Cukurova Medical Journal
ÇUKUROVA ÜNİVERSİTESİ TIP FAKÜLTESİ

Cukurova Med J 2023;48(2):715-722
DOI: 10.17826/cumj.1281955

RESEARCH

A K-nearest neighbors-based classification approach for automated detection of knee osteoarthritis

Diz osteoartritinin otomatik tespiti için K-en yakın komşuluk algoritmasına dayalı bir sınıflandırma yaklaşımı

Çağlar Cengizler[1], Ayşe Gül Kabakcı[2]

[1]Biomedical Device Technology Program, Vocational School of Health Services, Izmir Democracy University, Izmir, Turkey
[2]Cukurova University Faculty of Medicine Department of Anatomy, Adana, Turkey

Abstract

Purpose: Osteoarthritis is a serious condition that can significantly reduce a person's quality of life, causing pain and stiffness in the knees and limiting their mobility. The condition progressively worsens over time, emphasizing the importance of early diagnosis. This study implemented a computer-aided classification approach to reduce the time and effort required for diagnosing knee osteoarthritis while minimizing human errors.
Materials and Methods: Data analyzed in this study was obtained from the Osteoarthritis Initiative. A total of 165 samples were used in the study. All abnormal samples were graded as severe osteoarthritis. While 78 samples were used to test the implemented algorithm, the training process of the algorithm was completed with 87 samples. The proposed approach involves three main stages: segmenting the cartilage region through a series of image-processing operations, extracting morphological features from the defined region, and classifying samples based on these features. In the classification stage, morphological features characterizing the cartilage region were classified in the observation space, and the k-nearest neighbors algorithm was applied for automated discrimination. Accordingly, the computer utilizes the previously classified sample features to estimate the presence of pathology.
Results: Test classifications were completed with 78 samples; 28 were previously diagnosed with osteoarthritis. Morphological measures of the training samples were accepted as a reference for abnormality. The applied classification scheme can distinguish severed cartilage regions with a 0.95% accuracy.
Conclusion: This study demonstrates the potential effectiveness of a computer-aided approach in diagnosing knee osteoarthritis with high accuracy. The developed

Öz

Amaç: Osteoartrit, kişinin yaşam kalitesini düşüren, dizlerde hissedilen ağrı ve sertlik ile kişinin hareket kabiliyetini kısıtlayabilen ve zamanla şiddetini arttıran ciddi bir rahatsızlıktır. Hastalığın ilerleyici karakteri erken tanının önemini artırmaktadır. Röntgen görüntüleri bu hastalığım teşhisi için klinisyenler tarafından en çok tercih edilen araçlardan biridir. Çalışmada bilgisayar destekli sınıflandırma yaklaşımı ile diz osteoartritinin teşhisi için gereken zaman ve iş gücünün azaltılarak insan kaynaklı hataların minimize edilmesi hedeflenmiştir.
Gereç ve Yöntem: Bu çalışmada analiz edilen veriler, Osteoartrit Girişimi'nden elde edilmiştir. Çalışmada toplamda 165 örnek kullanılmıştır. Tüm anormal örneklerde şiddetli kıkırdak zararı gözlenmiştir. Uygulanan algoritmanın test safhasında 78, eğitim aşamasında ise 87 örnek kullanılmıştır. Önerilen yaklaşım üç ana aşama içermektedir: kıkırdak bölgesinin görüntü işleme yöntemleri ile bölütlenmesi, sınırları çizilen bölgeden morfolojik özelliklerin çıkarılması ve bu özelliklerin değerlendirilerek örneklerin sınıflandırılması. Sınıflandırma aşamasında, gözlem uzayı kıkırdak bölgesini karakterize eden morfolojik özellikler ile oluşturulmuş, otomatik sınıflandırma işlemi için k-en yakın komşu algoritması uygulanmıştır. Buna göre, bilgisayar önceden sınıflandırılmış örnek özelliklerini kullanarak patolojinin varlığını tahmin etmektedir.
Bulgular: Test sınıflandırmaları 28'ine osteoartrit teşhisi konmuş toplam 78 örnekle tamamlanmıştır. Eğitim örneklerinin sayısal özellikleri kıkırdak hasarının otomatik tespiti için referans olarak kabul edilmiştir. Uygulanan sınıflandırma mekanizması yıpranmış kıkırdak bölgelerini %0.95 doğrulukla ayırt edebilmiştir.

