Full Length Article

Machine-learning models for diagnosis of rotator cuff tears in osteoporosis patients based on anteroposterior X-rays of the shoulder joint

Yu Zhao, Jingjing Qiu, Yang Li, Muhammad Attique Khan, Lei Wan, Lihua Chen

Objective: This study aims to diagnose Rotator Cuff Tears (RCT) and classify the severity of RCT in patients with Osteoporosis (OP) through the analysis of shoulder joint anteroposterior (AP) X-ray-based localized proximal humeral bone mineral density (BMD) measurements and clinical information based on machine learning (ML) models.

Methods: A retrospective cohort of 89 patients was analyzed, including 63 with both OP and RCT (OPRCT) and 26 with OP only. The study analyzed a series of shoulder radiographs from April 2021 to April 2023. Grayscale values were measured after plotting ROIs based on AP X-rays of shoulder joint. Five kinds of ML models were developed and compared based on their performance in predicting the occurrence and severity of RCT from ROIs' grayscale values and clinical information (age, gender, advantage side, lumbar BMD, and acromion morphology (AM)). Further analysis using SHAP values illustrated the significant impact of selected features on model predictions.

Results: R1-6 had a positive correlation with BMD respectively. The nine variables, including grayscale R1-6, age, BMD, and AM, were used in the prediction models. The RF model was determined to be superior in effectively diagnosing RCT in OP patients, with high AUC scores of 0.998, 0.889, and 0.95 in the training, validation, and testing sets, respectively. SHAP values revealed that the most influential factors on the diagnostic outcomes were the grayscale values of all cancellous bones in ROIs. A column-line graph prediction model based on nine variables was constructed, and DCA curves indicated that RCT prediction in OP patients was favored based on this model. Furthermore, the RF model was also the most superior in predicting the types of RCT within the OPRCT group, with an accuracy of 86.364% and 73.684% in the training and test sets, respectively. SHAP values indicated that the most significant factor affecting the predictive outcomes was the AM, followed by the grayscale values of the greater tubercle, among others.

Conclusions: ML models, particularly the RF algorithm, show significant promise in diagnosing RCT occurrence and severity in OP patients using conventional shoulder X-rays based on the nine variables. This method presents a cost-effective, accessible, and non-invasive diagnostic strategy that has the potential to substantially enhance the early detection and management of RCT in OP patient population.

1. Introduction

Osteoporosis (OP) is one of the most common aging-related systemic metabolic diseases, characterized by decreased bone mass and bone microarchitectural deterioration, with reduced bone strength leading to increased fracture risk[1]. A meta-analysis of 40 studies from Asia,
Europe, and the Americas, covering 79,127 people aged 50 to 85, found that the global prevalence of OP in older adults was 21.7%[2]. Due to decreased bone mineral density (BMD) and vascular distribution of the tendons, rotator cuff tears (RCT) are also common in the elderly population[3]. RCT is one of the most clinically common musculoskeletal disorders in middle-aged and elderly patients, often resulting in shoulder pain, weakness, and dysfunction[4], which progressively worsen with age[5]. Diebold et al.[6] conducted a retrospective study and found that the prevalence of RCT increased with age. The incidence of RCT increases linearly at 5% per year between the ages of 50 and 69. The prevalence of RCT is up to 23% of patients over 50 and substantially increases after the age of 70, reaching as high as 62% over 80.

As the world’s population ages, RCT and OP are public health issues that cannot be ignored. OP is strongly associated with RCT. In terms of disease risk, a cohort study revealed that individuals diagnosed with OP face a 1.79-fold increased risk of RCT compared to those without OP. Anatomically, the greater and lesser tuberosities of the humerus serve as attachment points for all rotator cuff tendons in the proximal humerus. The strength of fixation of the rotator cuff attachment points are influenced by the proximal humerus sclerotonin[7]. Proximal humeral OP may contribute as one of the etiological factors for RCT[8]. Age over 70 and RCT are contributing factors to OP of the greater tuberosity of the humerus (GTH) in women[9]. Compared to individuals aged 40, those aged 70 or older experience a reduction of over 40% in BMD at the GTH[10]. Several studies have observed a significant decrease in BMD at the GTH in patients with RCT, correlating the extent of this reduction with the severity of the injury[11,12]. Pre-existing OP in patients with RCT also significantly contribute to localized OP of the proximal humerus in these patients[9].

RCT is conventionally managed via arthroscopic repair techniques. However, postoperatively, OP emerges as a distinct risk factor that compromises tendon-bone integration and predisposes to re-tearing[13]. In the GTH, localized OP escalates the likelihood of anchor nail dislodgment and loosening. This condition further modifies the microenvironment essential for tendon-bone healing, potentially resulting in postoperative rotator cuff re-tears. This may result in secondary surgery, adding to the economic burden of the patient[13]. It is estimated that the re-tear rate after repairing RCT is as high as 90% in individuals over 50[14,15] and even higher in females over 70[16]. Additionally, the failure rate of RCT repair is as high as 68%, and the bone quality of the proximal humerus is considered one of the factors that affect the outcome of the repair[17]. In turn, after a period of RCT, many patients develop localized OP secondary to RCT, especially in the region of the GTH[18]. Accordingly, the early detection of RCT in patients with OP is imperative to alleviate suffering and enhance prognostic outcomes.

