Unsupervised Machine Learning-Based Analysis of Clinical Features, Bone Mineral Density Features and Medical Care Costs of Rotator Cuff Tears

Tong-Fu Wang
De-Sheng Chen
Jia-wang Zhu
Bo Zhu
Zeng-Liang Wang
Jian-Gang Cao
Cai-Hong Feng
Jun-Wei Zhao

Department of Sports Medicine and Arthroscopy, Tianjin Hospital of Tianjin University, Tianjin, People’s Republic of China

*These authors contributed equally to this work

Purpose: We aim to present unsupervised machine learning-based analysis of clinical features, bone mineral density (BMD) features, and medical care costs of Rotator cuff tears (RCT).

Patients and Methods: Fifty-three patients with RCT were reviewed, the clinical features, BMD features, and medical care costs were collected and analyzed by descriptive statistics. Furtherly, unsupervised machine learning (UML) algorithm was used for dimensionality reduction and cluster analysis of the RCT data.

Results: There were 26 males and 27 females. The patients were divided into four subgroups using the UML algorithm. There were significant differences among four subgroups regarding trauma exposure, full-thickness supraspinatus tendon tears, infraspinatus tendon tear, subscapularis tendon tear, BMD distribution, medial row anchors, lateral row anchors, total medical care costs, and consumables costs. We observed the highest frequency of trauma exposure, infraspinatus tendon tear, subscapularis tendon tear, osteoporosis, the highest number of medial row anchors, lateral row anchors, total medical care costs, and consumables costs in subgroup II.

Conclusion: The unsupervised machine learning-based analysis of RCT can provide clinically meaningful classification, which shows good interpretability and contribute to a better understanding of RCT. The significance of the results is limited due to the small number of samples, a larger follow-up study is needed to confirm the encouraging results.

Keywords: unsupervised machine learning, clinical features, bone mineral density, medical care costs, rotator cuff tears

Introduction
Rotator cuff tears (RCT) are a common cause of shoulder pain. The prevalence of RCT is 34% in the general population and 54% in people over 60 years old. There are more than 4.5 million hospital visits for RCT in the United States each year.

The demographic and professional features of RCT patients have been investigated. In a study conducted by Yamamoto et al, shoulder ultrasound in 683 patients was analyzed, they found that the incidence of RCT increased with age, the risk factors for RCT included age, history of trauma, and the dominant arm. In the study conducted by Josserand et al, 254 patients with RCT were reviewed, they found that 69% of patients had right-side RCT, 6.1% patients involved in non-manual work, 25.5% patients involved in manual labor, and 68.3% patients involved in heavy manual labor.
Although remarkable achievements in RCT have been obtained, some questions remain to be answered. Rotator cuff tendon-bone healing is an essential component for RCT therapy, however, few studies have investigated the relationship between bone mineral density (BMD) and RCT. Meanwhile, there are few studies that involved medical care costs of RCT. Furthermore, in the era of big data, the requirements for efficient and intelligent processing of massive RCT data are continuously increasing.

An approach for big data analysis is machine learning (ML). ML is commonly divided into supervised machine learning and unsupervised machine learning (UML). In supervised machine learning algorithm, training datasets are labeled, and the algorithm learns from the labeled dataset and predicts the unlabeled dataset. In contrast, the UML algorithm does not require a labeled dataset and tries to work on its own to find information. Currently, more and more studies focus on the application of the UML algorithm in healthcare big data analysis.

In this study, we aim to present conventional descriptive analysis and novel unsupervised machine learning-based analysis of clinical features, BMD features, and medical care costs of RCT.

Materials and Methods

Participants

53 patients with RCT were reviewed. The inclusion criteria were: (1) diagnosis was confirmed by arthroscopy; (2) first surgery for RCT; (3) retrospective study. The exclusion criteria were: (1) simultaneous bilateral shoulder surgery; (2) revision surgery; (3) follow-up data missing. The study was approved by the ethical committee of Tianjin Hospital of Tianjin University and followed the Declaration of Helsinki, the informed consent was obtained from all patients.

