Cui *et al. Journal of Orthopaedic Surgery and Research* (2025) 20:503
https://doi.org/10.1186/s13018-025-05901-1

Journal of Orthopaedic Surgery and Research

RESEARCH Open Access

Noninvasive prediction of failure of the conservative treatment in lateral epicondylitis by clinicoradiological features and elbow MRI radiomics based on interpretable machine learning: a multicenter cohort study

Jianing Cui[1], Ping Wang[1], Xiaodong Zhang[2], Ping Zhang[3], Yuming Yin[4] and Rongjie Bai[1*]

Abstract

Objectives To develop and validate an interpretable machine learning model based on clinicoradiological features and radiomic features based on magnetic resonance imaging (MRI) to predict the failure of conservative treatment in lateral epicondylitis (LE).

Methods This retrospective study included 420 patients with LE from three hospitals, divided into a training cohort ($n=245$), an internal validation cohort ($n=115$), and an external validation cohort ($n=60$). Patients were categorized into conservative treatment failure ($n=133$) and conservative treatment success ($n=287$) groups based on the outcome of conservative treatment. We developed two predictive models: one utilizing clinicoradiological features, and another integrating clinicoradiological and radiomic features. Seven machine learning algorithms were evaluated to determine the optimal model for predicting the failure of conservative treatment. Model performance was assessed using ROC, and model interpretability was examined using SHapley Additive exPlanations (SHAP).

Results The LightGBM algorithm was selected as the optimal model because of its superior performance. The combined model demonstrated enhanced predictive accuracy with an area under the ROC curve (AUC) of 0.96 (95% CI: 0.91, 0.99) in the external validation cohort. SHAP analysis identified the radiological feature "CET coronal tear size" and the radiomic feature "AX_log-sigma-1-0-mm-3D_glszm_SmallAreaEmphasis" as key predictors of conservative treatment failure.

Conclusions We developed and validated an interpretable LightGBM machine learning model that integrates clinicoradiological and radiomic features to predict the failure of conservative treatment in LE. The model demonstrates high predictive accuracy and offers valuable insights into key prognostic factors.

Keywords Lateral epicondylitis, Machine learning, Magnetic resonance imaging, Radiomics, SHapley additive explanation

*Correspondence:
Rongjie Bai
bairongjie@126.com

Full list of author information is available at the end of the article

Introduction

Lateral epicondylitis (LE) is caused by a chronic degenerative lesion at the origin of the common extensor tendon (CET) of the elbow and is one of the most common conditions leading to lateral elbow pain and limitation of upper limb function [1]. Conservative treatment has a high success rate in the early stages of LE, but some patients still require surgical management after long-term conservative treatment fails [2]. There is a lack of standardization of treatment decisions for LE, which relies heavily on the experience of the clinician [2]. Without timely and appropriate treatment, patients are at higher risk of progressive tendon degeneration and potential rupture. The longer the symptoms persist, the lower the likelihood of successful treatment and recovery [3, 4]. Therefore, early and accurate prediction of failure of conservative treatment and early surgical intervention is essential to improve the prognosis of patients with LE.

Current research has recognized the importance of accurate early prognostic prediction and has found that persistent pain, CET abnormalities, and muscle edema are associated with failure of conservative treatment [4–7]. However, the predictive performance of traditional models still needs to be improved. In addition, although magnetic resonance imaging (MRI) is recognized as a high-precision detection method [8]. Currently, determining the extent of elbow injury primarily relies on the individual radiologist's work experience and interpretation skills. It is more difficult to capture subtle changes in the tissue features shown on MRI that may be correlated with treatment outcomes.

At present, medical image data has grown explosively, in which there is a large amount of interference and useless information, which invariably brings challenges to the professional ability of doctors and restricts the accuracy of judgment. Therefore, it is particularly important to conduct scientific, accurate, and effective information mining. In recent years, radiomics is an advanced quantitative image analysis method that utilizes computer algorithms to extract a large number of quantitative features from medical images to detect information not visible to human observers. Radiomics has been effective in improving the accuracy of clinical diagnosis and prognosis [9–11], and has been progressively applied to musculoskeletal radiology in recent years, demonstrating high performance and efficiency in the diagnosis of structural tears in the anterior cruciate ligament, meniscus, and supraspinatus tendon [12–14]. However, radiomics studies on lateral elbow tendon and ligament injuries are relatively few and short-lived. Furthermore, the process developed in machine learning algorithms is opaque and has been likened to a "black box". That is, researchers cannot know how the algorithm works. This is generally regarded as unfortunate because it is not conducive to clinical application. The Shapley Additive exPlanations (SHAP) is an emerging interpretability method that uses Shapley values to measure how much each feature contributes to the model's prediction [15–17]. The SHAP analysis provides case-specific (local) and population-level (global) explanations for "black-box" problems while maintaining a high level of consistency and diagnostic accuracy compared to other explainable AI methods [18].

Therefore, we aimed to construct a machine learning model combining MRI radiomics and clinicoradiological features to predict the failure of conservative treatments and validate the generality of the model through external validation, followed by further interpretation and visualization of the model using the SHAP method.

Materials and methods

The Institutional Review Board (IRB) approval was obtained for this study, and informed consent was waived owing to the study's retrospective nature.

Patient selection

This is a multicenter study in includes 3 hospitals. A total of 420 consecutive LE patients between January 2017 and January 2023 were retrospectively recruited in this study. The inclusion criteria were as follows: (1) The patient presented with typical symptoms of lateral elbow pain and was diagnosed with lateral epicondylitis by physical examination; (2) the patient underwent an MRI scan of the elbow. Exclusion criteria were as follows: (1) systemic comorbidities; (2) history of elbow surgery or arthroscopic procedure; (3) insufficient clinical information; (4) loss of follow-up; (5) poor MRI quality. Figure 1 depicts the study workflow. Initially, all patients with LE were treated conservatively, including rest, nonsteroidal anti-inflammatory drugs (NSAIDs), force-resistant stent protection, and physical therapy. Failure of conservative treatment is defined as no improvement in pain or dysfunction after 6 to 12 months of nonsurgical treatment. Patients were categorized by experienced orthopedic surgeons specializing in upper limb corrections into the conservative treatment success group and the conservative treatment failure group based on whether or not they received surgical treatment. The same classification criteria were used in all three hospitals. Patients from our hospital were randomly assigned to the training set, and the internal validation was set at a ratio of 7:3, while patients from the other two hospitals were used as the external validation set. The clinical variables included age, affected side, and duration of pain.

MR imaging examinations and analysis

MRI of the elbows of patients at the three hospitals was performed using a 3.0T MRI scanner (Philips Medical Systems) equipped with a dedicated extremity coil.