Researchers have focused on the localized BMD of the proximal humerus in RCT studies for a long time. Currently, there is no standardization for measuring proximal humerus BMD, including the selected imaging modality, the method of measurement, and the localization of the region of interest (ROI). Traditional methods of BMD measurement, such as measuring lumbar or hip BMD, may underestimate localized BMD in the proximal humerus[13]. What’s more, the measurements and microstructural analyses, such as dual-energy X-ray absorptiometry (DXA), quantitative computed tomography (QCT), and micro-computed tomography scan (CT), are not widely used for measuring bone density in the shoulder due to their high cost and the requirement for specialized software[19-21].

Therefore, it is essential to investigate a straightforward and efficient approach for the early detection of RCT in OP patients. Since the proximal humeral sclerotomy is strongly associated with the degree of RCT, we attempted to predict the likelihood of developing RCT in patients with OP by evaluating the bone density of the shoulder joint.

Currently, standard X-ray examinations remain the most widely distributed and frequently used examination and are the first examination performed in patients with RCT. We proposed using the X-ray-based localized proximal humeral BMD as a diagnostic tool for the early detection of RCT in OP patients. Based on previous studies, we chose anteroposterior (AP) X-ray of the shoulder joint as the source of proximal humeral BMD measurements. We modified the localization method of Regions of Interest (ROIs) and then measured the gray values of the corresponding sites[11,12]. We hypothesized that proximal humeral BMD based on greyscale values of shoulder radiographs was directly related to the occurrence and type of RCT. Bone density of the proximal humerus can be quantified by the greyscale values of the ROIs, and machine learning methods are used to predict the RCT in OP patients. In quantifying image information, machine learning methods are employed to process and interpret X-ray image data for diagnostic purposes[22,23]. Machine learning is particularly effective at extracting critical features from complex image data and analyzing them quantitatively. This is especially important for features that are difficult to assess quantitatively using traditional imaging[24]. In conclusion, we aim to investigate the machine-learning models for the diagnosis of RCT in OP Patients and the assessment of RCT severity in these patients based on AP X-rays of the shoulder joint.

2. Methods

This study was reviewed and approved by the Medical Ethics Committee of the Third Affiliated Hospital of Guangzhou University of Chinese Medicine (TAHGZUCM) (No. PJ-KY-20210401-008) and registered in the Chinese Clinical Trial Registry (ChiCTR) (No. ChiCTR2300074461). The retrospective analysis was conducted on cases at TAHGZUCM from April 2021 to April 2023. We obtained the information through the electronic hospital information system. The information includes gender, age, whether the affected shoulder is dominant, lumbar BMD, surgical records, and documents of shoulder joint AP X-rays, etc.

2.1. Inclusion and exclusion criteria

2.1.1. Patients

After rigorous screening, we included 89 patients in our study. Of these, 63 had both OP and RCT (OPRCT) (48 females and 15 males; mean age, 64.86 ± 8.79 years; age range, 47-88 years), while 26 had OP only (OPO) (21 females and 5 males; mean age, 69.35 ± 9.60 years; age range, 54-88 years).

The inclusion criteria for the patients are as follows: (1) Individuals aged 40 years or older who have completed a medical history collection; (2) Those who have undergone standard AP X-ray of the shoulder joint; (3) Participants who have completed a DXA scan, with a BMD measurement result of T≤-2.5, according to the diagnostic criteria for OP by the Clinical Guidelines From the American College of Physicians[25]; (4) Patients diagnosed with RCT according to the diagnostic criteria provided in the Clinical Practice Guidelines by the American Academy of Orthopaedic Surgeons (AAOS)[26], and who have undergone arthroscopic surgery of the shoulder joint. (5) The radiological examinations and surgical procedures were all conducted at TAHGZUCM. The exclusion criteria for this study are as follows: (1) Age <40; (2) RCT caused by acute trauma; (3) Concurrent conditions such as shoulder arthritis, fractures, bone tumors, other diseases affecting bone metabolism, long-term use of medications influencing bone metabolism, and congenital or acquired metabolic bone diseases; (4) History of abuse or dependence on alcohol, tobacco, drugs, narcotics, etc. within the past year.

2.1.2. AP X-rays of the Shoulder Joint

For this study, we measured and evaluated AP X-rays of the shoulder joint primarily. The AP X-rays of the patients must meet the following criteria. Inclusion criteria: (1) Standard standing AP X-ray of the affected shoulder (preoperative); (2) Radiograph background (excluding the human body) grayscale value = 0; (3) Complete, clear, and unobstructed visualization of the greater tubercle, lesser tubercle, humeral head,
The shoulder arthroscopy procedure records included the size of the RCT, whether anchors were used, and the number of anchors used. All shoulder arthroscopy procedures were performed by the same senior physician in the sports medicine department. The size of the RCT was recorded using the Cofield classification method[27], which was used to classify RCT based on the extent of the tears observed during repair surgery. The tears were classified as small (<1 cm), moderate (1-3 cm), large (3-5 cm), or massive (>5 cm) and recorded as 1, 2, 3, and 4, respectively.