Address for Correspondence: Ayşe Gül Kabakcı, Cukurova University Faculty of Medicine Department of Anatomy, Adana, Turkey E-mail; akabakci@cu.edu.tr
Received: 13.04.2023 Accepted: 22.06.2023

approach offers a promising solution for early and efficient diagnosis, enabling more timely and effective treatment strategies for osteoarthritis patients. The progressive nature of the disease makes these advancements in diagnostic methods invaluable. Future studies may focus on expanding the sample size and further refining the model for enhanced precision and broad applicability in clinical settings.
Keywords:. automated; knee; osteoarthritis; KNN; classification

Sonuç: Hastalığın ilerleyici doğası, teşhis yöntemlerindeki ilerlemeleri oldukça değerli kılmaktadır. Sonraki çalışmalar örneklem büyüklüğünün genişletilerek ümit vadeden modelin klinik ortamlarda yüksek doğrulukla uygulanabilirlik kazanması için geliştirilmesine odaklanabilir.
Anahtar kelimeler: Otomatik; diz; osteoartrit; KNN; sınıflama

INTRODUCTION

Knee osteoarthritis is characterized by the gradual breakdown of cartilage in the knee joint, which is one of the most common conditions that affect people worldwide[1]. Usually, it takes 10 to 15 years to develop this disorder and interfere with life[2]. This condition limits mobility while causing pain and stiffness[3]. The risk group of knee osteoarthritis covers certain occupations which require repetitive motions or heavy liftings, such as farming and construction[4]. Treatment options for the disease include both non-pharmacologic and pharmacologic interventions. Diagnosis and treatment of the disease are mainly based on physical examination (relies heavily on subjective experience) and plain radiography[5]. Moreover, early diagnosis is vital in treating the disease and preventing progression. At that point, the computer-aided diagnosis would decrease the time, effort, and human resources required for early diagnosis and intervention. Several studies have investigated the use of computer-aided diagnosis systems for knee osteoarthritis. One study by Kokkotis et al. proposed a fully automated system for accurate identification of the outlines of the bones of the knee joint. They have utilized two random forest classifiers to identify abnormal texture and morphology[6]. Another study by Shamir et al. includes an implementation of a simple weighted nearest neighbor to grade the osteoarthritis existence[7]. Also, Suresha et al. have utilized a deep neural network for severity assessment of the disease[8]. Moreover, Saleem et al. have proposed a computer-aided mechanism for assisting radiologists. They have segmented the knee regions by template matching to calculate the joint space[9]. Convolutional neural networks have also been previously adapted for automatic quantification of the severity of the disorder[10]. Brahim et al. also applied multivariate linear regression to reduce the variability between normal and abnormal cases in one study. They have examined 1024 publicly available samples and presented a random forest-based classifier[11]. Machine learning approaches have also been applied for proactive predictions. In one study, the algorithm implemented by Shamir et al. predicted severity change with 72% accuracy[12]. Previous studies have shown that computer-based rapid and automated diagnosis plays an effective role in minimizing the damage caused by the disease's progressive nature. Implementing a lazy learning approach using the k-nearest neighbors' algorithm can enable the computer to accurately distinguish between damaged and healthy knee cartilage based on morphological features extracted from X-ray images. This novel machine-learning approach is expected to contribute to the automatic diagnosis of the disease by providing an efficient and reliable method for identifying abnormalities in knee cartilage. Accordingly, the cartilage area, which usually covers the ends of bones in the knee joint, is extracted by processing X-ray images. Then the k-nearest neighbors' algorithm is trained to classify morphological features extracted from the cartilage area. For forming a ground truth, all of the samples belonging to test and train groups were previously classified and assigned to a class as normal or abnormal. The proposed system's performance is measured by comparing estimated classes assigned to training samples and actual classes.