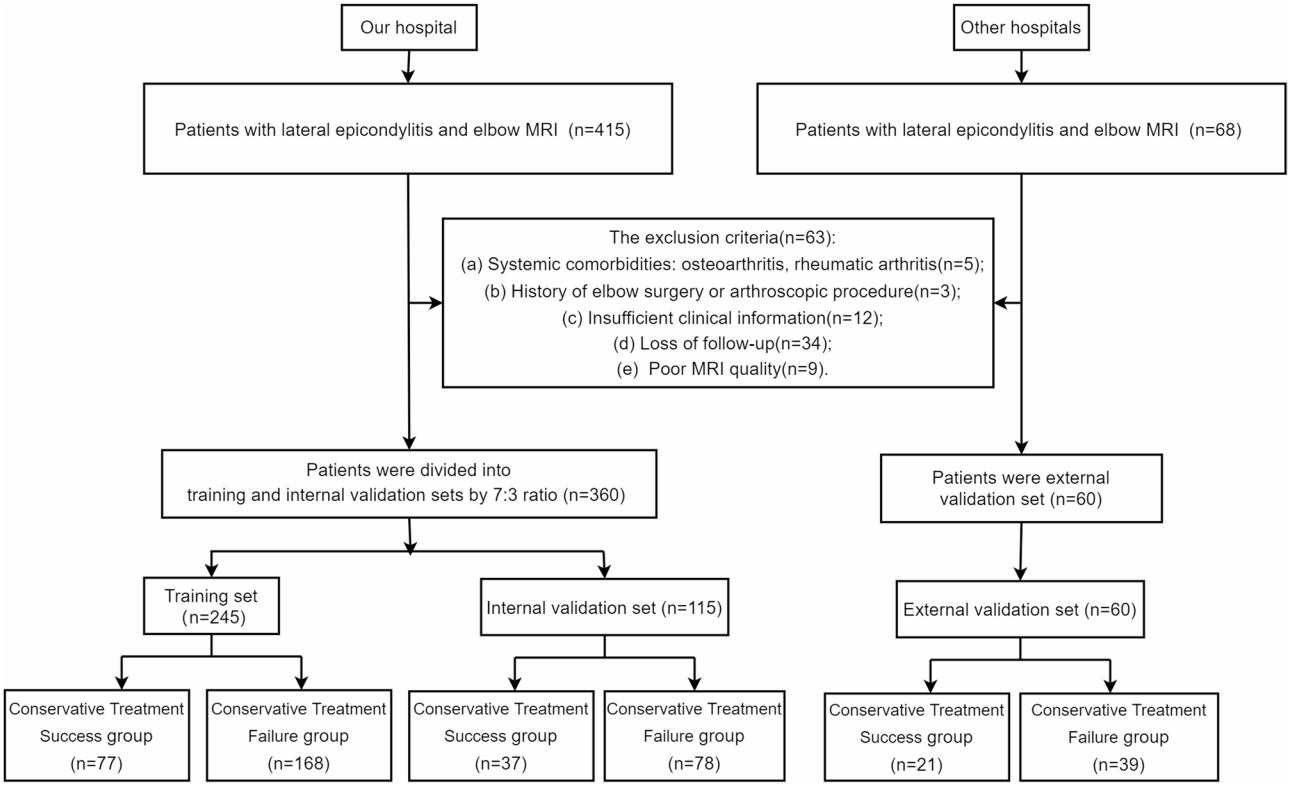

Fig. 1 Flowchart depicts the patient's enrollment process in the study. MRI: magnetic resonance imaging

The MRI protocol of the elbow included T1-weighted image (T1WI) sequences in coronal and proton density-weighted imaging with fat suppression (PD-FS) sequences in axial, coronal, and sagittal planes. The MRI protocols of the three hospitals are detailed in Table S1.

An experienced musculoskeletal radiologist reviewed the elbow MRI blinded to clinical data. CET abnormalities are categorized as tendinosis and tendon tears. Tendinosis was characterized by increased signal intensity within the tendon, lower than the fluid signal, and/or an increased tendon thickness. Determination of the presence of a tear based on fluid signal intensity in the tendon region [5]. The maximum extent of the CET tear was measured on coronal and axial PD-FS images. Lateral collateral ligament (LCL) complex insufficiency was classified into three grades: Grade 0 was normal, Grade 1 was a partial tear, peri-ligamentous edema, thickening or thinning, and Grade 2 was a complete tear [5]. Joint effusions, muscle edema, subchondral cysts, subchondral bone marrow edema, and cartilage defects, were classified as present or absent.

Radiomics workflow

The comprehensive radiomics workflow encompassed four stages: MRI preparation and Regions of interest (ROI) Segmentation, Feature Extraction and subsequent Selection, Model Construction, and Model evaluation (Fig. 2).

Scar segmentation

ROI were manually outlined on coronal and axial PD-FS images by a musculoskeletal radiologist using ITK SNAP software (version 3.6.0, www.itksnap.org). The ROI region is the portion of the CET and LCL complex originating at the lateral epicondyle of the humerus and contains 4–5 consecutive MRI slices (Fig. 2). The segmentation protocol is described in the Supplementary material.

Feature extraction and feature selection

A total of 2260 features were initially extracted from the coronal and axial PD-FS images, covering shape, first-order statistics, and texture matrices. After conducting robustness testing, which included Pearson correlation and intraclass correlation coefficient (ICC) for absolute agreement, and additional selection procedures, the final radiomics features were retained for combined model construction [19]. Detailed procedures are available in the Supplementary Material.

Model construction, evaluation, and interpretability

We constructed two models to evaluate the incremental value of MRI radiomics in predicting the failure of