Data Collection

16 clinical features including gender, age, marriage, trauma exposure were collected. 18 BMD features including BMD values, T-score, and Z-score of the lumbar spine were collected. 6 medical care costs indicators including total medical care costs, consumables costs, and surgical costs were collected for statistical analysis.

UML Algorithm

The UML algorithm was performed on preprocessed input dataset employing principal component analysis (PCA). The input dataset was defined as a collection of clinical features, BMD features, and medical care costs indicators. The PCA was performed to project the input dataset into a two-dimensional data space. Then, the UML clustering algorithm was applied to identify subgroups, in which the K-means clustering algorithm was used and the best value of k was determined automatically. First, we use the UML algorithm which has been validated and extensively tested in other diseases to convert the input dataset into a compact representation space.

Statistical Analysis

The UML algorithms and the statistical analysis were implemented in Python (Python Software Foundation). Descriptive analysis was reported as mean ± SD and frequency. Comparisons among sub-groups were analyzed by chi-square test or Fisher exact test for categorical variables, and ANOVA for continuous variables. A two-sided P value < 0.05 was considered significant.

Results

Descriptive Analysis of RCT

There were 26 males and 27 females with a mean age of 55.5 ± 8.6 years. The clinical features of RCT are shown in Table 1 and Figure 1.

The mean BMD values of the lumbar spine were 1.01 ± 0.19 g/cm2, T-score was −0.59 ± 1.69, and Z-score was 0.30 ± 1.61. The T-score of the proximal femur was −0.59 ± 1.69 and the hip T-score was −0.41 ± 0.96. There were 23 patients (43.40%) with osteopenia and 4 patients (7.55%) with osteoporosis. The BMD features of RCT were shown in Figure 3.

The total medical care costs were 58,612.60 ± 9,823.01 yuan, the consumables costs were 38,557.87 ± 9,253.44 yuan (52.69%), the surgical costs were 7,278.49 ± 227.09 yuan (9.95%), the drug costs were 3,990.47 ± 812.69 yuan (5.45%), the costs of medical service were 2,330.49 ± 90.69 yuan (9.95%), and the other costs were 21,012.25 ± 5,438.74 yuan (28.72%). The medical care costs of RCT are shown in Figure 4.

Table 1 5 Clinical Features of Patients

| Clinical Features             | Value         |
|------------------------------|---------------|
| Age (years)                  | 55.47 ± 8.6   |
| Duration of symptoms (months)| 9.86 ± 22.82  |
| Medial row anchors           | 2.13 ± 0.81   |
| Lateral row anchors          | 1.51 ± 0.54   |
| Length of hospitalization (days) | 7.13 ± 1.69 |
Clustering results of UML Algorithm

The input dataset was divided into four subgroups via the UML algorithm. As shown in Figure 5, there were 13 patients in subgroup I, 8 patients in subgroup II, 24 patients in subgroup III, and 8 patients in subgroup IV.

Comparisons Among Subgroups

The mean age was 52.0 ± 8.3 years in subgroup I, 55.8 ± 9.8 years in subgroup II, 56.8 ± 9.1 years in subgroup III, and 57.0 ± 5.2 years in subgroup IV. There was no significant difference in age among the four subgroups (P = 0.412).

There were 4 males (30.8%) in subgroup I, 7 males (87.5%) in subgroup II, 12 males (50.0%) in subgroup III, and 3 males (37.5%) in subgroup IV. There was no significant difference in gender among four subgroups (P = 0.074).

As shown in Figure 6, there were significant differences among four subgroups regarding trauma exposure (P = 0.035), full-thickness (FT) supraspinatus (SSP) tendon tears (P = 0.015), infraspinatus (ISP) tendon tear (P < 0.001), and subscapularis (SCP) tendon tears (P = 0.049).