MATERIALS AND METHODS

Dataset

The data utilized in this research was obtained from the Osteoarthritis Initiative (OAI), a longitudinal, prospective study conducted in the USA that focuses on knee osteoarthritis (OA) [13]. A committee of radiologists evaluated several knee radiographs as part of the OAI. It should be noted that all medical data used in this study is published online by OAI under the Creative Commons license [14]. A total of 165 bilateral fixed-flexion plain film radiographs were analyzed in the study. Out of these, 78 radiographs were used to evaluate the implemented algorithm, with 28 classified as abnormal. The algorithm was trained using 37 abnormal and 50 normal samples during the training process. Significantly, all the abnormal samples were classified as exhibiting severe sclerosis. Two samples from the data set are shown in Figure 1.

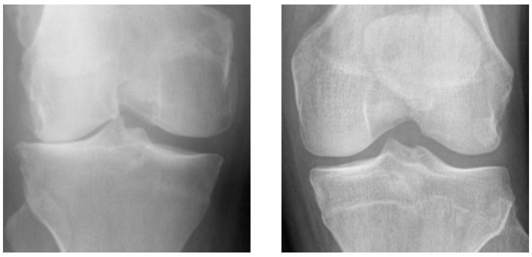

Figure 1. Two samples from the dataset are shown (abnormal on the left, normal on the right)

This study aimed to utilize the learned numerical features from pre-classified examples to make predictions about unknown samples using a computer. At this point, samples that possess the most distinctive features of their class were specifically selected for the training phase. Additionally, the impurity of the dataset would negatively affect the efficiency of the implemented machine-learning model. Therefore, while preserving the data's diversity, artifact-free, not excessively dark, blurry, or noisy examples have been included in the dataset to maintain data purity.

The used data set is collected under the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. OAI has received approval from the Institutional Review Board (IRB) at the University of California, San Francisco (UCSF) and its affiliated institutions (approval number: FWA00000068). Additionally, all individual participants included in the study provided informed consent prior to their involvement.

Procedure

Segmentation of cartilage area

The segmentation stage of the study involves morphological operations to segment cartilage areas between bone regions. The process starts with the extraction of complement images from grayscale samples. Blurring followed extraction, which helps suppress the local high-frequency noise. Then a top-hat filter with a disk-shaped structuring element and thresholding was applied to recover space remaining from bones, including the area of cartilage tissue as binary blobs [15]. Finally, each image was windowed to crop extra blank regions from the cartilage area. It should be noted that the whole process is completely automated, and all parameters were fixed to a specific value. The visual output of each step is shown in Figure 2.

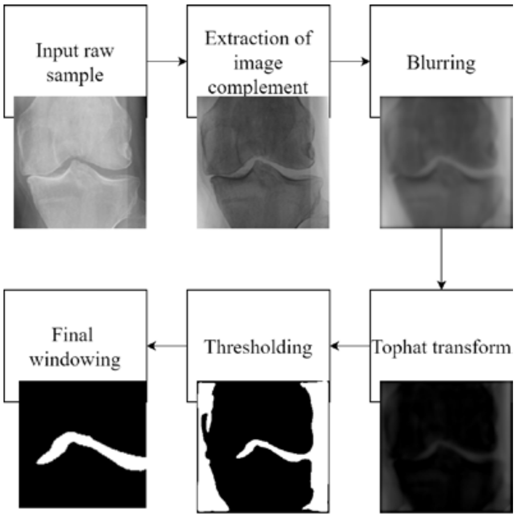

Figure 2. Steps of the segmentation stage is show as block diagram.

Feature extraction

The feature extraction stage aims to measure the best attributes that characterize the healthy cartilage region. It is known that osteoarthritis causes a decrease in cartilage amount [16]. Accordingly, the size and shape of the cartilage region vary in case of the condition's existence. Therefore, three morphology-based features are extracted from automatically defined cartilage regions to classify the shape and amount of variance caused by osteoarthritis. Extracted features are explained below.

Major axis length

It was measured as the length (in pixels) of the longest line segment that can be drawn within the cartilage area.

Minimum feret distance

It was calculated by measuring the smallest distance between two parallel lines perpendicular to the longest axis of the cartilage area.