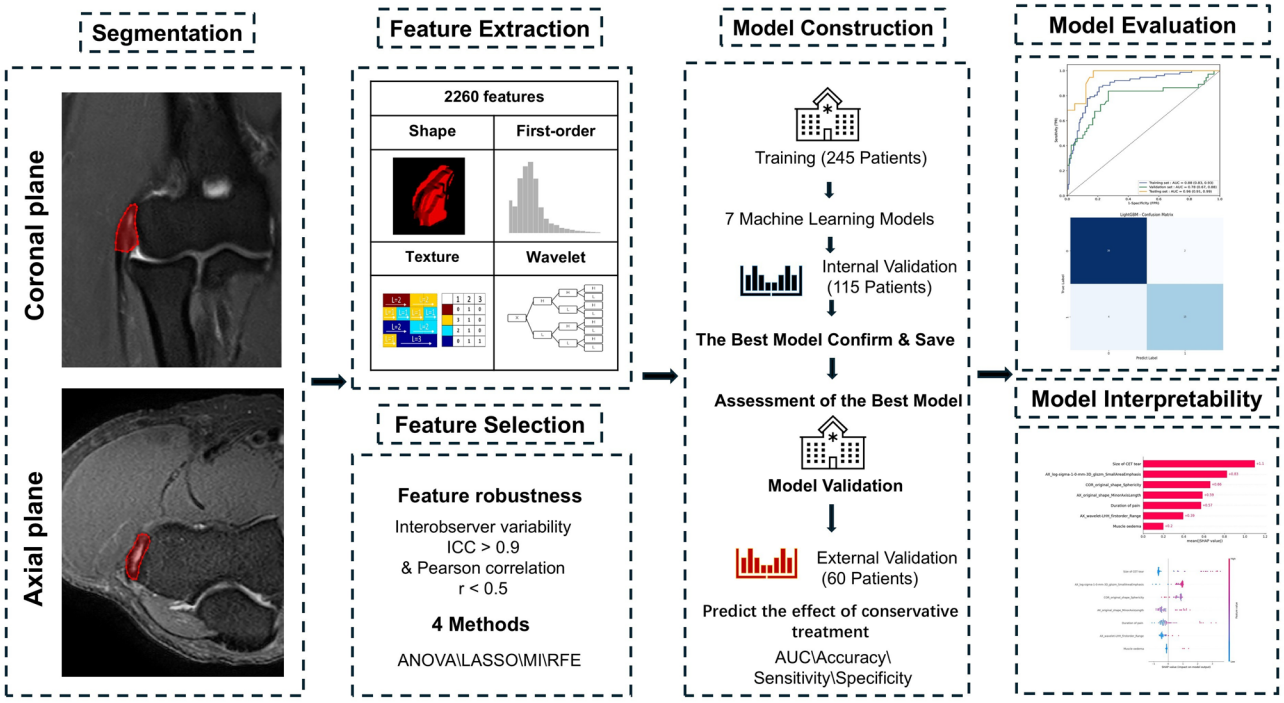

Fig. 2 Workflow of the radiomic analysis in this study. ICC: intraclass correlation coefficient for absolute agreement; ANOVA: analysis of variance; LASSO: least absolute shrinkage and selection operator; MI: mutual information; RFE: recursive feature elimination; ROC: receiver operating characteristic curve; AUC: area under the ROC curve; SHAP: Shapley additive explanations

conservative treatment of LE. Clinicoradiological features comprise conventional MRI findings (e.g., CET tear size and muscle edema) and traditionally clinically collected parameters (e.g., age, sex, and pain duration). In contrast, radiomic features are quantitative features extracted from MRI through computational algorithms that capture imaging characteristics beyond human visual assessment (e.g., first-order statistics and texture matrices). The clinicoradiological model was developed based on univariable and multivariable logistic regression analyses of clinicoradiological characteristics. The combined model was based on univariable and multivariable logistic regression analyses of clinicoradiological model features and selected radiomics features. These features are used in seven machine learning algorithms to develop machine learning models, including support vector machine (SVM), logistic regression (LR), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), XGBoost, and Light Gradient Boosting Machine (LightGBM). Each machine learning model's predictive performance was evaluated by comparing the accuracy, sensitivity, and specificity in predicting the failure of conservative treatments.

SHAP analysis

After identifying the best predictive model, we turned our focus to understanding the contribution of each variable to the prediction. SHAP quantifies the contribution of each feature to a model's prediction, thereby improving global and local interpretability [20, 21]. Specifically, global interpretation aims to assign meaning to each model feature. This process involves calculating the SHAP value for each feature (higher magnitude indicates greater influence on model results), plotting a SHAP beeswarm plot (showing the distribution of SHAP values across individual patients), and plotting a SHAP bar chart (showing the features' mean absolute SHAP values across patients, ranked). The purpose of local interpretation is to emphasize the impact of features on determining the outcomes of the model in individual patients by constructing SHAP force plots that show the features that contribute most to the predicted probability of an individual patient in the model.

Statistical analysis

Continuous variables were expressed as median and interquartile range (IQR), while categorical variables were expressed as counts and percentages. Differences between the groups were analyzed using the Mann–Whitney U test, chi-square test, or Kruskal-Wallis H-test as appropriate. Univariable analysis of candidate variables was performed using univariable logistic regression. Parameters that demonstrated statistical significance ($P < 0.05$) in the univariable analysis were then examined using a backward stepwise multivariable logistic regression analysis. DeLong's test was used to compare the

AUC of the combined model and the clinicoradiological model. A *p*-value less than 0.05 indicated statistical significance. All figures presented in this work were created using R software (Version 4.2.1, available on the R Project website). For statistical analysis and machine learning analysis, the R packages "ggstatsplot" (version 0.9.1.9000), "ggcor", the open-source Python packages SciPy, scikit-learn, glmnet, caret, and SHAP were used.

Result

Patient population

A total of 483 patients with LE underwent elbow MRI in 3 hospitals, of whom 63 were excluded. Therefore, 420 patients were included in the final study, including 245 patients in the training set, 115 in the internal validation set, and 60 in the external validation set (Fig. 1). There were no significant differences in clinical and MRI characteristics between the training, internal validation, and external validation sets (all $P > 0.05$) (Table 1). The conservative treatment failure group had fewer men than the conservative treatment success group ($P = 0.031$). Pain duration was longer in the conservative treatment failure group compared to the conservative treatment success group (both $P < 0.001$). CET tears were significantly larger in the conservative treatment failure group (both $P < 0.001$). The conservative treatment failure group had significantly higher CET and LCL complex abnormality grades (both $P < 0.001$). In addition, muscle oedema, cartilaginous defect, and subchondral cyst were more common in the conservative treatment failure group ($P < 0.001$, $P = 0.009$, and $P < 0.001$) (Table S2).

Radiomics feature extraction, selection, and development of models

A total of 473 radiomics features demonstrated robustness against resolution resampling (ICC > 0.9). The remaining radiomics features were reduced by four feature selection methods, and finally, we selected 12 radiomics features using the RFE method.

Clinicoradiological model- Univariable logistic regression analysis revealed significant variables. After multivariable analysis, only duration of pain[OR: 1.146 (1.082–1.213), $P < 0.001$], CET coronal tear size [OR: 1.760 (1.424–2.175), $P < 0.001$], Grade 1 of LCL complex abnormality [OR: 2.658 (1.272–5.553), $P = 0.009$], and presence of muscle oedema [OR: 9.840 (2.766–35.013), $P = 0.002$] were independent predictors of conservative treatment failure (Table 2).

Combined model- Univariate analysis was based on the 4 features of the clinicoradiological model and selected 12 radiomics features using the RFE method. In univariable logistic analysis, 13 characteristics (4 clinicoradiological and 9 radiomics) had a *P* value of less than 0.05 and were therefore included in the multivariable logistic analysis. The final analysis identified 7 factors associated with the prognosis of conservative treatment, including 3 clinicoradiological features and 4 radiomics features (all $P < 0.05$) (Table S3).

