91±46.53	107.59±45.18	131.84±45.52	t=-3.58	<.001

<b>Table 1</b> Baseline characteristics of all patients in collapse and no-collapse g	roup
---	------

BMI body mass index, ARCO Association Research Circulation Osseous classification

Z: Mann–Whitney test,  $\chi^2$ : Chi-square test

M: Median, Q1: 1st Quartile, Q3: 3st Quartile

model exhibited a negative net benefit at higher threshold risks (Fig. 5a). For risk thresholds below 80%, the radiomics, DL, and DLRC models achieved higher net benefits compared to the clinical model (Fig. 5b). Among these, the DLRC model showed the most consistent performance across varying thresholds and is presented as a nomogram (Fig. 1b). The Kaplan–Meier survival curves (Fig. 5c) highlight significant differences in no-collapse survival outcomes between high-risk and low-risk cohorts as predicted by the DLRC model. For the low-risk cohort, the 3-year no-collapse survival rate reached 93.96%, compared to 51.2% in the high-risk cohort. Similarly, the 5-year no-collapse survival rates were 85.7% for the low-risk group and 15.64% for the high-risk group.

## Discussion

In this retrospective study, a Deep Learning-Radiomics-Clinical (DLRC) predictive model was developed and validated using radiomics and deep learning signatures, along with three clinical characteristics that

**Table 2** Cox proportional hazards regression analysis of factors predicting no-collapse survival following multiple drilling

	Univariate analysis		Multivariate analysis		
Variables	Ρ	HR (95%CI)	Р	HR (95%CI)	
Bisphosphonate me	edicatior	1			
No		Ref			
Yes	0.602	1.14(0.69—1.89)			
Gender					
Female		Ref			
Male	0.545	1.21(0.65—2.28)			
Laterality					
Unilateral MD		Ref			
and					
	0 202	1 22/0 77 2 21)			
Dilatoral	0.502	1.33(0.77 - 2.31)			
Etiology	0.252	0.74(0.45—1.24)			
Idiopathic		Dof			
Corticoctoroid	0.151	Rei 1 47(0.97 - 2.5)			
	0.101	1.47(0.67 - 2.5)			
	0.167	0.57(0.24—1.52)			
		Dof		Dof	
	0.157	REI 0.7(0.42 1.15)	0.157		
	0.137	0.7(0.42-1.13)	0.137	2 12 (1 25 2 59)	
Ropo marrow odom	0.005	2.12(1.25-5.59)	0.037	2.12 (1.23—3.36)	
No.	la	Pof		Pof	
No	0.011	nei 0.12(1.10, 2.01)	0.010	12/110 201	
Location	0.011	2.15(1.19-5.01)	0.010	2.13 (1.19—3.01)	
Type B		Rof			
Type D	0.124	0.64(0.36 1.13)			
Type C1	0.124	1 73(1-2 98)			
Ago	0.001	1.75(1-2.90)			
< 12.00		Rof			
< 42.00 > 42.00	0.174	1 /2/0.86 2.37)			
242.00	0.174	1.42(0.00-2.37)			
23.71		Rof			
> 23.71	0.238	1 35(0 82-2 22)			
Symptom duration	0.250	1.55(0.02 2.22)			
< 2.00		Ref			
> 2.00	0.526	0.85(0.52-1.4)			
Combined necrotic	angle	0.00(0.02 1.4)			
< 210.50	angic	Ref		Ref	
> 210.50	0.001	2 42(1 44-4 07)	< 0.001	2 41 (1 43 4 07)	
Necrosis dearee inc	lex	2.12(1.17 7.07)	< 0.001	2.11(1.13 1.07)	
< 10967		Ref			
> 109.07	0.001	2 35(1 30 3 05)			
2109.07	0.001	2.33(1.39-3.95)			

BMI body mass index, ARCO Association Research Circulation Osseous classification, HR Hazard Ratio

independently predict femoral head no-collapse survival after multiple drilling. The DLRC model demonstrated good discrimination ability, with an AUC of 0.83 for 3-year no-collapse survival and 0.87 for 5-year no-collapse survival. The C-index was 0.78 (95% CI: 0.73— 0.84), outperforming other models. The model exhibited favorable calibration and demonstrated clinical utility. The calibration curve showed good concordance between predicted and observed outcomes across various probability ranges. Moreover, the DCA curve illustrated the model's clinical usefulness by demonstrating a positive net benefit across a range of threshold probabilities.