As shown in Figure 7, there were significant differences among four subgroups in BMD distribution classified as normal bone mass, osteopenia, and osteoporosis (P = 0.044).

As shown in Table 2, there were significant differences among four subgroups regarding medial row anchors (P < 0.001), lateral row anchors (P < 0.001), total medical care costs (P < 0.001), and consumables costs (P < 0.001). There were no significant differences among sub-groups for other indicators.

Characteristics of Subgroups

We observed the highest frequency of trauma exposure, ISP ten-don tear, SCP tendon tear, osteoporosis in subgroup II, as well as, we observed the highest number of medial row anchors, lateral row anchors, total medical care costs, and consumables costs.
costs, and consumables costs in subgroup II. Correspondingly, subgroup IV showed the lowest frequency of trauma exposure, FT-SSP tendon tears, the lowest number of medial row anchors, lateral row anchors, total medical care costs, consumables costs, and the highest frequency of osteopenia. There were no ISP ten-don tear, SCP tendon tear, and osteoporosis in subgroup I and sub-group IV (Figure 8).

Discussion

Rotator cuff tears (RCT) are a common cause of shoulder pain.1–3 Yamaguchi et al pointed out the mean age was 58.7 years for unilateral RCT and 67.8 years for bilateral RCT.8 In a study conducted by Andrew et al, 6653 individuals were reviewed, hypertension was involved in 48.3% of patients, diabetes mellitus was involved in 20.1% of patients, and dominant shoulder was involved in 50.4% of patients.29 In our study, the mean age of RCT patients was 55.5 ± 8.6 years, hypertension was involved in 32.08% of patients, diabetes mellitus was involved in 5.66% of patients, and dominant shoulder was involved in 58.49% of patients.

RCT occurs as a result of a broken tendon-bone interface, which suggests that rotator cuff tendon-bone healing is an essential component for RCT therapy.30,31 BMD is an important factor linked with rotator cuff tendon-bone healing, recent studies have shown that the failure rate of rotator cuff healing was 7.25 times in patients with osteoporosis than patients with normal bone mass.9 In our study, osteoporosis was involved in 7.55% of patients and all patients with osteoporosis were accompanied by FT-SSP tendon tears.

In the costs analysis performed by Savoie et al, 50 patients with successful rotator cuff repair surgery were reviewed, the mean medical care costs were $50,302.25 if the patients were referred to a specialist immediately following the diagnosis of RCT, the mean medical care costs were $25,870.64.12 In our study, the total medical care costs were 58,612.60 ± 9,823.01 yuan, the consumables costs were 38,557.87 ± 9,253.44 yuan with the highest proportion.

Recent advances in UML have shown that it can help physicians to analyze massive information and is crucial in medical practice optimization.32–35
Figure 3 The T-score of patients.
Abbreviation: PF, proximal femur.

Figure 4 The distribution of medical care costs. The consumables costs contributed the highest proportion (52.69%).
Figure 5 The input dataset was divided into four subgroups. (A) Dimensionality reduction of the input dataset into a two-dimensional data space using principal component analysis. (B) The k-means clustering algorithm was used to divide the dataset into 4 subgroups, with red dots representing subgroup I (n=13), blue dots representing subgroup II (n=8), green dots representing subgroup III (n=24), and yellow dots representing subgroup IV (n=8).

Figure 6 There were significant differences among four subgroups regarding trauma exposure, FT-SSP, ISP, SCP.
Abbreviations: FT-SSP, full-thickness supraspinatus tendon tears; ISP, infraspinatus tendon tear; SCP, subscapularis tendon tears.
study of Baud et al, UML can provide a powerful approach for the automatic identification of recurrent pathological neurophysiological signals, which is important for the accurate, quantitative, and individualized evaluation of focal epilepsy. In a study conducted by Yang et al, the UML algorithm was used to reveal phenotypically distinct cell populations and determines whether these populations stratify patient survival with glioblastoma. In the study conducted by Cikes et al, 1106 patients with heart failure were reviewed, UML can provide a clinically meaningful classification and may help optimize response rates to specific treatments.