Model evaluation and model interpretability

Among the 7 different machine learning models, we selected LightGBM as our algorithm for constructing the combined model owing to its superior performance in

Table 1 Clinical and MRI characteristics of patients in the training and validation sets

Characteristics	Training ($n = 245$)	Internal ($n = 115$)	External ($n = 60$)	*P* value
Sex, *n* (male, %)	100(41)	53(46)	25(42)	0.636
Age (years)	47(41,53)	47(41,52)	48.50(41,54.75)	0.697
Laterality, *n* (left, %)	73(30)	37(32)	18(30)	0.897
Duration of pain (months)	10(8,12)	10 (8,13)	10 (7.25,12.75)	0.708
CET abnormality, *n* (%)				0.570
Tendinosis	171(70)	85(74)	39(65)	
Tear	74(30)	30(26)	21(35)	
Size of CET tear (mm)				
Coronal plane	0(0,2.15)	0 (0,1)	0 (0,3.37)	0.360
Axial plane	0 (0,2.87)	0 (0,1.93)	0 (0,4.19)	0.212
LCL complex abnormality, *n* (%)				0.487
Grade 0	136(56)	104(42)	5(2)	
Grade 1	64(56)	50(44)	1(1)	
Grade 2	34(57)	23(38)	3(5)	
Muscle oedema, *n* (%)	24(10)	12(10)	6(10)	0.982
Joint effusion, *n* (%)	226(92)	107(93)	57(95)	0.756
Cartilaginous defect, *n* (%)	76(31)	30(26)	16(27)	0.572
Subchondral cyst, *n* (%)	96(39)	38(33)	20(33)	0.448
Subchondral marrow oedema, *n* (%)	19(8)	9(8)	1(2)	0.224

Data are presented as number (percentage) for dichotomous variables and median (interquartile range) for continuous variables. MRI: Magnetic resonance imaging; LCL: lateral collateral ligament; CET: common extensor tendon. *Statistically significant difference

Table 2 Clinical and MRI features associated with failed Conservative treatment for lateral epicondylitis

Characteristics	Univariable analysis			Multivariable analysis		
	OR	95%CI	*P* value	OR	95%CI	*P* value
Sex, *n* (male, %)	0.548	0.310–0.969	0.039*	-	-	-
Age (years)	1.006	0.975–1.038	0.720			
Laterality, *n* (left, %)	0.761	0.416–1.393	0.377			
Duration of pain (months)	1.169	1.109–1.232	<0.001*	1.156	1.093–1.223	<0.001*
CET abnormality, *n* (%)	5.589	3.082–10.136	<0.001*	-	-	-
Size of CET tear (mm)						
Coronal plane	1.757	1.476–2.091	<0.001*	1.728	1.397–2.138	<0.001*
Axial plane	1.519	1.326–1.739	<0.001*	-	-	-
LCL complex abnormality, *n* (%)						
Grade 0						
Grade 1	3.489	1.961–6.206	<0.001*	2.522	1.202–5.288	0.014*
Grade 2	16.923	1.815-157.692	0.013*	4.203	0.297–59.490	0.288
Muscle oedema, *n* (%)	10.679	3.814–29.904	<0.001*	9.530	2.621–34.653	<0.001*
Joint effusion, *n* (%)	2.596	0.734–9.191	0.139			
Cartilaginous defect, *n* (%)	1.693	0.957–2.993	0.070			
Subchondral cyst, *n* (%)	1.462	0.845–2.528	0.174			
Subchondral marrow oedema, *n* (%)	2.091	0.813–5.377	0.126			

MRI: Magnetic resonance imaging; LCL; lateral collateral ligament; CET: common extensor tendon; OR; odds ratio; CI: confidence interval

Table 3 Performance of different machine learning combined models in the external validation set

Machine learning Model	AUC	95% CI	Accuracy	Sensitivity	Specificity	PPV	NPV
SVM	0.79	0.64–0.92	0.75	0.68	0.78	0.59	0.84
LR	0.80	0.66–0.92	0.73	0.68	0.76	0.57	0.84
KNN	0.61	0.45–0.76	0.43	0.74	0.29	0.33	0.71
Decision Tree	0.43	0.31–0.54	0.33	0.68	0.17	0.28	0.54
RF	0.93	0.86–0.99	0.85	0.84	0.85	0.73	0.92
XGBoost	0.91	0.82–0.99	0.92	0.74	1	1	0.89
LightGBM	0.96	0.91–0.99	0.87	0.74	0.93	0.82	0.88

AUC: area under the curve; CI: confidence interval; PPV: positive predictive value; NPV: negative predictive value; SVM: Support Vector Machine; LR: logistic regression; KNN: K-Nearest Neighbors; RF: Random Forest; LightGBM: Light Gradient Boosting Machine

Table 4 Accuracy and predictive value of the clinicoradiological model and combined model

	Accuracy	Sensitivity	Specificity	PPV	NPV
Training set					
Clinicoradiological model	0.784	0.363	0.976	0.892	0.772
Combined model	0.841	0.726	0.893	0.768	0.881
Internal validation set					
Clinicoradiological model	0.835	0.595	0.949	0.846	0.832
Combined model	0.757	0.487	0.885	0.668	0.784
External validation set					
Clinicoradiological model	0.867	0.790	0.902	0.790	0.902
Combined model	0.867	0.73	0.927	0.824	0.884

PPV: positive predictive value; NPV: negative predictive value

the external validation set (Table 3). The implementation of the best machine learning model source code is available in a public repository with the URL: https://github.com/yxz95/ElbowMRImodel. Table 4 shows the predictive performances of the clinicoradiological and combined models. In the external validation set, the combined model significantly outperformed the clinicoradiological model (AUC: 0.96 [0.91–0.99] vs. 0.83 [0.67–0.96] by DeLong's test) (Fig. 3).

SHAP provided a quantitative explanation for the combined model. Figure 4 shows global interpretability, whereas Fig. 5 shows local interpretability (explanation of an individual case). The feature importance rankings of the 7 predictors were shown in Fig. 4A. The "CET tear size in the coronal plane" and