2.3. Image Preprocessing

2.3.1. The Criteria for Selecting ROIs

RCT generally occurs in the supraspinatus tendon (which attaches to GTH), followed by the infraspinatus tendon (which attaches to lesser tuberosity of humerus (LTH)). Typically, tears begin in the anterior portion of the supraspinatus, below the acromion, and near the GTH, and gradually extend to the other tendons[7]. Additionally, the humeral head (HH) has a high BMD in the cancellous bone of the proximal humerus. A study utilizing OP rabbit models found that increasing the bone mass of the HH can enhance the strength of the rotator cuff insertion site, thereby reducing the likelihood of RCT[28]. Other important areas for the study include the surgical neck of the humerus and the cortical bone. Therefore, we established 6 ROIs (Fig. 2) to select locations that comprehensively and representatively reflect the bone density of the proximal humerus.

2.3.2. Definition and Measurement of ROIs

All participants’ AP X-rays of the shoulder joint in the original Digital Imaging and Communications in Medicine (DICOM) format were exported. Auxiliary lines and circles were added to delineate the ROIs by using Photoshop 2020, as illustrated in Fig. 2. The depth of images was...
set to 8-bit, with a default grayscale value range of 0-255 in the IMAGE J software (https://imagej.nih.gov/ij/). The grayscale values of regions 1-6 were represented as R1-6.

2.3.3. Acromion Morphology

The diagnostic contents of the acromion morphology (AM) were obtained from the X-ray reports, and the classification of the AM was recorded as Type I (flat), Type II (curved), and Type III (hooked) [29], and recorded as 1, 2, and 3, respectively.

2.3.4. Imaging Equipment and Parameters

To ensure standardization of individual x-rays and grayscale values, our hospital maintains consistent imaging conditions for patients undergoing medical imaging of the same anatomical structures. The shoulder joint X-rays were conducted using a medical X-ray device: Shimadzu (RAD SPEED M), a 500mA digital radiography (DR) system. Imaging parameters included a voltage of 72kV, current of 400mA, exposure time of 28ms, current per second of 11.2mAs, and a photography distance of 100cm. BMD measurements were performed using an Italian Unigamma X-ray Plus dual-energy X-ray bone densitometer (Device No. 021300200001), with measurement parameters of 86kV voltage, 0.4mA current, a current duration of 51s, and a testing duration of 99s.

2.4. Establishment and Development of Machine Learning Models

Develop and compare prediction models using five algorithms. The methods used to build prediction models include Logistic Regression, Random Forest (RF), Support Vector Machine, Decision Tree, and MLP model. All prediction models were performed by Python 3.7. The patients were randomly divided into a training group and a testing group (7:3). The validation dataset was used to evaluate and compare the performance of each model. The ability of the model to predict patients with different Cofield classification was evaluated using the area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1 score. Use the "prediction and observation" chart to evaluate model calibration.

2.5. Model Optimization and Evaluation

Double cross-validation was used to evaluate the predictive ability of the model to ensure its stability. The training set was randomly divided into two groups. In each iteration of the second cross-validation, one group was randomly selected for training, and the remaining was considered as the validation set. In each training section of the model, randomly select a 30% dataset from the training dataset to test the model's performance. Subsequently, we quantified the model discrimination using ROC curve analysis and evaluated its predictive accuracy using the obtained AUC and calibration plots. Decision Curve Analysis (DCA) evaluates clinical usefulness and net benefits. Use Shapley Additive Interpretation (SHAP) to evaluate feature importance. A higher absolute SHAP value is associated with the feature that has the most significant impact on the predicted score of the model. In addition, the distribution of eigenvalues and their correlation with model predictions.

Fig. 2. Schematic diagram of ROI. Note: 1: GTH; 2: LTH, 3: a square area on the head of the humerus (1cm²); 4: a square area on the surgical neck of the humerus (1cm²); 5/6: proximal humeral cortex (lateral/medial); a/b: circles tangent to the lateral aspect of the humeral stem; line A: the median axis of the humeral stem; lines B/C/D parallel to line A; straight line B over the endpoint of the orthogonal view of the anatomical neck of the humerus; line C over the edge of the orthogonal view of LTH; and line D tangent to the orthogonal view edge of the articular glenoid.