In our study, 53 patients were reviewed and up to 40 indicators per patient were collected. Finally, the input dataset was divided into four subgroups. Subgroup II had the highest frequency of osteoporosis, ISP tendon tear, and SCP tendon tear, which indicated that decreased BMD downregulated the union strength of tendon-bone interface and decreased the threshold of ISP tendon tear and SCP tendon tear. Meanwhile, subgroup II had the highest number of medial row anchors and lateral row anchors, which could be due to ISP tendon tear and SCP tendon tear increased the area requiring repair, or due to more anchors were required to maintain the union strength of tendon-bone interface which downregulated by decreased BMD.

**Table 2** Comparisons Among Subgroups Regarding Medial Row Anchors, Lateral Row Anchors, Total Medical Care Costs, Consumables Costs

| Group    | Medial Row Anchors | Lateral Row Anchors | Total Medical Care Costs (Yuan) | Consumables Costs (Yuan) |
|----------|--------------------|---------------------|---------------------------------|--------------------------|
| Subgroup I | 1.92 ±0.28        | 1.08±0.28           | 5.16±1.51                      | 3.22±0.10                |
| Subgroup II | 3.25 ±1.04        | 2.13±0.35           | 7.63±0.52                      | 5.53 ±0.39               |
| Subgroup III | 2.25 ±0.44        | 1.71±0.46           | 6.09±0.24                      | 4.07 ±0.22               |
| Subgroup IV | 1.00 ±0.00        | 1.00±0.00           | 4.53±0.14                      | 2.57±0.17                |
| F value   | 26.73              | 20.52               | 212.02                          | 273.64                   |
| P value   | < 0.001            | < 0.001             | < 0.001                         | < 0.001                  |

**Figure 7** There were significant differences among four subgroups in BMD distribution.

**Abbreviation:** Nor, normal.
Interestingly, Subgroup IV had the highest frequency of osteopenia, but there were no ISP tendon tear and SCP tendon tear in all patients, which indicated that the decreased BMD taking osteopenia as the lower limit did not reach the threshold of ISP and SCP tendon tear. Meanwhile, subgroup IV had the lowest number of medial row anchors and lateral row anchors, which indicated that the decreased BMD taking osteopenia as the lower limit did not downregulate the union strength of the tendon-bone interface.

This study has some limitations: (1) It is a small retrospectively study; (2) There are few indicators in this study, furthermore indicators were needed for a systematic and comprehensive assessment of RCT; (3) The research is a single-center study, the results may be biased and further larger, multicenter research was needed to confirm our findings.

Figure 8 Summary of clinical features, BMD features, and medical care costs among four subgroups.
Abbreviations: FT-SSP, full-thickness supraspinatus tendon tears; ISP, infraspinatus tendon tear; SCP, subscapularis tendon tears; MRA, medial row anchors; LRA, lateral row anchors.

Conclusion
The unsupervised machine learning-based analysis of RCT can provide clinically meaningful classification, which shows good interpretability and contribute to a better understanding of RCT. The significance of the results is limited due to the small number of samples, a larger follow-up study is needed to confirm the encouraging results.

Author Contributions
All authors contributed to data analysis, drafting or revising the article, gave final approval of the version to be published, agreed to the submitted journal, and agree to be accountable for all aspects of the work.

Disclosure
The authors report no conflicts of interest in this work.
References

1. Toyoda H, Ito Y, Tomo H, et al. Evaluation of rotator cuff tears with magnetic resonance arthrography. Clin Orthop Relat Res. 2005;439:109–115. doi:10.1097/01.blo.0000176142.01262.35

2. Osthsy R, Lindsey Derek P, Giori Nicholas J, et al. Bioabsorbable tricalcium phosphate bone cement strengthens fixation of suture anchors. Clin Orthop Relat Res. 2010;468(12):3406–3412. doi:10.1007/s11999-010-1412-7