Perimeter

This feature was based on the distance around the boundary of the cartilage region, which also varies with thickness.

Classification

Each cartilage region investigated in this study was accepted as an observation in search space. Moreover, each observation was represented by shape and morphology characterizing attributes. Accordingly, it was aimed to train a supervised machine learning approach to discriminate morphological deformation with extracted features. The k-nearest neighbors (KNN) algorithm was innovatively implemented as a classifier to accomplish such a task.

Training stage of the KNN algorithm is based on forming search space with previously classified observations. Accordingly, KNN stores all training observations to make predictions at runtime [17,18]. When there is an input of an unknown class to be identified, the algorithm searches for the k number of nearest neighbors to the new input in the search space created by training observations [19]. The distance of neighbors is calculated based on the preferred distance metric [20]. Euclidean distance was preferred in this study. After the k nearest neighbors found, the label of the new entry was decided according to the class of the nearest neighbors [21].

RESULTS

Experimental algorithms were implemented in the MATLAB environment. Test classifications were completed with 78 samples; 28 were previously diagnosed with osteoarthritis. Numerical features of the training samples were accepted as a reference for abnormality. The accordingly Implemented algorithm has divided the test set into two clusters by assigning each sample to an abnormal or healthy class.

Feature distinctiveness and variance in the dataset

In the study, three morphological features that best characterize cartilage damage in X-ray images were preferred for the automated diagnosis of the disease. To demonstrate the discriminative power of the extracted features, the means and standard deviations of the healthy and abnormal classes were calculated for all features. Additionally, an independent t-test was applied to compare the means of the two classes. Cohen's d was then computed to determine the effect size of the difference between the two means (Table 1). Furthermore, Figure 3 displays histograms of the features, illustrating the distribution of the extracted features according to classes.

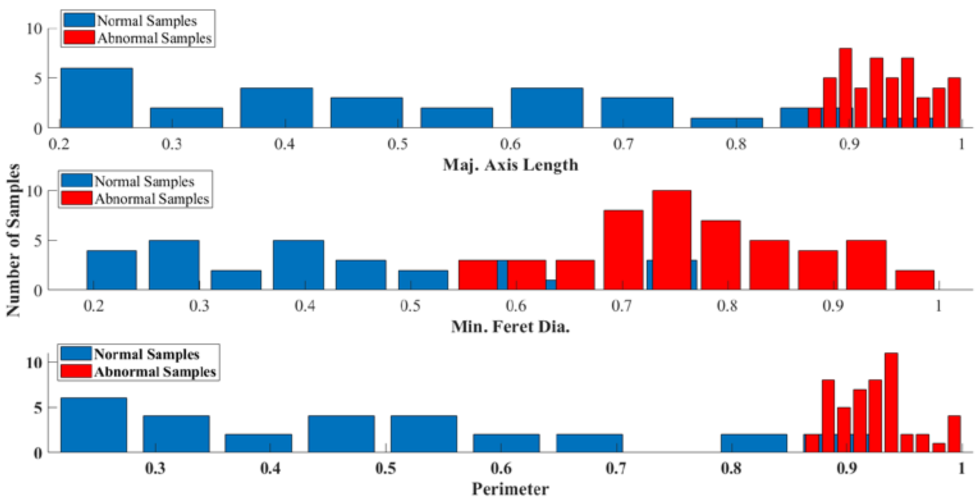

Figure 3. Test data feature distribution is plotted as histogram. Each class is indicated with a different color.

Classification performance

Performance evaluation was based on a comparison of predicted labels and actual labels. The following formulations were used as an objective measure of performance:

$$Precision = \frac{Tp}{Tp+Fp} \quad (1a)$$

$$Recall = \frac{Tp}{Tp+Fn} \quad (1b)$$

$$Specificity = \frac{Tn}{Tn+Fp} \quad (1c)$$

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \quad (1d)$$

$$FScore = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (1e)$$

Where, Tn, Tp, Fp, and Fn stand for true negative, true positive, false positive, and false negative, respectively. Classification performance measures are given in Table 2. In addition to numerical results, visual outputs of predictions are generated for empirical evaluation. Figure 4 showcases the classification performance by visually marking abnormal and normal cartilage regions.