To the best of our knowledge, this is the first radiomics and deep learning model developed to predict nocollapse survival in patients with osteonecrosis after multiple drilling. Previous studies have mainly focused on clinical characteristics and radiographic parameters [10, 16, 27].Wei et al. developed a prognostic system for avascular necrosis of the femoral head after CD and identified seven independent risk factors [17]. Although the system achieved an AUC of 0.935, it is noteworthy that the prognostic system was constructed and evaluated within the same patient cohort and survival time was not accounted for [17]. Zhao et al. proposed a predictive model for the collapse of avascular necrosis of the femoral head after CD, but this study solely conducted internal validation and the C statistic was 0.82 [18]. We established a combined nomogram that incorporates the radiomics and DL signature and clinical factors for prognostic prediction, which is a novel approach compared to previous studies. The radiomics, DL, and DLRC models demonstrated superior performance compared to the clinical model, according to the IDI results. This finding aligns with previously reported prognostic models. Despite the absence of significant differences between DLRC and DL models, the DLRC model outperforms the radiomics model.

This study identified disease stage, bone marrow edema, and the combined necrotic angle as independent predictors of femoral head collapse, aligning with findings from previous research. Based on these factors, a clinical model was subsequently developed. Mont et al. [28] and Song et al. [13] observed a higher success rate in Ficat stage I and II compared to stage III after MD. Lieberman et al. [12] observed that in pre-collapse hips, there was a 19% conversion to total hip arthroplasty (THA) and a 31% radiographic progression, and in postcollapse hips, these rates were 30% and 49%, respectively. Within our study, the risk of collapse in ARCO stage IIIA (HR 2.12, 95% CI: 1.25-3.58) was higher compared to ARCO stage I, with no significant difference observed between ARCO stage I and ARCO stage II. However, controversy exists regarding the best classification system due to suboptimal interobserver and intraobserver



Fig. 2 Radiomics and DL feature selection using the LASSO-Cox regression model. a The partial likelihood deviance was plotted versus log (lambda) with radiomics features: (b) LASSO coefficient profiles of the 184 radiomics features: (c) The partial likelihood deviance was plotted versus log (lambda) with DL features; (d) LASSO coefficient profiles of the 512 DL features

model	C-index (95% Cl)	3-year AUC	5-year AUC			
Clinical	0.64(0.56-0.71)	0.65	0.77			
Radiomics	0.73(0.70-0.82)	0.80	0.87			
DL	0.76(0.70-0.82)	0.80	0.84			
DLRC	0.78(0.73-0.84)	0.83	0.87			

 Table 3
 Model performance of different models

C-index concordance index, AUC area under receiver operating characteristic (ROC) curve

agreement [20]. In the present study, the 2019 ARCO classification is adopted due to its simplicity, utilizing a 2 mm threshold to define both early and late collapse.

The extent of necrosis, size and location were identified as risk factors for collapse in previous researches. Various methods have been employed to quantify necrotic lesions, including the Steinberg classification, the Kerboul combined necrotic angle, the index of necrosis, the modified index of necrosis, and 3-D MRI measurements [22, 29]. However, consensus regarding the most reliable and valid method remains elusive [30]. Studies have suggested a higher success rate in cases with a smaller affected area (<15%) or a necrotic angle less than 200°. In this study, we chose to use the combined necrotic angle and the index of necrosis as alternatives. The combined necrotic angle (HR 2.41, 95% CI:1.43-4.07, p<0.05) was identified as an independent risk factor, whereas the index did not prove to be an independent risk factor in the multivariate Cox model (p > 0.05). A plausible reason for this discrepancy may be attributed to the insufficient precision of our measurement method. Research has indicated the significance of the necrotic location in relation to the weight-bearing zone, with the JIC proving to be a reliable method. Specifically, JIC C1 and JIC C2 have been associated with higher rates of collapse. However, in the current study, necrosis location did not emerge as an independent risk factor (p > 0.05). A possible explanation for this may include inadequacies in sample size.