Table 1

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>OPCT group</th>
<th>OPO group</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>63</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td>0.638</td>
</tr>
<tr>
<td>Female</td>
<td>48 (53.9%)</td>
<td>21 (23.6%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>15 (16.9%)</td>
<td>5 (5.6%)</td>
<td></td>
</tr>
<tr>
<td>Age, mean ± sd</td>
<td>62.857 ± 8.7949</td>
<td>69.346 ± 9.6039</td>
<td>0.003</td>
</tr>
<tr>
<td>Advantage side, n (%)</td>
<td></td>
<td></td>
<td>0.502</td>
</tr>
<tr>
<td>Yes</td>
<td>34 (38.2%)</td>
<td>12 (13.5%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>29 (32.6%)</td>
<td>14 (15.7%)</td>
<td></td>
</tr>
<tr>
<td>R1, mean ± sd</td>
<td>139.95 ± 9.3076</td>
<td>145.59 ± 10.432</td>
<td>0.014</td>
</tr>
<tr>
<td>R2, mean ± sd</td>
<td>137.97 ± 9.923</td>
<td>149.96 ± 8.2358</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>R3, mean ± sd</td>
<td>145.23 ± 7.9256</td>
<td>160.23 ± 9.7899</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>R4, mean ± sd</td>
<td>149.27 ± 10.666</td>
<td>157.16 ± 9.7511</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>R5, mean ± sd</td>
<td>190.49 ± 11.08</td>
<td>206 ± 12.285</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>R6, median (IQR)</td>
<td>198.23 (192.97, 204.39)</td>
<td>212.6 (204.03, 220.04)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>BMD, median (IQR)</td>
<td>-3.37 (-3.945, -2.915)</td>
<td>-2.85 (-3.1, -2.7)</td>
<td>0.009</td>
</tr>
<tr>
<td>AM, n (%)</td>
<td></td>
<td></td>
<td>0.049</td>
</tr>
<tr>
<td>1</td>
<td>30 (33.7%)</td>
<td>19 (21.3%)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>15 (16.9%)</td>
<td>4 (4.5%)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>18 (20.2%)</td>
<td>3 (3.4%)</td>
<td></td>
</tr>
</tbody>
</table>

Y. Zhao et al.
Fig. 3. Construction and evaluation of RF model 1. Note. ROC curves using 2x cross validation on the (A-C) (training set (A), validation set (B), and test set (C)). (D) machine learning curve. (E) calibration chart. The (F)-decision curve analysis chart shows the net benefit and threshold probability of the decision based on the model output. Summary chart and bar chart of (G-I) feature importance SHAP. The left plot on (G) represents the contribution direction of each value of each variable, with red indicating larger values and blue indicating lower values. The bar chart on the right side of (H) represents the importance of variables and their overall contribution to model predictions. (I). The SHAP score explains the ability to predict the risk of RCT in groups OPRCT and OPO.
were also evaluated. We have chosen the most reliable model.

2.6. Development of the Nomogram

After deriving several independent variables with more accurate predictive efficacy for diagnosing RCTs based on the above machine learning, column line plots as well as associated DCA curves and calibration curves are plotted.

2.7. Statistical Analysis

Continuous variables were represented as mean ± standard deviation or median ± interquartile range and analyzed using unpaired t-tests or Mann-Whitney U-tests. The categorical variables were described as absolute numbers (n) and relative frequencies (%) and analyzed using chi-square or Fisher’s exact test. Spearman correlation analysis was used to analyze the correlations between variables. SPSS Statistics for Windows (Version 25, IBM Corp., Armonk, NY, USA) was used for all statistical analyses. The difference with p<0.05 was considered statistically significant.

3. Results

3.1. Patients’ Characteristics

Table 1 presents the baseline characteristics between 63 OPRCT patients and 26 OPO patients. Compared to OPO patients, OPRCT patients are younger. There is no difference in gender and advantage measures. The data (R1-6, BMD, and AM) differed between the two groups. Nine variables, including Age, R1, R2, R3, R4, R5, R6, BMD, and AM, were used in the prediction models.

3.2. Correlation between R1-6 and BMD

The Pearson correlation test was utilized to examine the correlation between R1-6 and BMD. This method determined the degree of linear correlation between two continuous variables. Correlation scatter plots were plotted to visualize the relationship among them (Fig. 3). The analysis results indicated a positive correlation between R1-6 and BMD. All Pearson correlation coefficients (R-values) were positive, ranging from 0.177 to 0.228, suggesting that as the values of R1-6 increased, BMD also tended to increase. Additionally, all correlations had p-values less than 0.05, indicating statistical significance and strongly supporting our hypothesis. These results highlighted the potential of R1-6 as a valid biomarker for assessing local BMD changes.

3.3. Machine Learning Models for Diagnosis of RCT in OP Patients

The purpose of these models is to diagnose whether OP patients have RCT based on the 9 variables between OPRCT and OPO groups. Nine variables, including Age, R1, R2, R3, R4, R5, R6, BMD, and AM, were used in the prediction models.

3.3.1. Selection of Models

The summary of training/validation/testing set results are shown in Tables 2-4. Based on the results of the training, validation, and testing
sets obtained from five different algorithms, the RF model was the optimal model, with AUCs of 0.998, 0.889, and 0.95 in the training, validation, and testing sets, respectively.

3.3.2. Model Optimization and Interpretation

The RF model was chosen as optimal, and its parameters were automatically adjusted to enhance its performance. We employed the nine variables as input features, enabling the model to achieve optimal performance with AUC scores of 0.998 on the training set, 0.889 on the validation set, and 0.95 on the test set, as illustrated in Fig. 4 (A-C).