Table 2. Classification performance of the implemented algorithm

F-measure	Accuracy	Sensitivity	Specificity	Precision	Recall
0.923	0.949	0.857	1.000	1.000	0.857

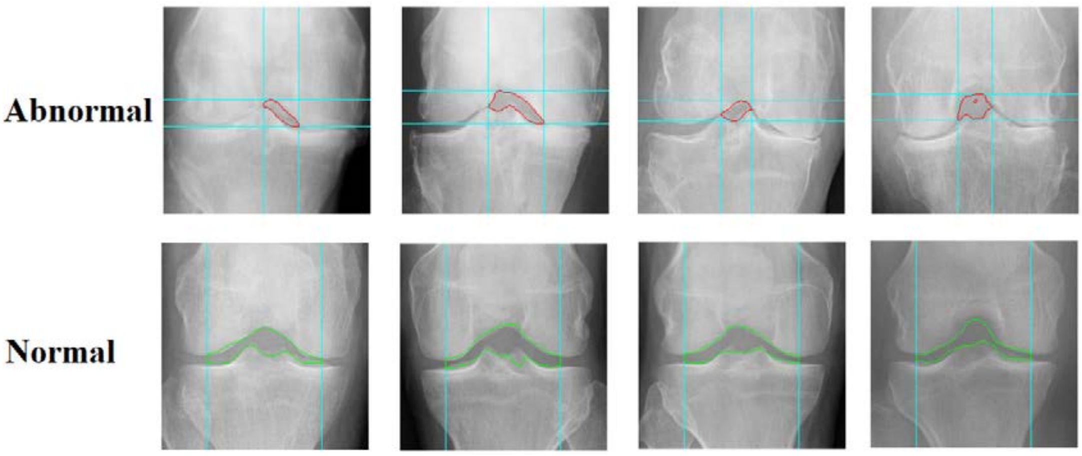

Figure 4. Classified test samples are shown. Segmented cartilage area is seen on each sample. Cartilage region boundaries of abnormal samples are indicated with red.

Table 1. The means and standard deviations of the two classes are presented, along with the p-value and Cohen's d, to demonstrate the variance of features between the two classes.

	Mean		Standard Deviation			
	Abnormal	Healthy	Abnormal	Healthy	Cohen's d	p
MajorAxis Length	0.513	0.930	0.228	0.037	2.555	1.71E-20
Minimum Feret Distance	0.425	0.769	0.170	0.109	2.398	4.60E-17
Perimeter	0.486	0.924	0.213	0.033	2.876	2.76E-23

DISCUSSION

This study presented a machine learning-based classification approach that requires relatively lower processing power, enabling the automated detection of abnormalities in knee cartilage. The applied algorithm learns the characteristics of the previously classified samples in order to distinguish them. Accordingly, the computer was taught the distinctive features that would characterize the abnormality in the cartilage. Statistical measures, including mean and standard deviation (Table 1), were calculated for healthy and abnormal classes for each feature. The comparison of the means of the two classes was achieved using an independent t-test, providing a clear indication of the distinctiveness of these features in differentiating abnormality in the cartilage region.

Additionally, the effect size of the difference between the means was calculated using Cohen's d. This measure gives a number that shows how big the difference is between the healthy and abnormal groups. Bigger numbers mean a larger difference. Figure 3 highlights the notable variations in extracted morphological features associated with Osteoarthritis. Based on this, it may be possible to say that these features can be combined with other machine-learning approaches, further enhancing the accuracy and effectiveness of the diagnosis process.

It should be noted that the innovative approach proposed in the study has been tested with a data set purified from noisy images with low representation ability of their class. One reason for this is to optimize the learning ability of the applied machine learning approach. This situation may require additional methods and algorithms to reduce noise sensitivity before the field use of the proposed mechanism. Bayramoglu et al. have reported that they have developed a region segmentation and ROI selection approach for adaptive segmentation of knee radiographs. Their approach similarly combines machine learning with image processing for robust segmentation. They have reported that their method would be sensitive to varying imaging conditions such as rotations, beam angle, and noise[22]. Olsson et al. have argued that their CNN-based method does not require a preprocessing phase, and no elimination is made in the data set, even though the images in the data sets show high variations[23]. However, the method they propose requires relatively more processing power.