Clinical factors such as age, BMI, etiology, and laterality were not identified as risk factors for collapse in this study. However, Wei et al. identified age, male gender,



Fig. 3 Model evaluation of different models. a Receiver operating characteristic curves for predicting 3-year no-collapse survival after multiple drilling; (b) Receiver operating characteristic curves for predicting 5-year no-collapse survival after multiple drilling



Fig. 4 Model validation of different models. a Calibration curves for 3-year no-collapse survival after multiple drilling; (b) Calibration curves for 5-year no-collapse survival after multiple drilling

etiology, and a prolonged disease duration as prognostic factors influencing the outcome following CD [17]. Disparities in findings may stem from various factors, including differences in sample sizes used for Cox regression analysis and variations in patient characteristics such as alcohol or corticosteroid intake. Moreover, discrepancies in the accuracy of patients' recollection of symptom duration could also contribute to differing results. Some studies have suggested that bisphosphonates may play a role in preventing femoral head collapse [31, 32]. Agarwala et al. found that alendronate administration resulted in satisfactory outcomes for most patients with earlystage avascular necrosis. The results revealed that 364 hips (92.2%) achieved satisfactory clinical outcomes,



**Fig. 5** Clinical utility of different models. **a** The decision curves indicated that the radiomics, DL and DLRC model showed better clinical utility than the clinical model for 3-year no-collapse survival; (**b**) The decision curves indicated that the radiomics, DL and DLRC model showed better clinical utility than the clinical model for 5-year no-collapse survival when the risk threshold was less than 80%. At higher risk thresholds, the curves did not intersect with the X-axis, which may be due to factors such as the relatively small sample size, the lower 5-year survival rate, and the reduced number of high-risk individuals at 5 years.; (**c**) No-collapse survival comparison between patients classified by the DLRC nomogram into high-risk and low-risk cohorts

obviating the need for surgical intervention [31]. Similarly, Kang et al. reported successful pain relief and delayed progression of necrosis with a combination of multiple drilling and systematic alendronate treatment [32]. However, our study did not find bisphosphonates to have a significant impact on prognosis after MD. This discrepancy may be attributed to our focus on radiological progression as an outcome measure, whereas prior research primarily relied on clinical assessments.

There are several strengths to this study. As a rapidly evolving discipline in medical imaging, radiomics and deep learning facilitates high-throughput analysis of medical image data [33]. Both radiomics and deep learning play pivotal roles in aiding diagnosis, prognosis, and prediction, with significant advancements observed in oncology studies [34]. Yet, their application in orthopedic research has been rarely reported. In the present study, various radiomics and deep learning features were extracted, and we employed several dimensionality reduction techniques to select the top 14 radiomics features and 21 DL features. Furthermore, multivariate Cox analysis demonstrated that both the radiomics and deep learning signatures were identified as independent risk factors, exhibiting a higher Hazard Ratio (Supplemental Table 2). Radiomics and deep learning methodologies may offer superior approaches for discerning the extent of necrotic regions and identifying lesion heterogeneity. DL directly generated and extracted optimal features from raw data, whereas radiomics relies on predefined features [35]. In this study, no significant differences were observed between the radiomics and DL models, which contrasts with findings from other studies. This distinction may be attributed to the specific DL architecture utilized in this investigation.

This model can stratify patients into high-risk and lowrisk groups with significant statistical differences, thus enabling prediction of no-collapse survival. This study confirms the applicability of radiomics and deep learning in the field of orthopedics, thereby aiding in treatment and prognosis. Personalized assessment of the collapse risk after osteonecrosis of the femoral head (ONFH) can guide treatment decisions, enabling high-risk patients to consider adjunctive therapies such as cell therapy.

This study has several limitations. First, it is a retrospective analysis with a relatively small patient cohort from a single center, which inherently limits the diversity of the patient population. The small sample size reduces the statistical power of the analysis, increasing the risk of Type II errors and hindering the detection of subtle or less common risk factors. Additionally, the limited sample size increases the likelihood of overfitting. This overfitting may result in inflated performance during internal validation but significantly reduced predictive accuracy on independent datasets. The single-center nature of the study further restricts generalizability, as the patient cohort may reflect regional clinical practices and demographic characteristics that are not representative of broader populations. The lack of external validation exacerbates these issues, making it challenging to robustly evaluate the model's applicability across diverse patient populations. Second, the postoperative followup data are incomplete, lacking detailed information on parameters such as activity levels, scoring metrics, and continued exposure to preoperative etiological factors. These omissions limit the depth of the analysis and may obscure important associations. Third, there are limitations related to the imaging data. Radiographic parameters were measured using sagittal-plane MRI rather than coronal-plane MRI, which differs from approaches commonly used in prior research. Moreover, radiomics analysis was conducted using X-ray imaging because MRI images in DICOM format were unavailable for some patients from external institutions. Finally, the ResNet18 architecture employed in this study may have inherent limitations, underscoring the need for refinement in future deep learning frameworks to improve performance. Future research should address these limitations by incorporating multi-center data and external validation cohorts to enhance the model's generalizability and robustness. In addition, it is essential to include more comprehensive follow-up data, such as postoperative activity levels, scoring metrics, and continued exposure to preoperative etiological factors, to provide deeper insights into patient outcomes. Further efforts should also focus on integrating advanced imaging modalities, such as MRI, to extract richer radiomic and deep learning features. Finally, exploring more sophisticated deep learning architectures could optimize predictive accuracy and improve the model's overall efficiency.