The learning curve (Fig. 4 (D)) is shown before evaluating the model’s accuracy, indicating that the error difference between the training and validation sets in our model converges with an increase in the number of training samples, indicating that our model is not overfitting. The calibration chart (Fig. 4 (E)) evaluates the model’s accuracy. This indicates that the RF model exhibits excellent consistency in predicting the probability of observed patients. Subsequently, we constructed the DCA of the model in our study (Fig. 4 (F)), indicating that the RF model exhibited a risk threshold of <70%.

To explain our model, we further analyzed the results of the RF algorithm using SHAP. This ubiquitous feature importance measure uses game theory-based Shapley values to calculate the commitment of each feature to the model output. Fig. 4 (G) reveals the relationship between high and low eigenvalues and the SHAP values in the dataset. Each point represents the characteristic values and SHAP values of each patient. Fig. 4 (H) displays a bar chart of feature importance evaluated using SHAP values. The impact of each feature on the prediction model is represented as a bar graph of the average absolute SHAP value. The graph shows the ranking of variables based on their contribution to the model. Fig. 4 (I) shows an individual-level segmentation, highlighting how model features significantly influence the final risk prediction score by identifying which feature values are adjusted. These adjustments, either upwards or downwards, directly impact the risk prediction scores for individuals, particularly affecting patients categorized as OPRCT and OPO.

3.4. Predictive Modeling of RCT in Patients with OP using Column

Given that the combined diagnosis of a total of nine independent variables, R1-6, BMD, Age, and AM, had a more accurate predictive efficacy for OP patients suffering from RCT, the study utilized a column to construct a predictive model (Fig. 5(A)). The graph (Fig. 5(B)) displayed predicted values on the horizontal axis and actual values on the vertical axis. The 45° ideal line served as a reference standard, while the ‘apparent’ line reflects the degree of fit between predicted and actual values. The bias-corrected solid line shows the fit between the corrected predicted and actual values. Closer proximity of the Bias-corrected or Apparent line to the Ideal line indicates better consistency between predicted and actual values. The calibration curve revealed the accuracy of the constructed prediction model based on the column. In the decision curve, the DCA curve was located above the two baseline lines "None" and "All”, indicating that the performance of the model is acceptable within this range (Fig. 5(C)).

3.5. Machine Learning Models for Predicting the Severity of RCT

These models aimed to predict the severity of RCT based on Cofield classification among OPRCT patients.

3.5.1. Selection of Models

The optimal model training sets are shown in Tables 5–9. According to the results of five different algorithms, the RF model is optimal, with an accuracy of 86.364% in the training set and 73.684% in the testing set.

3.5.2. Model Optimization and Interpretation

We utilized nine variables as input variables, resulting in a model accuracy of 86.364% on the training set and 73.684% on the test set, as shown in Tables 10-11. The learning curve (Fig. 6 (A)) is shown before evaluating the model’s accuracy, indicating that the error difference between the training and validation sets in our model converges with an increase in the number of training samples, indicating that our model is not overfitting.
Fig. 6 (B, C) represents the SHAP summary charts. Fig. 6 (B) unveils the relationship between high and low feature values alongside the SHAP values within the dataset, with each point representing the feature values and SHAP values for individual patients. Fig. 6 (C) presents a bar chart of feature importance evaluated using SHAP values, where the influence of each feature on the predictive model is depicted as a bar chart of the mean absolute SHAP values, displaying the ranking of variables based on their contribution to the model.
Table 5
Logistic Regression Model Training Set Analysis Table.

<table>
<thead>
<tr>
<th>name</th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.2</td>
<td>0.286</td>
<td>10.0</td>
</tr>
<tr>
<td>2</td>
<td>0.385</td>
<td>0.714</td>
<td>0.5</td>
<td>7.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>accuracy</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
</tr>
<tr>
<td>macro avg</td>
<td>0.295</td>
<td>0.305</td>
<td>0.262</td>
<td>19.0</td>
</tr>
<tr>
<td>weighted avg</td>
<td>0.405</td>
<td>0.368</td>
<td>0.335</td>
<td>19.0</td>
</tr>
</tbody>
</table>

Table 6
RF Model Training Set Analysis Table.

<table>
<thead>
<tr>
<th>name</th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.889</td>
<td>0.8</td>
<td>0.842</td>
<td>10.0</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
<td>0.857</td>
<td>0.706</td>
<td>7.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>accuracy</td>
<td>0.737</td>
<td>0.737</td>
<td>0.737</td>
<td>0.737</td>
</tr>
<tr>
<td>macro avg</td>
<td>0.496</td>
<td>0.552</td>
<td>0.516</td>
<td>19.0</td>
</tr>
<tr>
<td>weighted avg</td>
<td>0.689</td>
<td>0.737</td>
<td>0.703</td>
<td>19.0</td>
</tr>
</tbody>
</table>

Table 7
Decision Tree Model Training Set Analysis Table.