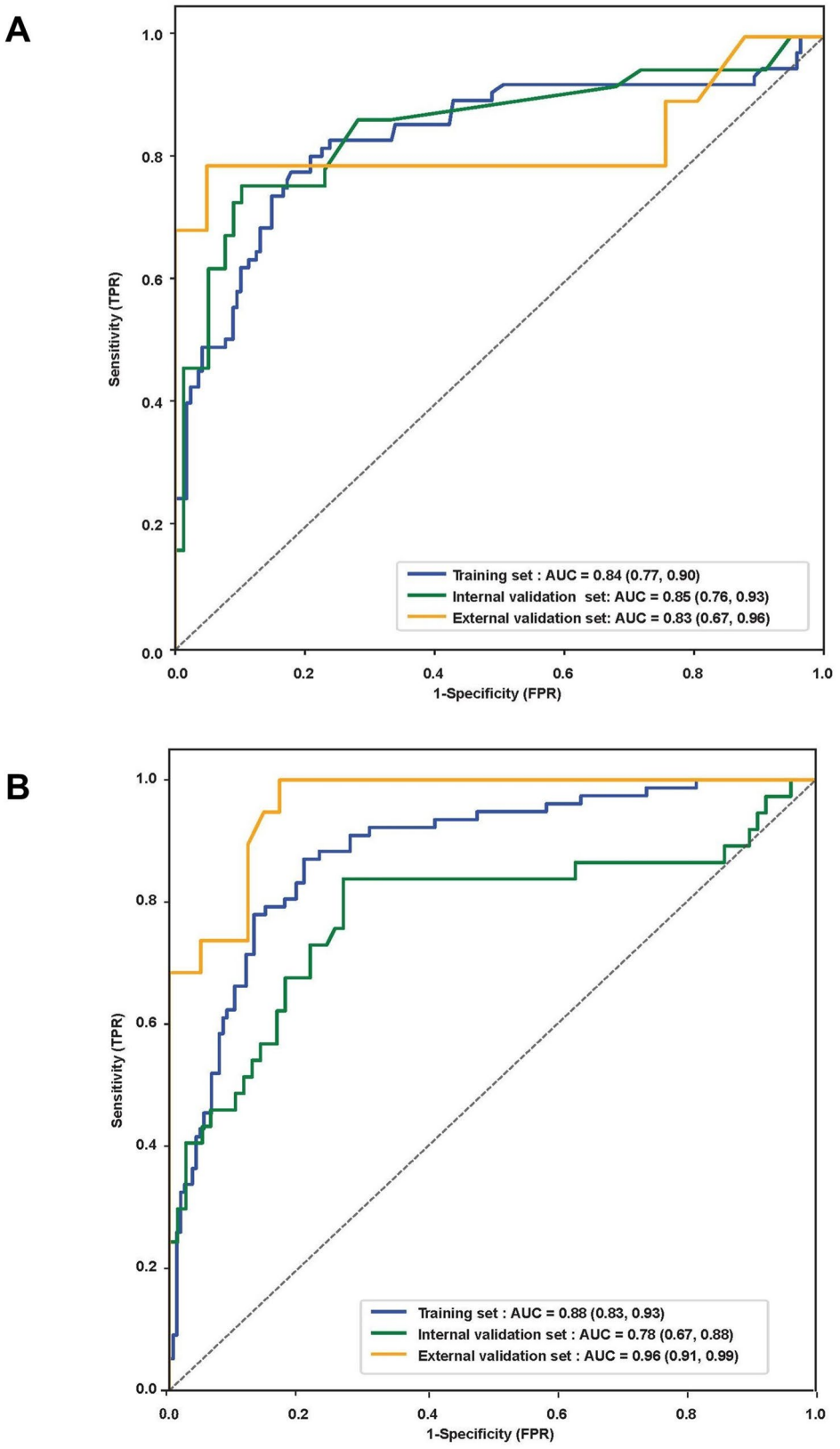

Fig. 3 The ROC curves in the training, internal, and external validation sets. (**A**) ROC curve of the clinical model. (**B**) ROC curve of the combined model. ROC: receiver operating characteristic

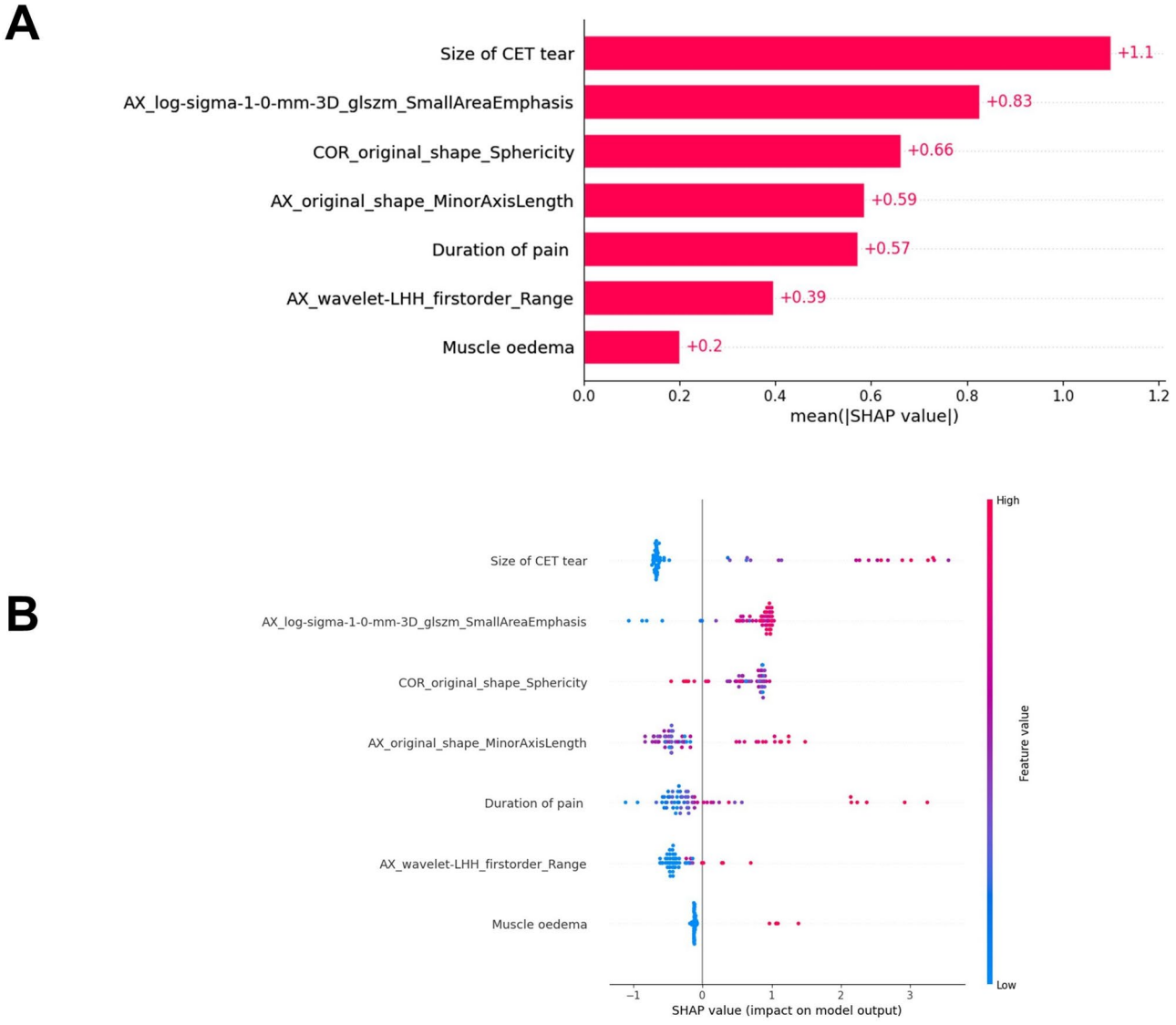

Fig. 4 SHAP summary plots of the combined model. (**A**) The SHAP bar chart. Ranking the importance of each feature in the final model output. Higher SHAP values indicate a more significant contribution of the model to a particular feature. (**B**) The SHAP beeswarm plot. A comprehensive visualization of the cumulative influence of each feature. Each dot represents an individual sample, and the color represents the value of each feature. The x-axis represents the SHAP values, and a positive SHAP value indicates that it has positive effects on the predictions of the model and vice versa. SHAP: Shapley additive explanation

"AX_log-sigma-1-0-mm-3D_glszm_SmallAreaEmphasis" ranked in the top two in terms of their average impact on the combined model's prediction of prognosis for conservative treatment. Figure 4B shows the distribution of the contribution of each feature to the model output. We can see that increases in the "CET tear size in the coronal plane" and "AX_log-sigma-1-0-mm-3D_glszm_SmallAreaEmphasis" have a positive impact and push the prediction toward conservative treatment failure. Figure 5 shows two typical examples of correctly predicting whether a patient will fail conservative treatment. The SHAP force plot shows the positive and negative impacts of all the features in the combined model on the predicted outcome in a single case.