3. Millar Neal L, Wu X, Tantau R, et al. Open versus two forms of arthroscopic rotator cuff repair. Clin Orthop Relat Res. 2009;467(4):966–978. doi:10.1097/BLO.0b013e318182fc175

4. Sher JS, Uribe JW, Posada A, et al. Abnormal findings on magnetic resonance images of asymptomatic shoulders. J Bone Joint Surg Am. 1995;77(1):10–15. doi:10.2106/00004623-199501000-00002

5. Oh Luke S, Wolf Brian R, Hall Michael P, et al. Indications for rotator cuff repair: a systematic review. Clin Orthop Relat Res. 2007;455:52–63. doi:10.1097/BLO.0b013e31806f706-5

6. Yamamoto A, Takagishi K, Osawa T, et al. Prevalence and risk factors of a rotator cuff tear in the general population. J Shoulder Elbow Surg. 2010;19(1):116–120. doi:10.1016/j.jse.2009.04.006

7. Nové-Josserand L, Liotard JP, Godeneche A, et al. Occupational outcome after surgery in patients with a rotator cuff tear due to a work-related injury or occupational disease. A series of 262 cases. Orthop Traumatol Surg Res. 2011;97(4):361–366. doi:10.1016/j.otsr.2011.01.012

8. Chung SW, Oh JH, Gong HS, et al. Factors affecting rotator cuff healing after arthroscopic repair: osteoporosis as one of the independent risk factors. Am J Sports Med. 2011;39(10):2099–2107. doi:10.1177/0005144411415659

9. Cancio Jourdan M, Brockmeier Stephen F, Kew Michelle E, et al. The association of osteoporosis and bisphosphonate use with revision shoulder surgery after rotator cuff repair. Arthroscopy. 2016;32(9):1771–1780. doi:10.1016/j.arthro.2016.01.063

10. Christope C, Laurence B, Kunal K, et al. Arthroscopic repair of full-thickness rotator cuff tears: is there tendon healing in patients aged 65 years or older? Arthroscopy. 2010;26(3):302–309. doi:10.1016/j.arthro.2009.08.027

11. Savoie FH, Field LD, Jenkins RN. Costs analysis of successful rotator cuff repair surgery: an outcome study. Comparison of gatekeeper system in surgical patients. Arthroscopy. 1995;11(6):672–676. doi:10.1016/0749-8063(95)90107-8

12. 13. Makhni Eric C, Swart E, Steinhaus Michael E, et al. Cost-effectiveness of reverse total shoulder arthroplasty versus arthroscopic rotator cuff repair for symptomatic large and massive rotator cuff tears. Arthroscopy. 2016;32(9):1771–1780. doi:10.1016/j.arthro.2016.01.063

14. Ghani Khurshid R, Zheng K, Wei John T, et al. Harnessing big data for health care and research: are urologists ready? Eur Urol. 2014;66 (6):975–977. doi:10.1016/j.eururo.2014.07.032

15. Zhongheng Z, Lei M, Chen H, et al. Binary logistic regression modeling with TensorFlow™. Ann Trans Med. 2019;7(20):591. doi:10.21037/atm.2019.09.125

16. Andersson Claes R, Gustafsson Mats G, Helena S. Quantitative chemogenomics: machine-learning models of protein-ligand interaction. Curr Top Med Chem. 2011;11(15):1978–1993. doi:10.2174/1568026117936931249

17. McDermid A, Chen X, Zhang Y, et al. A new machine learning-based framework for mapping uncertainty analysis in ma-seq read alignment and gene expression estimation. Front Genet. 2018;9:313. doi:10.3389/fgene.2018.00313

18. Agany Diing DM, Pietro Jose E, Gnimpiela Etienne Z. Assessment of vector-host-pathogen relationships using data mining and machine learning. Comput Struct Biotechnol J. 2017;2020(18):1704.