<table>
<thead>
<tr>
<th>name</th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.875</td>
<td>0.7</td>
<td>0.778</td>
<td>10.0</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>0.714</td>
<td>0.588</td>
<td>7.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>accuracy</td>
<td>0.632</td>
<td>0.632</td>
<td>0.632</td>
<td>0.632</td>
</tr>
<tr>
<td>macro avg</td>
<td>0.458</td>
<td>0.471</td>
<td>0.455</td>
<td>19.0</td>
</tr>
<tr>
<td>weighted avg</td>
<td>0.645</td>
<td>0.632</td>
<td>0.626</td>
<td>19.0</td>
</tr>
</tbody>
</table>

Table 8
MLP Model Training Set Analysis Table.

<table>
<thead>
<tr>
<th>name</th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>10.0</td>
</tr>
<tr>
<td>2</td>
<td>0.368</td>
<td>1.0</td>
<td>0.538</td>
<td>7.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>accuracy</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
</tr>
<tr>
<td>macro avg</td>
<td>0.123</td>
<td>0.333</td>
<td>0.175</td>
<td>19.0</td>
</tr>
<tr>
<td>weighted avg</td>
<td>0.136</td>
<td>0.368</td>
<td>0.198</td>
<td>19.0</td>
</tr>
</tbody>
</table>

Table 9
SVM Model Training Set Analysis Table.

<table>
<thead>
<tr>
<th>name</th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>10.0</td>
</tr>
<tr>
<td>2</td>
<td>0.444</td>
<td>0.571</td>
<td>0.5</td>
<td>7.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>accuracy</td>
<td>0.526</td>
<td>0.526</td>
<td>0.526</td>
<td>0.526</td>
</tr>
<tr>
<td>macro avg</td>
<td>0.348</td>
<td>0.39</td>
<td>0.367</td>
<td>19.0</td>
</tr>
<tr>
<td>weighted avg</td>
<td>0.48</td>
<td>0.526</td>
<td>0.5</td>
<td>19.0</td>
</tr>
</tbody>
</table>

Table 10
Analysis Table of Optimal Model Training Set.

<table>
<thead>
<tr>
<th>name</th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.778</td>
<td>0.737</td>
<td>0.757</td>
<td>19.0</td>
</tr>
<tr>
<td>2</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>20.0</td>
</tr>
<tr>
<td>3</td>
<td>0.667</td>
<td>0.8</td>
<td>0.727</td>
<td>5.0</td>
</tr>
<tr>
<td>accuracy</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>macro avg</td>
<td>0.731</td>
<td>0.762</td>
<td>0.745</td>
<td>44.0</td>
</tr>
<tr>
<td>weighted avg</td>
<td>0.753</td>
<td>0.75</td>
<td>0.75</td>
<td>44.0</td>
</tr>
</tbody>
</table>

Table 11
Analysis Table of Optimal Model Test Set.

<table>
<thead>
<tr>
<th>name</th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>0.6</td>
<td>0.75</td>
<td>10.0</td>
</tr>
<tr>
<td>2</td>
<td>0.538</td>
<td>1.0</td>
<td>0.7</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>accuracy</td>
<td>0.684</td>
<td>0.684</td>
<td>0.684</td>
<td>0.684</td>
</tr>
<tr>
<td>macro avg</td>
<td>0.513</td>
<td>0.533</td>
<td>0.483</td>
<td>19.0</td>
</tr>
<tr>
<td>weighted avg</td>
<td>0.725</td>
<td>0.684</td>
<td>0.653</td>
<td>19.0</td>
</tr>
</tbody>
</table>

4. Discussion

OP patients are prone to RCT, and the presence of RCT may further exacerbate the negative impacts of OP and vice versa. There is an urgent need for early screening and identification of risk factors for RCT in the OP population and proactive preventive measures for them[30,31]. Diagnostic imaging modalities for RCT, including X-ray, ultrasound and MRI were commonly used in clinic[32–34]. However, ultrasound and MRI are significantly scarce, especially for low- and middle-income countries[35], whereas X-ray examinations are readily available and are the preferred method for diagnosing shoulder joint disorders[36]. Therefore, a comprehensive analysis of shoulder X-ray data holds significant implications for the early detection of RCT. We investigated into novel diagnostic models based on the correlation between proximal humerus BMD and RCT, employing ML techniques to enhance the diagnostic efficacy of conventional radiography in ruling out RCT in OP patients.

ML is an artificial intelligence (AI) technique that allows for more flexible predictive models than traditional statistical methods and has been applied to predict clinical outcomes[37,38]. Clinical prediction models based on medical imaging and ML have been widely used [39–42]. Researchers have been actively exploring machine learning-based diagnostic models for RCT using shoulder X-rays. Notably, Lio and colleagues[36] developed a deep learning algorithm for diagnosing RCT and assessing its severity with high accuracy. At the same time, Cho et al.[43] enhanced the diagnosis of RCT types through ML and the use of TensorFlow and KERAS for image processing, achieving superior accuracy compared to human vision. Despite these advancements, these studies primarily focused on the visual quality of the images, neglecting the extraction of key informational features and the assessment of data information quality. They also did not extensively explore the pre-processing of shoulder X-rays or select ROIs in a targeted manner to improve the models’ quality and clinical applicability.