## Conclusions

This study introduces a novel approach that integrates radiomics and deep learning techniques to predict femoral head collapse following multiple drilling for osteonecrosis. The DLRC model, which combines radiomics and deep learning features with clinical factors, demonstrated superior predictive performance compared to traditional clinical model. These findings highlight the potential of personalized risk assessment in guiding treatment decisions and improving outcomes for patients with osteonecrosis of the femoral head.

#### Abbreviations

 THA
 Total hip arthroplasty

 ONFH
 Osteonecrosis of the femoral head

 CD
 Core decompression

 MD
 Multiple drilling

 DL
 Deep Learning

 DLRC
 Deep Learning-Radiomics-Clinical model

#### Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-025-02859-2.

Supplementary Material 1.

## Acknowledgements

Not applicable.

#### Authors' contributions

All authors contributed to the study conception and design. Material preparation and data collection were performed by Fan Liu, De-bao Zhang, Shi-huan Cheng, and analysis were performed by Fan Liu and Gui-shan Gu. The first draft of the manuscript was written by Fan Liu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

#### Funding

This research received no external funding.

#### Data availability

Availability of data and materials: Due to privacy considerations, the datasets generated during the current study are not publicly available. However, data may be available from the corresponding author (email: gugs@jlu.edu.cn) upon reasonable request.

#### Declarations

#### Ethics approval and consent to participate

This retrospective study was approved by the Medical Ethics Committee of the first hospital of Jilin University (registration number 2020–607, approved on 15 Dec 2020), and written informed consent was waived. Because the medical records used in this study were obtained from past clinical diagnosis and treatment, this clinical study did not directly involve the subjects, and the results were not used for subject diagnosis. Therefore, it will not have any

adverse effects on the subjects. The privacy and personal identity information of the subjects are protected. Furthermore, the study adheres to the Helsinki Declaration.

## Consent for publication

Not applicable.

## **Competing interests**

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Orthopedics, the First Hospital of Jilin University, Changchun, Jilin Province 130021, China. <sup>2</sup>Department of Rehabilitation, the First Hospital of Jilin University, Changchun, Jilin Province 130021, China.

Received: 7 April 2024 Accepted: 6 January 2025 Published online: 15 January 2025