19. Camino A, Jia Y, Yu J, Wang J, Liu L, Huang D. Automated detection of shadow artifacts in optical coherence tomography angiography. Biomed Opt Express. 2019;10(3):1514–1531. doi:10.1364/BOE.10.001514

20. Go T, Byeon H, Lee SJ. Label-free sensor for automatic identification of erythrocytes using digital in-line holographic microscopy and machine learning. Biosens Bioelectron. 2018;103:12–18. doi:10.1016/j.bios.2017.12.020

21. Libbrecht Maxwell W, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet. 2015;16(6):321–332. doi:10.1038/nrg3920

22. Beatrui R, Laura H, Geras Krzysztof J, et al. Machine learning in breast MRI. J Mag Reson Imaging. 2020;52(4):998–1018.

23. Thomas B, Robert A, Alexander W, et al. Extracting aerobic system dynamics during unsupervised activities of daily living using wearable sensor machine learning models. J Appl Physiol. 2018;124(2):473–481.

24. Hamadheh L, Imran S, Bencsik M, et al. Machine learning analysis for quantitative discrimination of dried blood droplets. Sci Rep. 2020;10(1):3313. doi:10.1038/s41598-020-59847-x

25. Cikes M, Sanchez-Martinez S-M, Claggett B, et al. Machine-learning-based phenotyping in heart failure to identify responders to cardiac resynchronization therapy. Eur J Heart Fail. 2019;21(1):74–85. doi:10.1002/ejhf.1333

26. Wei W, Bleecker E, Moore W, et al. Unsupervised phenotyping of severe asthma research program participants using expanded lung data. J Allergy Clin Immunol. 2014;133(5):1280–1288. doi.org/10.1016/j.jaci.2013.11.042

27. Li H, Xu C, Xin B, et al. F-FDG PET/CT radiomic analysis with machine learning for identifying bone marrow involvement in the patients with suspected relapsed acute leukemia. Theranostics. 2019;9(16):4730–4739. doi:10.7150/thno.33841

28. Li H, Galperin-Aizenberg M, Pryma D, et al. Unsupervised machine learning of radiomic features for predicting treatment response and overall survival of early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy. Radiother Oncol. 2018;129(2):218–226. doi:10.1016/j.radonc.2018.06.025

29. Arjun SA, Chye ATH. Systematic review on risk factors of rotator cuff tears. J Orthop Surg. 2017;25(1):230949916684318.

30. Cheung EV, Luz S, Sperling JW. Strategies in biologic augmentation of rotator cuff repair: a review. Clin Orthop Relat Res. 2010;468 (6):1476–1484. doi:10.1007/s11999-010-1323-7

31. Chen P, Cui L, Chen G, et al. The application of BMP-12-overexpressing mesenchymal stem cells loaded 3D-printed PLGA scaffolds in rabbit rotator cuff repair. Int J Biol Macromol. 2019;138:79–88. doi:10.1016/j.ijbiomac.2019.07.041

32. Cearns M, Hahn T, Baune BT. Recommendations and future directions for supervised machine learning in psychiatry. Transl Psychiatry. 2019;9(1):271. doi:10.1038/s41398-019-0667-2

33. Nalin L, Justine S, Mistry AM, et al. Unsupervised machine learning reveals risk stratifying glioblastoma tumor cells. eLife. 2020;9: e56879.

34. Saboo KV, Varartharajah Y, Berry BM, et al. Unsupervised machine-learning classification of electrophysiologically active electrodes during human cognitive task performance. Sci Rep. 2019;9(1):17390. doi:10.1038/s41598-019-53925-5

35. Hussain S, Le Guezennec X, Yi W, et al. Digging deep into golgi phenotypic diversity with unsupervised machine learning. Mol Biol Cell. 2017;28(25):3686–3698. doi:10.1091/mbc.e17-06-0379

36. Baud MO, Kleen JK, Anumanchipalli GK, et al. Unsupervised learning of spatiotemporal interictal discharges in focal epilepsy. Neurosurgery. 2018;83(4):683–691. doi:10.1093/neuros/nyx480