Inspired by the comprehensive work of Yang Li, Xuewei Chao, and colleagues[44–46], our study incorporates advanced image processing techniques and strategic ROI selection to enhance data quality and extract pivotal informational features.

Our study presented ML models based on AP shoulder X-rays for diagnosing RCT and predicting the type of RCT in OP patients. By analyzing the grayscale values in these X-rays, our models estimated the bone density variations in the proximal humerus of OP patients. We successfully demonstrated the effectiveness and accuracy of ML models in diagnosing RCT in OP patients. In particular, the RF model shows extremely high predictive performance in the training, validation, and test sets with AUC values of 0.998, 0.889, and 0.95, respectively. Among the above results indicated that these models are highly accurate and reliable in predicting the presence of RCT and the type of RCT in OP patients. The significance of our study lay in its contributions to the early diagnosis of RCT and in understanding its relationship with OP. Given that RCT and OP are common conditions among the older adults and there is a significant correlation between the two, early diagnosis is essential to prevent recurrence after surgery and improve patient prognosis.
an effective method for early diagnosis and type identification of RCT in OP patients in the real line with low-cost, simple equipment. Additionally, our results revealed the key variables that affected the prediction of RCT risk through a detailed analysis of the characteristics’ significance. Using SHAP value analysis, we determined which features had the most significant impact on the model’s predictive results.

Among the models for diagnosing RCT in patients with OP, the factors that most influence the risk of developing RCT are the grayscale values of the ROIs set at the GTH, LTH, the humeral head, and the surgical neck of the humerus (R1-4). The tendons of the rotator cuff muscles each attach to GTH and LTH. Due to anatomical characteristics among other factors, RCT generally occurs in the supraspinatus tendon (which attaches to GTH), followed by the infraspinatus tendon (which attaches to LTH). Typically, tears begin in the anterior portion of the supraspinatus, below the acromion, and near the GTH, and gradually extend to the other tendons[7]. In contrast to the more studied site of the GTH, bony changes in other sites such as the LYH are easily overlooked in related studies. They were also included in the ROIs, and the results proved that they also have a significant impact on the occurrence of RCT, which need to be paid attention to.

Among the models for predicting the severity of RCT in patients with OP, AM had the most significant impact on prediction. Consistent with previous studies, there was a direct correlation between different AM and RCT. Among the three types of acromion, type III acromion exerted the most pressure and friction on the rotator cuff ligament, and was the most prone to subacromial impingement and RCT[29]. Our results showed that the greyscale value of the GTH was second only to AM in influencing the severity of RCT. Cadet et al[11]. showed that the lower the greyscale value of the GTH, the more severe the RCT retraction. What’s more, osteoporotic changes in the GTH itself may also affect the severity of rotator cuff dysfunction. In the study of human shoulder specimens, Jiang et al. [47] found that specimens with partial or complete tears had significant trabecular bone loss at the GTH compared to structurally intact rotator cuff specimens. The degree of bone loss also depended on the extent of the RCT. Bone loss was more severe in total tears compared to specimens with partial tears. Meyer et al. [48] found that shoulder specimens with a full RCT had much lower bone density than specimens without a full RCT at GTH. This suggests that the sclerotin of the GTH is strongly associated with the type of RCT. In addition, assessment of the bone density of GTH is critical to the success or failure of shoulder arthroscopy and to postoperative rehabilitation and prognostic outcomes. In arthroscopic rotator cuff repair, anchors to fix torn rotator cuff tendons are inserted mainly in the GTH. The bone quality of GTH plays an important role in postoperative retear and

Fig. 6. Construction and evaluation of RF model 2. Note. (A) machine learning curve. (B-C) feature importance SHAP summary chart and bar chart. The left dot graph represents the contribution direction of each value of each variable, with red indicating larger values and blue indicating lower values. The bar chart on the right side of (E) represents the importance of variables and their overall contribution to model predictions.
anchor loosening[49]. Applying the ML model to the early diagnosis and type identification of RCT in OP patients enables clinicians to more accurately identify those suffering from RCT. In addition, physicians can intervene promptly to take preventive measures to reduce the further development of RCT and develop personalized treatment plans for the extent of the injury, thereby avoiding complex surgical interventions and long-term rehabilitation processes.