#### References

- Grecula MJ. CORR Insights<sup>®</sup>: Which Classification System Is Most Useful for Classifying Osteonecrosis of the Femoral Head? Clin Orthop Relat Res. 2018;476(6):1250–2. https://doi.org/10.1097/01.blo.0000533640.75452.45.
- Atilla B, Bakircioglu S, Shope AJ, Parvizi J. Joint-preserving procedures for osteonecrosis of the femoral head. EFORT Open Rev. 2019;4(12):647–58. https://doi.org/10.1302/2058-5241.4.180073.
- Zhao DW, Yu M, Hu K, Wang W, Yang L, Wang BJ, Gao XH, Guo YM, Xu YQ, Wei YS, et al. Prevalence of Nontraumatic Osteonecrosis of the Femoral Head and its Associated Risk Factors in the Chinese Population: Results from a Nationally Representative Survey. Chin Med J (Engl). 2015;128(21):2843–50. https://doi.org/10.4103/0366-6999.168017.
- Gosling-Gardeniers AC, Rijnen WHC, Gardeniers JWM. The Prevalence of Osteonecrosis in Different Parts of the World. In: Koo K-H, Mont MA, Jones LC, editors. Osteonecrosis. Heidelberg: Springer; 2014. p. 35–7.
- Villa JC, Husain S, van der List JP, Gianakos A, Lane JM. Treatment of Pre-Collapse Stages of Osteonecrosis of the Femoral Head: a Systematic Review of Randomized Control Trials. HSS J. 2016;12(3):261–71. https:// doi.org/10.1007/s11420-016-9505-9.
- Sodhi N, Acuna A, Etcheson J, Mohamed N, Davila I, Ehiorobo JO, Jones LC, Delanois RE, Mont MA. Management of osteonecrosis of the femoral head. Bone Joint J. 2020;102-B(7\_Supple\_B):122–8. https://doi.org/10. 1302/0301-620X.102B7.BJJ-2019-1611.R1.
- Huang ZQ, Fu FY, Li WL, Tan B, He HJ, Liu WG, Chen WH. Current Treatment Modalities for Osteonecrosis of Femoral Head in Mainland China: A Cross-Sectional Study. Orthop Surg. 2020;12(6):1776–83. https://doi.org/ 10.1111/os.12810.
- Johnson AJ, Mont MA, Tsao AK, Jones LC. Treatment of femoral head osteonecrosis in the United States: 16-year analysis of the Nationwide Inpatient Sample. Clin Orthop Relat Res. 2014;472(2):617–23. https://doi. org/10.1007/s11999-013-3220-3.
- Mont MA, Carbone JJ, Fairbank AC. Core decompression versus nonoperative management for osteonecrosis of the hip. Clin Orthop Relat Res. 1996;324:169–78. https://doi.org/10.1097/00003086-199603000-00020.
- Zalavras CG, Lieberman JR. Osteonecrosis of the Femoral Head: Evaluation and Treatment. J Am Acad Orthop Sur. 2014;22(7):455–64. https:// doi.org/10.5435/Jaaos-22-07-455.
- Mont MA, Salem HS, Piuzzi NS, Goodman SB, Jones LC. Nontraumatic Osteonecrosis of the Femoral Head: Where Do We Stand Today?: A 5-Year Update. J Bone Joint Surg Am. 2020;102(12):1084–99. https://doi.org/10. 2106/jbjs.19.01271.
- Lieberman JR, Engstrom SM, Meneghini RM, SooHoo NF. Which Factors Influence Preservation of the Osteonecrotic Femoral Head? Clin Orthop Relat R. 2012;470(2):525–34. https://doi.org/10.1007/s11999-011-2050-4.
- Song WS, Yoo JJ, Kim YM, Kim HJ. Results of multiple drilling compared with those of conventional methods of core decompression. Clin Orthop Relat Res. 2007;454:139–46. https://doi.org/10.1097/01.blo.0000229342. 96103.73.
- 14. Al Omran A. Multiple drilling compared with standard core decompression for avascular necrosis of the femoral head in sickle cell disease

patients. Arch Orthop Trauma Surg. 2013;133(5):609–13. https://doi.org/10.1007/s00402-013-1714-9.