A lazy learning system was adapted in the study for classification, allowing continuous updates with new entries. The results in Table 1 demonstrate the classification scheme's ability to accurately distinguish severe cartilage regions with a precision of 94%. In comparison, Shamir et al. reported a maximum accuracy of 91.5% for their mechanism in classifying test samples[12], and Thomson et al.'s random forest-based classifier reportedly achieved an AUC of 0.849 in discriminating abnormal samples[24].These findings indicate that the novel combination of image processing and machine learning approaches presented in this study is well-suited for automating the diagnosis of the disorder. Moreover, the proposed implementation offers a low computational cost and outperforms previous attempts at classifier adaptations.

However, it is essential to recognize the importance of further research and testing to validate these results and ensure their applicability to larger and more diverse datasets. Furthermore, the visual results presented in this study demonstrate the effectiveness of the applied fully automated unsupervised segmentation method, even in severe cases of Knee osteoarthritis. The segmentation capability of an automated diagnosis system significantly influences the final classification performance. Consequently, incorporating a machine learning-based adaptive segmentation approach would enhance the functional performance of the classifier.

In conclusion, this study showcases the successful adaptation of a lazy learning system for classification, achieving high accuracy in distinguishing severe cartilage regions. The combination of image processing and machine learning approaches proves to be compatible with automating the diagnosis of the disorder. While the presented results and visual findings are encouraging, additional research is necessary to validate their generalizability across diverse datasets.

The k-Nearest Neighbors approach was adapted to classify shape and morphology characterizing features of severed knee cartilage in this study. Accordingly, a fully automated segmentation algorithm was developed, and several morphological measures were extracted from the automatically segmented cartilage region. The obtained classification results are promising, suggesting that the novel combination of image-processing and machine-learning approaches proposed in this study would be effective and efficient for real-life practical applications. Furthermore, as the methodology evolves, its capabilities can be further enhanced by incorporating a more extensive, diverse dataset.

Author Contributions: Concept/Design : ÇC; Data acquisition: ÇC; Data analysis and interpretation: ÇC, AK; Drafting manuscript: ÇC, AK; Critical revision of manuscript: ÇC; Final approval and accountability: ÇC, AK; Technical or material support: ÇC; Supervision: ÇC, AK; Securing funding (if available): n/a.
Ethical Approval: OAI has received approval from the Institutional Review Board (IRB) at the University of California, San Francisco (UCSF) and its affiliated institutions (approval number: FWA00000068).
Peer-review: Externally peer-reviewed.
Conflict of Interest: The author(s) received no financial support for the research, authorship, and/or publication of this article.
Financial Disclosure: The authors declare that they have no conflict of interests.
Availability of data and materials: All samples used in this study are available publicly online at https://data.mendeley.com/datasets/56rmx5bjcr/1