5. Limitations

Although our results are encouraging, several limitations warrant acknowledgment and deeper discussion. Firstly, the study sample was limited by its small size and single center origin, which may restrict the generalizability of the findings. To improve the reliability and applicability of the model, future studies should consider using multicenter data and increasing the sample size. Secondly, although our model demonstrated high accuracy, further validation and optimization are required to adapt to different patient populations and clinical environments before it can be widely used. This is crucial, as single-center studies often fail to represent the broader demographic and clinical variations found in general populations. Moreover, purely data-driven AI methods suffer from high data acquisition costs, poor interpretability, and susceptibility to noise interference[50–52]. These issues can significantly impact the practical deployment of AI solutions in healthcare. The high cost of data acquisition limits the availability of large, diverse datasets, which are essential for developing robust AI models. Poor interpretability of AI models can hinder their acceptance by healthcare professionals who may be reluctant to rely on ‘black-box’ solutions. Furthermore, the susceptibility of these models to noise interference necessitates the development of algorithms that are more robust and capable of handling real-world data variations. Addressing these challenges is crucial for the successful integration of AI technologies in clinical settings.

6. Conclusion

In conclusion, our study successfully applied the RF algorithm, a generalizability model, to predict the occurrence and severity of RCT in OP patients. It used greyscale values from specific ROIs and clinical data, achieving high AUC scores and accuracy. This method demonstrates the potential of integrating conventional shoulder X-rays with ML to create a cost-effective, accessible, and non-invasive approach for early RCT detection and management in OP patients, offering clinicians a personalized diagnostic tool. Future research should expand the sample size, validate the models’ clinical use, and explore additional diagnostic tools to enhance early RCT detection in OP patients.

Funding

This work was funded and supported by the Innovation Foundation of the Third Affiliated Hospital of Guangzhou University of Chinese Medicine (No. SY2021005).

Ethics Approval and Consent to Participate

The study involving human participants was approved by the Medical Ethics Committee of the Third Affiliated Hospital of Guangzhou University of Chinese Medicine (No. PJ-KY-202010401-008). Patients included in the study provided written consent to participate after being fully informed of the study’s requirements.

CRediT authorship contribution statement

Yu Zhao: Conceptualization, Data curation, Formal analysis, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing, Funding acquisition, Project administration, Resources. Jingjing Qiu: Data curation, Formal analysis, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. Yang Li: Data curation, Formal analysis, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. Muhammad Attique Khan: Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. Lei Wan: Conceptualization, Investigation, Methodology, Project administration, Writing – review & editing. Lihua Chen: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

The authors will provide the raw data without reservation supporting the conclusions of this study.

Acknowledgments

The authors would like to express our sincere gratitude to Sports Medicine, Osteoporosis, Medical Informatics, and Imaging Departments of the Third Affiliated Hospital of Guangzhou University of Chinese Medicine, as well as to all contributors who contributed to this article.

References

csm.2022.08.001, Journal Article; Review.
Bigliani LU, Ticker JB, Flatow EL, et al. The relationship of acromial architecture to
Chen X, Giambini H, Ben-Abraham E, et al. Effect of Bone Mineral Density on
Qaseem A, Hicks LA, Etxeandia-Ikobaltzeta I, et al. Pharmacologic Treatment of
Tingart MJ, Bouzouzn Ml, Zurkowski D, et al. Three-dimensional distribution of
Oh JH, Song BW, Lee YS. Measurement of volumetric bone mineral density in
proximal humerus using quantitative computed tomography in patients with
https://doi.org/10.1038/s41596-023-03859-z. Journal Article.
Park JY, Kim MH. Changes in bone mineral density of the proximal humerus in
Mall NA, Lee AS, Chahal J, et al. An evidenced-based examination of the
epidemiology and outcomes of traumatic rotator cuff tears. Arthroscopy 2013;29:
Matharu K, Mittal SK, Ashok KS, et al. Selectivity enhancement of Arsenazo(III)
reagent towards heavier lanthanides using polyaminocarboxylic acid: a
Li Y, Ercidi S. Explainable human-in-the-loop healthcare image information quality
Jabben K, Khan MA, Alhassoni M, et al. Breast Cancer Classification from
Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion.
Bayesian Optimization and Fusion-Assisted Optimal Deep Features for COVID-19
Sekeroglu B, Ozashin I. Detection of COVID-19 from Chest X-Ray Images Using
images based on a developed deep neural network. Slas Technol 2022;27:63–75.
images and machine learning techniques.Acm International Conference Proceeding
10.1145/3497623.3497661.
Li Y, Chao X. Distance-Entropy: An Effective Indicator for Selecting Informative
Journal Article.
Li Y, Chao X, Ercidi S. Disturbed-entropy: a simple data quality assessment
Introductory Journal Article.
Chao X, Giambini H, Ben-Abraham E, et al. Effect of Bone Mineral Density on
https://doi.org/10.1371/journal.pone.013934. Journal Article; Research
Support, N.I.H., Extramural.
Bigliani LU, Ticker JB, Flateau EL, et al. The relationship of acromial architecture to
Support, Non-U.S. Gov’t; Review.
clinician needs to know. Radiographics 2006;26:1045–65. https://doi.org/
10.1148/rg.265055087. Journal Article; Review.
Minagawa H, Yamamoto N, Abe H, et al. Prevalence of symptomatic and
asymptomatic rotator cuff tears in the general population: From mass-screening in
Journal Article.
Chin K, Chowdhury A, Levidiotou D, et al. The accuracy of plain radiographs in