- Liu Z, Yang X, Li Y, Zeng WN, Zhao E, Zhou Z. Multiple drilling is not effective in reducing the rate of conversion to Total hip Arthroplasty in earlystage nontraumatic osteonecrosis of the femoral head: a case-control comparative study with a natural course. BMC Musculoskelet Disord. 2021;22(1):535. https://doi.org/10.1186/s12891-021-04418-y.
- Migliorini F, Maffulli N, Baroncini A, Eschweiler J, Tingart M, Betsch M. Prognostic factors in the management of osteonecrosis of the femoral head: A systematic review. Surgeon. 2023;21(2):85–98. https://doi.org/10. 1016/j.surge.2021.12.004.
- Wei C, Yang M, Chu K, Huo J, Chen X, Liu B, Li H. The indications for core decompression surgery in patients with ARCO stage I-II osteonecrosis of the femoral head: a new, comprehensive prediction system. BMC Musculoskelet Disord. 2023;24(1):242. https://doi.org/10.1186/ s12891-023-06321-0.
- Zhao EZ, Liu ZH, Zeng WN, Ding ZC, Luo ZY, Zhou ZK. Nomogram to predict collapse-free survival after core decompression of nontraumatic osteonecrosis of the femoral head. J Orthop Surg Res. 2021;16(1):519. https://doi.org/10.1186/s13018-021-02664-3.
- Vial A, Stirling D, Field M, Ros M, Ritz C, Carolan M, Holloway L, Miller AA. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. Transl Cancer Res. 2018;7(3):803–16. https://doi.org/10.21037/tcr.2018.05.02.
- Yoon BH, Mont MA, Koo KH, Chen CH, Cheng EY, Cui Q, Drescher W, Gangji V, Goodman SB, Ha YC, et al. The 2019 Revised Version of Association Research Circulation Osseous Staging System of Osteonecrosis of the Femoral Head. J Arthroplasty. 2020;35(4):933–40. https://doi.org/10. 1016/j.arth.2019.11.029.
- Sugano N, Atsumi T, Ohzono K, Kubo T, Hotokebuchi T, Takaoka K. The 2001 revised criteria for diagnosis, classification, and staging of idiopathic osteonecrosis of the femoral head. J Orthop Sci. 2002;7(5):601–5. https:// doi.org/10.1007/s007760200108.
- Cherian SF, Laorr A, Saleh KJ, Kuskowski MA, Bailey RF, Cheng EY. Quantifying the extent of femoral head involvement in osteonecrosis. J Bone Joint Surg Am. 2003;85(2):309–15. https://doi.org/10.2106/00004623-20030 2000-00019.
- 23. Fleisch H. Bisphosphonates in bone disease: from the laboratory to the patient. 4th ed. San Diego: Academic Press; 2000.
- Chughtai M, Piuzzi NS, Khlopas A, Jones LC, Goodman SB, Mont MA. An evidence-based guide to the treatment of osteonecrosis of the femoral head. Bone Joint J. 2017;99-B(10):1267–79. https://doi.org/10.1302/0301-620X.99B10.BJJ-2017-0233.R2.
- Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. Userguided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage. 2006;31(3):1116– 28. https://doi.org/10.1016/j.neuroimage.2006.01.015.
- van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts H. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017;77(21):e104–7. https://doi.org/10.1158/0008-5472.CAN-17-0339.
- Banerjee S, Kapadia BH, Jauregui JJ, Cherian JJ, Mont MA. Natural history of osteonecrosis. In: Koo KH, Mont MA, Jones LC, editors. Osteonecrosis. Heidelberg: Springer; 2014. p. 161–4.
- Mont MA, Ragland PS, Etienne G. Core decompression of the femoral head for osteonecrosis using percutaneous multiple small-diameter drilling. Clin Orthop Relat R. 2004;429:131–8. https://doi.org/10.1097/01.blo. 0000150128.57777.8e.
- Steinberg ME, Oh SC, Khoury V, Udupa JK, Steinberg DR. Lesion size measurement in femoral head necrosis. Int Orthop. 2018;42(7):1585–91. https://doi.org/10.1007/s00264-018-3912-0.
- Hines JT, Jo WL, Cui Q, Mont MA, Koo KH, Cheng EY, Goodman SB, Ha YC, Hernigou P, Jones LC, et al. Osteonecrosis of the Femoral Head: an Updated Review of ARCO on Pathogenesis, Staging and Treatment. J Korean Med Sci. 2021;36(24):e177. https://doi.org/10.3346/jkms.2021.36. e177.
- Agarwala S, Shah S, Joshi VR. The use of alendronate in the treatment of avascular necrosis of the femoral head: follow-up to eight years. J Bone Joint Surg Br. 2009;91(8):1013–8. https://doi.org/10.1302/0301-620x.91b8. 21518.

- 32. Kang P, Pei F, Shen B, Zhou Z, Yang J. Are the results of multiple drilling and alendronate for osteonecrosis of the femoral head better than those of multiple drilling? A pilot study Joint Bone Spine. 2012;79(1):67–72. https://doi.org/10.1016/j.jbspin.2011.02.020.
- Avanzo M, Wei L, Stancanello J, Vallieres M, Rao A, Morin O, Mattonen SA, El Naga I. Machine and deep learning methods for radiomics. Med Phys. 2020;47(5):e185–202. https://doi.org/10.1002/mp.13678.
- Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue R, Even AJG, Jochems A, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749–62. https://doi.org/10.1038/nrclinonc.2017. 141.
- 35. Scalco E, Rizzo G, Mastropietro A. The stability of oncologic MRI radiomic features and the potential role of deep learning: a review. Phys Med Biol. 2022;67(9):09TR3. https://doi.org/10.1088/1361-6560/ac60b9.

## **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.