REFERENCES

1. Xing D, Xu Y, Liu Q, Ke Y, Wang B, Li Z et al. Osteoarthritis and all-cause mortality in worldwide populations: grading the evidence from a meta-analysis. Sci Rep. 2016;6:24393.
2. Lespasio MJ, Piuzzi NS, Husni ME, Muschler GF, Guarino A, Mont MA. Knee osteoarthritis: a primer. Perm J. 2017;21.
3. Risberg MA, Oiestad BE, Gunderson R, Aune AK, Engebretsen L, Culvenor A et al. Changes in knee osteoarthritis, symptoms, and function after anterior cruciate ligament reconstruction: a 20-year prospective follow-up study. Am J Sports Med. 2016;44:1215-24.
4. Dulay GS, Cooper C, Dennison E. Knee pain knee injury knee osteoarthritis & work. Best Pract Res Clin Rheumatol. 2015;29:454-61.
5. Katz JN, Arant KR, Loeser RF. Diagnosis and treatment of hip and knee osteoarthritis: a review. JAMA. 2021;325:568-78.
6. Schock J, Truhn D, Abrar DB, Merhof D, Conrad S, Post M et al. Automated analysis of alignment in long-leg radiographs by using a fully automated support system based on artificial intelligence. Radiol Artif Intell. 2020;3:e200198.
7. Shamir L, Ling SM, Scott WW, Bos A, Orlov N, Macura TJ et al. Knee x-ray image analysis method for automated detection of osteoarthritis. IEEE Trans Biomed Eng. 2008;56:407-15.
8. Suresha S, Kidzin´ski L, Halilaj E, Gold G, Delp S. Automated staging of knee osteoarthritis severity using deep neural networks. Osteoarthritis Cartilage. 2018;26:S441.
9. Saleem M, Farid MS, Saleem S, Khan MH. X-ray image analysis for automated knee osteoarthritis detection. Signal Image Video Process. 2020;14:1079-87.
10. Chen P, Gao L, Shi X, Allen K, Yang L. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. Comput Med Imaging Graph. 2019;75:84-92.
11. Brahim A, Jennane R, Riad R, Janvier T, Khedher L, Toumi H et al. A decision support tool for early detection of knee osteoarthritis using X-ray imaging and machine learning: data from the osteoarthritis initiative. Comput Med Imaging Graph. 2019;73:11-8.
12. Shamir L, Ling SM, Scott W, Hochberg M, Ferrucci L, Goldberg IG. Early detection of radiographic knee osteoarthritis using computer-aided analysis. Osteoarthritis Cartilage. 2009;17:1307-12.
13. Alenazi AM, Alshehri MM, Alothman S, Alqahtani BA, Rucker J, Sharma N et al. The association of diabetes with knee pain severity and distribution in people with knee osteoarthritis using data from the osteoarthritis initiative. Sci Rep. 2020;10:3985.
14. Chen P. Knee Osteoarthritis Severity Grading Dataset. Mendeley Data. 2018.
15. Tolay P, Vajinepalli P, Bhattacharya P, Firtion C, Sisodia RS. Automated fetal spine detection in ultrasound images. Computer-Aided Diagnosis. vol. 7260. SPIE; 2009. p. 1054-63.
16. Davies-Tuck M, Wluka AE, Wang Y, Teichtahl A, Jones G, Ding C et al. The natural history of cartilage defects in people with knee osteoarthritis. Osteoarthritis and Cartilage. 2008;16:337-42.
17. Guo G, Wang H, Bell D, Bi Y, Greer K. KNN model-based approach in classification. In: On The Move to Meaningful Internet Syst 2003: CoopIS, DOA, and ODBASE: OTM Confed Int Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, Nov 3-7, 2003. Proc. Springer; 2003.p. 986-96.

18. Zhang ML, Zhou ZH. ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognit. 2007;40:2038-48.
19. Soucy P, Mineau GW. A simple KNN algorithm for text categorization. In: Proc 2001 IEEE Int Conf Data Min. IEEE; 2001:647-8.
20. Deng Z, Zhu X, Cheng D, Zong M, Zhang S. Efficient KNN classification algorithm for big data. Neurocomputing. 2016;195:143-8.
21. Xing W, Bei Y. Medical health big data classification based on KNN classification algorithm. IEEE Access. 2019;8:28808-19.
22. Bayramoglu N, Tiulpin A, Hirvasniemi J, Nieminen MT, Saarakkala S. Adaptive segmentation of knee radiographs for selecting the optimal ROI in texture analysis. Osteoarthritis Cartilage. 2020;28:941-52.
23. Olsson S, Akbarian E, Lind A, Razavian AS, Gordon M. Automating classification of osteoarthritis according to Kellgren-Lawrence in the knee using deep learning in an unfiltered adult population. BMC Musculoskelet Disord. 2021;22:1-8.
24. Thomson J, O'Neill T, Felson D, Cootes T. Automated shape and texture analysis for detection of osteoarthritis from radiographs of the knee. In: Med Image Comput Comput Assist Interv–MICCAI 2015: 18th Int Conf, Munich, Germany, October 5-9, 2015, Proceedings, Part II 18. Springer; 2015;127-34.