Discussion

While most LE patients recover within 6–12 months with conservative treatment, some patients suffer for 24 months or more, significantly affecting daily life [2, 22]. Improved prognostic tools are needed to predict treatment outcomes. In this study, we developed a machine learning model that integrates clinicoradiological and elbow MRI radiomic features to predict the failure of conservative treatment in patients with LE. The combined model demonstrated robust performance in the

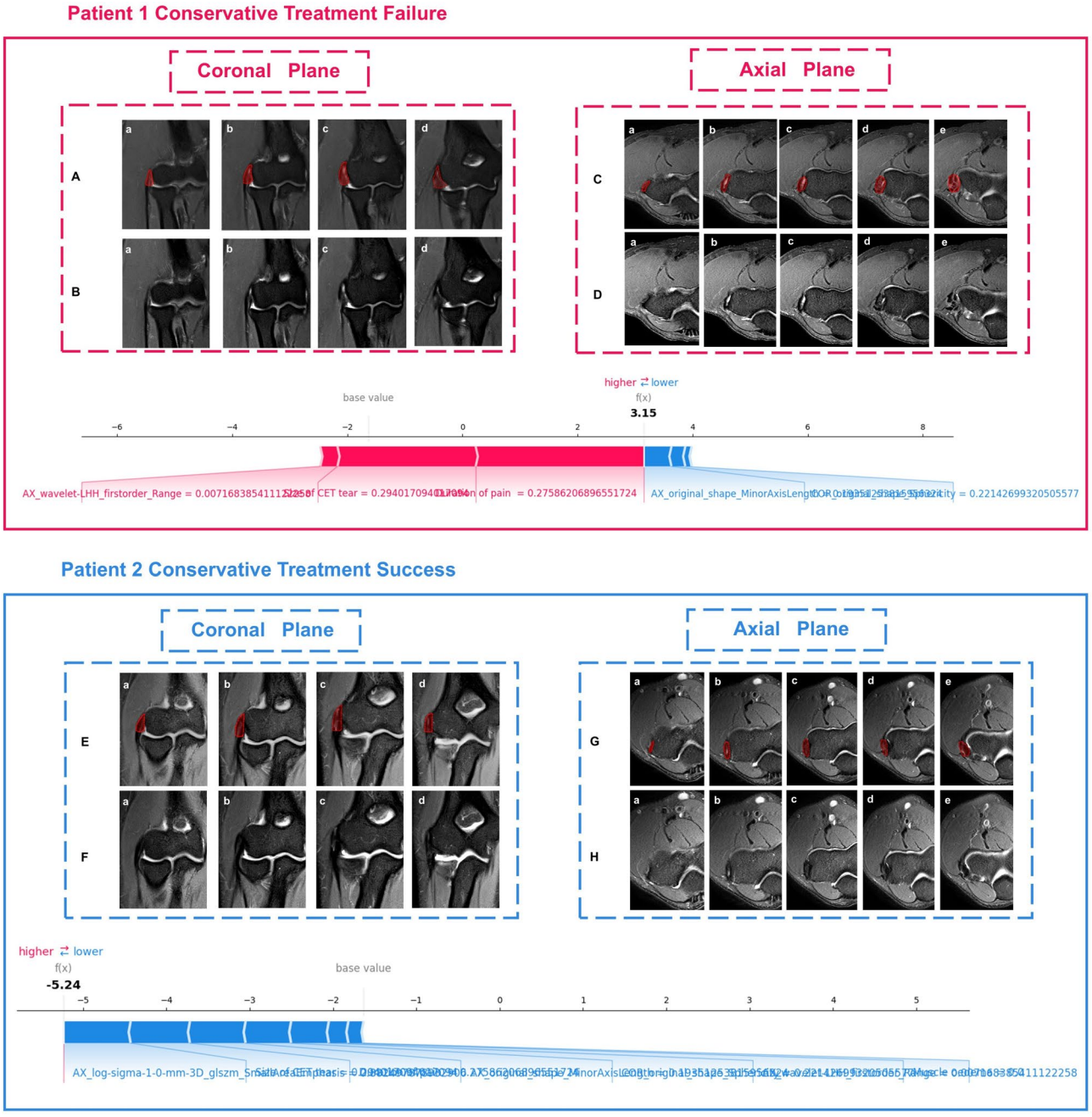

Fig. 5 Individual visualization of the combined model by SHAP force plots explains how the model predicts the outcome of two patients with lateral epicondylitis who were treated conservatively. Patient 1 shows an example of a patient who underwent surgery after conservative treatment failed. Patient 2 shows an example of a patient with effective conservative treatment. Images B_{a-d}, D_{a-e}, F_{a-d}, and H_{a-e} were consecutive layers of the original DICOM image in coronal and axial views, respectively. Images A_{a-d}, C_{a-e}, E_{a-d}, and G_{a-e} were manual annotation diagrams. The red area is the ROI area. The base value represents the average prediction of the combined model, and f (x) represents its final prediction probability. The width of the bar indicates how much the feature affects the conservative treatment outcome; the closer to the f (x) value, the wider the bar of the feature, and the greater the effect on the conservative treatment outcome. The red color indicates high feature values, which are usually associated with an increased likelihood of conservative treatment failure. Conversely, the blue color indicates low feature values, which are usually associated with a decreased likelihood of conservative treatment failure. The SHAP force plot for Patient 1 shows that pain duration was the feature that contributed most positively to predicting patient failure of conservative treatment in the combined model. The SHAP force diagram for Patient 2 shows that "AX_log-sigma-1-0-mm-3D_glszm_SmallAreaEmphasis" was the feature that contributed most negatively to predicting patient failure of conservative treatment in the combined model

external validation set and provided high interpretability through SHAP analysis, which offers valuable insights into the factors influencing the model's predictions.

Several studies have recognized that accurate early prognostic prediction and the development of a rational treatment plan are essential for treating LE and have analyzed the factors associated with the failure of conservative treatment. Elisa et al. [4] compared demographic and disease-specific characteristics of the non-surgical and surgical treatment groups and found that characteristics such as radial tunnel syndrome and symptom duration longer than 12 months were independent predictors of surgical treatment. Studies also analyze the relationship between MRI and prognosis in patients with LE [5, 7, 23, 24]. A study on 60 patients with LE developed a model that included persistent pain, CET abnormalities in the longitudinal plane, and muscle edema to predict operative treatment [5]. Ikeda et al. [7]concluded that treatment strategies including surgery should be considered for patients with severe MRI findings of CET and LCL complex. In contrast, Kessler et al. [24]concluded that the size of the CET defect size on MRI did not predict the success or failure of conservative treatment. Therefore, the clinical generalizability of its existing studies has been questioned due to limited patient samples, lack of internal and external validation, and controversial results of studies. In addition, the relationship between radiomics and the prognosis of LE has not been previously evaluated. Given that CET in patients with LE undergoes histologically progressive degenerative changes, the severity of which can only be definitively assessed by postoperative pathology. Radiomics is a field that extracts a large number of quantitative features from medical images to capture subtle changes in CET and LCL complex. Because radiomics extracts high-dimensional features on a spatial scale, it is easier to detect hidden information and reflect the underlying pathophysiology associated with lateral epicondylitis than traditional imaging features, thus providing a more objective and reproducible assessment method.

Radiomics has emerged as a powerful tool for evaluating tendon and ligament injuries, offering high diagnostic efficacy [12]. Cheng et al. [13] found that a machine-learning-based multi-sequence MRI radiomics model demonstrated superior performance in diagnosing anterior cruciate ligament tears. Zhan et al. [14] found that MRI-based radiomics could accurately determine the presence and extent of tears in the supraspinatus tendon. Droppelmann et al. [25] found that ultrasound image-based machine learning algorithms have high diagnostic accuracy for CET tears in patients with LE, with an AUC of 0.991. However, the application of radiomics to the elbow joint has been limited. To our knowledge, this is the first study to assess the ability of radiomic features to predict failure of conservative treatment for LE with further external validation, providing clinicians with a tool to make timely decisions regarding surgical intervention, thereby minimizing patient discomfort and reducing healthcare costs.

There are two main reasons why the ROI of our study included not only CET but also the LCL complex. (1) LE is a degenerative process that occurs at the origin of the CET [1]. However, previous studies have shown that CET is not always an isolated lesion of LE, and LCL complex injury often occurs with progressive injury of CET [23, 26]. This is further supported by our study, where the rate of LCL complex abnormalities was significantly higher in the conservative treatment failure group. (2) Given the close anatomical relationship between the LCL complex and CET, both originating from the lateral epicondyle, it is challenging to distinguish between the tendon and ligament on MRI.

In this research, the performance of five machine learning algorithms was compared. Eventually, LightGBM was chosen as the optimal algorithm, and it demonstrated excellent performance and generalizability in external validation. Our findings demonstrate that integrating clinicoradiology and radiomics features is critical for optimizing predictive performance. While traditional clinicoradiological features are effective in capturing macroscopic patterns of disease progression, they exhibit limited sensitivity in detecting subtle pathophysiologic changes. Radiomics analysis, while capable of quantifying imperceptible tissue features, lacks strong clinical relevance. Our combined model compensates for these respective limitations by incorporating quantitative imaging features while retaining clinically routine features, ultimately achieving higher predictive accuracy.

Lack of interpretability is a significant challenge associated with machine learning predictive models based on radiomics, which limits their application in clinical practice. We calculated the SHAP values to interpret the combined model and address this issue. SHAP recognizes specific patterns learned through complex machine learning algorithms and provides both global and local interpretability for radiomics models. SHAP enables interpretation and visualization of LightGBM models in a clinician-friendly manner through SHAP plots. From a global perspective, the SHAP summary plots provide an intuitive and concise graph that represents the range and importance distribution of the feature's impact on the output of the combined model (the more important the feature, the higher the SHAP value, and the wider the range of dots). In this study, clinicians could see how the features' value impacted the assessment through the dots' range and color. At the same time, clinicians were able to simultaneously identify the feature that contributed the most to the model as the radiologic feature "CET tear size

on the coronal plane." In LE, the CET is the primary site of stress injury caused by repetitive contractions of the forearm extensor muscles. Several studies have shown that patients with more extensive CET tears treated conservatively have a poorer prognosis and require further operative treatment. Furthermore, the radiomics feature "AX_log-sigma-1-0-mm-3D_glszm_SmallAreaEmphasis" had the second highest average impact on the combined model predictions. It describes the distribution of small-sized zones, which may correspond to areas of inflammation or tissue damage [27]. In our study, the 'AX_log-sigma-1-0-mm-3D_glszm_SmallAreaEmphasis' of the conservative treatment success group was significantly lower than that of the conservative treatment failure group, which implies that the conservative treatment success group was observed in axial images with CET and LCL complex areas that were smaller and finer textured.

By understanding the influence of features on the combined model and their underlying pathophysiological mechanisms, clinicians can utilize the SHAP force plot for a local interpretation of individual patient predictions. The clinician simply compares the SHAP value for a given case to a baseline threshold, with values above the baseline threshold indicating that the patient belongs to the conservative treatment failed group. This visualization method intuitively displays feature contributions through color-coded arrows (red indicates an increased probability that the patient will require surgery) and scale lengths representing the relative degree of impact. As shown in Fig. 5A (red), 'CET tear size' has a positive effect on the assessment of conservative treatment failure, and the length of the arrow for 'CET tear size' is longer than that for 'duration of pain' representing a greater effect on the model than that for 'duration of pain'.

Our predictive model enables clinicians to make more accurate and timely diagnostic assessments, facilitating prompt intervention and optimized management strategies. For high-risk patients identified by the model, clinicians may consider early surgical intervention with intensified monitoring to prevent disease progression. Conversely, low-risk patients can be managed with standard conservative protocols while avoiding unnecessary surgical procedures. This reduces pain and further damage and enhances patients' quality of life by expediting the rehabilitation process and enabling faster return to daily activities and physical function.

Our study has several limitations. First, as a retrospective study, our study inevitably presents potential selection bias. Although we used standardized image preprocessing and harmonization techniques, subtle differences in MRI protocols, scanner parameters, and clinical evaluation criteria among the three hospitals may affect the generalizability of feature extraction and models. Future prospective multicenter studies using standardized imaging protocols would be valuable in validating our findings and improving the robustness of our conclusions. Additionally, we recommend investigating ways to combine imaging modalities such as X-ray, CT, and ultrasound with clinical features to develop a more comprehensive predictive model that can significantly improve the clinical applicability of the model in different healthcare settings. Second, while our model is based on radiomic features, incorporating deep learning approaches may further enhance predictive accuracy. Finally, some patients had relatively severe symptoms as they had received poor results from conservative treatment in other hospitals, and then came to our hospital.

In conclusion, we developed and validated a noninvasive and robust LightGBM machine learning model that combines MRI radiomics and clinicoradiological features to predict the failure of conservative treatment in LE. The SHAP provides a bridge for personalized prediction, which may aid clinical decision-making for the individualized treatment of LE.

Abbreviations

MRI	Magnetic resonance imaging
SHAP	SHapley Additive exPlanations
CET	Common extensor tendon
LCL	Lateral collateral ligament
T1WI	T1-weighted image
PD-FS	Proton density-weighted imaging with fat suppression
ROI	Regions of interest
AUC	Area under the curve
ROC	Receiver operating characteristic curve
IQR	Interquartile range
ICC	Intraclass correlation coefficients

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13018-025-05901-1.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

RB was responsible for searching for research related to the subject content, designing the research proposal, and writing the introductory section of the manuscript. JC and PW performed the MRI image analysis, and ROI outlining, and wrote the discussion and introduction sections of the manuscript. XZ and PZ collected MR images from other hospitals, reviewed the image features, and revised the manuscript. YY was responsible for the statistical processing and writing the methods section.

Funding

This study was supported by the National Natural Science Foundation of China (Grant numbers 82171921, 81771809) and the Beijing Natural Science Foundation of China (Grant number 7202063).

Data availability

The data that support the findings of this study are available from Beijing Jishuitan Hospital but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly

available. Data are however available from the authors upon reasonable request and with permission of Beijing Jishuitan Hospital.

Declarations

Ethics approval and consent to participate
This study was approved by the ethics committee of Beijing Jishuitan Hospital. For this retrospective analysis informed consent is not required.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Radiology, Beijing Jishuitan Hospital, Capital Medical University, Beijing 100035, China
[2]Department of Radiology, The Third Affiliated Hospital, Southern Medical University, Zhongshan Avenue West, Tianhe District, Guangzhou 510515, China
[3]Department of Radiology, Beijing Geriatric Hospital, Beijing 100095, China
[4]Department of Radiology, Pomona Valley Hospital Medical Center, Pomona, CA 91767, USA

Received: 19 February 2025 / Accepted: 9 May 2025
Published online: 24 May 2025

References

1. Wolf JM. Lateral epicondylitis. N Engl J Med. 2023;388(25):2371–7.
2. Karjalainen T, Buchbinder R. Is it time to reconsider the indications for surgery in patients with tennis elbow. Bone Joint J. 2023;105–B:109–11.
3. Smidt N, Lewis M, VAN DER Windt DA, Hay EM, Bouter LM, Croft P. Lateral epicondylitis in general practice: course and prognostic indicators of outcome. J Rheumatol. 2006;33(10):2053–59.
4. Knutsen EJ, Calfee RP, Chen RE, Goldfarb CA, Park KW, Osei DA. Factors associated with failure of nonoperative treatment in lateral epicondylitis. Am J Sports Med. 2015;43(9):2133–7.
5. Jeon JY, Lee MH, Jeon IH, Chung HW, Lee SH, Shin MJ. Lateral epicondylitis: associations of MR imaging and clinical assessments with treatment options in patients receiving Conservative and arthroscopic managements. Eur Radiol. 2018;28(3):972–81.
6. Miyamura S, Temporin K, Miyata S, Miyake T, Shimada K. Arthroscopic debridement for refractory lateral epicondylitis results for substantial improvement in tendinosis scores and good clinical outcomes: qualitative and quantitative magnetic resonance imaging analysis. Arthroscopy. 2022;38(12):3120–9.
7. Ikeda K, Ogawa T, Ikumi A, et al. Magnetic resonance imaging predicts outcomes of Conservative treatment in patients with lateral epicondylitis. J Orthop Sci. 2024;29(3):795–801.
8. Dewan AK, Chhabra AB, Khanna AJ, Anderson MW, Brunton LM. MRI of the elbow: techniques and spectrum of disease: AAOS exhibit selection. J Bone Joint Surg Am. 2013;95(14):e991–13.
9. Zhou Y, Wu D, Yan S, et al. Feasibility of a Clinical-Radiomics model to predict the outcomes of acute ischemic stroke. Korean J Radiol. 2022;23(8):811–20.
10. Kim KM, Hwang H, Sohn B, et al. Development and validation of MRI-Based radiomics models for diagnosing juvenile myoclonic epilepsy. Korean J Radiol. 2022;23(12):1281–9.
11. Pak E, Choi KS, Choi SH, et al. Prediction of prognosis in glioblastoma using radiomics features of dynamic Contrast-Enhanced MRI. Korean J Radiol. 2021;22(9):1514–24.
12. Fritz B, Yi PH, Kijowski R, Fritz J. Radiomics and deep learning for disease detection in musculoskeletal radiology: an overview of novel MRI- and CT-Based approaches. Invest Radiol. 2023;58(1):3–13.
13. Cheng Q, Lin H, Zhao J, Lu X, Wang Q. Application of machine learning-based multi-sequence MRI radiomics in diagnosing anterior cruciate ligament tears. J Orthop Surg Res. 2024;19(1):99.
14. Zhan J, Liu S, Dong C, et al. Shoulder MRI-based radiomics for diagnosis and severity staging assessment of surgically treated supraspinatus tendon tears. Eur Radiol. 2023;33(8):5587–93.
15. Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. Clin Transl Sci. 2024;17(11):e70056.
16. Rodríguez-Pérez R, Bajorath J. Interpretation of compound activity predictions from complex machine learning models using local approximations and Shapley values. J Med Chem. 2020;63(16):8761–77.
17. Quinn KN, Wilber H, Townsend A, Sethna JP. Chebyshev approximation and the global geometry of model predictions. Phys Rev Lett. 2019;122(15):158302.
18. Taiyeb Khosroshahi M, Morsali S, Gharakhanlou S, et al. Explainable artificial intelligence in neuroimaging of Alzheimer's disease. Diagnostics (Basel). 2025;15(5):612.
19. Du R, Lee VH, Yuan H, et al. Radiomics model to predict early progression of nonmetastatic nasopharyngeal carcinoma after intensity modulation radiation therapy: A multicenter study. Radiol Artif Intell. 2019;1(4):e180075.
20. Chen Y, Chen S, Tang W, et al. Multiparametric MRI radiomics with machine learning for differentiating HER2-Zero, -Low, and -Positive breast cancer: model development, testing, and interpretability analysis. AJR Am J Roentgenol. 2025;224:e2431717.
21. Ning C, Ouyang H, Xiao J, et al. Development and validation of an explainable machine learning model for mortality prediction among patients with infected pancreatic necrosis. EClinicalMedicine. 2025;80:103074.
22. Soeur L, Desmoineaux P, Devillier A, Pujol N, Beaufils P. Outcomes of arthroscopic lateral epicondylitis release: should we treat earlier. Orthop Traumatol Surg Res. 2016;102(6):775–80.
23. Cha YK, Kim SJ, Park NH, Kim JY, Kim JH, Park JY. Magnetic resonance imaging of patients with lateral epicondylitis: relationship between pain and severity of imaging features in elbow joints. Acta Orthop Traumatol Turc. 2019;53(5):366–71.
24. Kessler RE, Day MS, Tyler TF, et al. Predictive value of magnetic resonance imaging in outcomes of nonsurgical treatment of lateral epicondylitis. JSES Int. 2022;6(2):305–8.
25. Droppelmann G, Tello M, García N, Greene C, Jorquera C, Feijoo F. Lateral elbow tendinopathy and artificial intelligence: binary and multilabel findings detection using machine learning algorithms. Front Med (Lausanne). 2022;9:945698.
26. Kholinne E, Liu H, Kim H, Kwak JM, Koh KH, Jeon IH. Systematic review of elbow instability in association with refractory lateral epicondylitis: myth or fact. Am J Sports Med. 2021;49(9):2542–50.
27. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for High-Throughput image-based phenotyping. Radiology. 2020;295(2):328–38.

Publisher's note