81861_Auto_Edited.docx

Retrospective Study

Deep learning-assisted diagnosis of femoral trochlear dysplasia based on MRI measurements

Xu SM *et al*. AI diagnosis of femoral trochlear dysplasia.

Sheng-Ming Xu, Dong Dong, Wei Li, Tian Bai, Ming-zhu Zhu, Gui-Shan Gu

Abstract

BACKGROUND

Femoral trochlear dysplasia(FTD) is an important risk factor for patellar instability. Dejour classification is widely used at present and relies on standard lateral X-rays,which are not common in clinical work. Therefore, magnetic resonance imaging (MRI) has become the first choice for the diagnosis of FTD. However, manually measuring is tedious, time-consuming, and easily produces great variability.

AIM

The aim of this study was to use artificial intelligence(AI) to assist diagnosing FTD on MRI images and to evaluate its reliability.

METHODS

We searched 464 knee MRI cases between January 2019 and December 2020, including FTD (n = 202) and normal trochlea (n = 252). This paper adopts the heatmap regression method to detect the key points network. For the final evaluation, several metrics (accuracy, sensitivity, specificity,etc) were calculated.

RESULTS

The accuracy, sensitivity, specificity, positive predictive value and negative predictive value of the AI model ranged from 0.74-0.96. All values were superior to junior doctors and intermediate doctors, similar to senior doctors. However, diagnostic time was much lower than that of junior doctors and intermediate doctors.

CONCLUSION

The diagnosis of FTD on knee MRI can be aided by AI and can be achieved with a high level of accuracy.

Core Tip: Femoral trochlear dysplasia is an important risk factor for patellar instability.MRI has become the preferred method for evaluating FTD.However,manually measuring femoral trochlea parameters on MRI is tedious, time-consuming, and easily produces great variability.In this work,we propose an assisted diagnosis algorithm framework based on deep learning technology, which can quickly and accurately distinguish whether there is trochlear dysplasia in the femur. All values(The accuracy, sensitivity, specificity,etc) were superior to junior doctors and intermediate doctors, similar to senior doctors. Our model is beneficial to both orthopedic surgeons and radiologists,especially,the young front-line clinicians with less experience.

INTRODUCTION

Femoral trochlear dysplasia (FTD) was first proposed by Swedish professor Brattstrom in 1964[1]. It is an important risk factor for patellar instability[2]. It is defined as a shallow, flat or raised trochlear groove, with an incidence of 96% in patients with recurrent patellar instability[3-4]. Different degrees of FTD are treated differently. Dejour classification is widely used at present and relies on the morphological features of standard lateral X-rays of the knee to describe the increasing disease severity (type A-D)[5]. However, it is subjectively dependent on doctors and has the disadvantages of poor sensitivity and specificity, which may lead to the neglect of FTD and the formation of an incorrect treatment plan for patients.

Mathias *et al*[6]reviewed all the literature related to FTD measurement,

including X-ray CT and MRI examination, and summarized 33 FTD evaluation methods. However, compared with X-ray and CT, MRI has the advantage of protruding articular cartilage, so MRI has become the preferred method for evaluating femoral trochlear dysplasia[7]. Specifically, the measurement of femoral trochlear groove depth, lateral trochlear inclination and trochlear facet asymmetry by MRI is particularly good in distinguishing normal and dysplastic femur trochlea[6-11]. However, tedious and repeated measurement is essential to using these qualitative and quantitative parameters to diagnose FTD, and it easily produces considerable differences in intragroup consistency and intergroup consistency. For surgeons and radiology doctors, especially resident physicians with less experience, it not only increases the working hours and burden but also increases the occurrence of misdiagnosis.

Artificial intelligence (AI) has become more and more popular in medical research, because it can not only quickly and accurately assist the diagnosis of diseases, such as cancer, but also participate in the robot surgery system to accurately treat diseases, which has made a great contribution to the development of health care system[12,13].Recently, the use of deep learning methods in medical imaging has aroused much interest [14,15]. Seok *et al*[16] assessed the ability of artificial intelligence to detect and classify proximal humeral fractures using ordinary shoulder radiographs. Their results show that CNN outperforms general physicians and orthopaedists. Pranata *et al*[17] used a deep learning algorithm to automatically classify and detect fracture locations in CT of the calcaneus. Urakawa *et al*[18] trained a convolutional neural network to compare its performance with orthopaedists in diagnosing intertrochanteric fractures. Cheng *et al*[19] localized and detected hip fractures on pelvic X-rays by using a deep convolutional neural network (DCNN). The feasibility and effectiveness of deep neural networks for hip fracture screening were confirmed. Lindsey *et al*[20] developed a deep neural network to assist emergency physicians in detecting and locating fractures in X-ray photographs. These findings show that senior medical experts can share their expertise with young doctors at the forefront of medicine through deep learning networks.

In addition to X-rays, deep learning and appropriate construction models can also efficiently identify MRI images, and their role in assisting in diagnosis is obvious. Zhou et al[21] used a combination of CNN and 3D deformation modelling to perform fast and accurate comprehensive knee joint tissue segmentation. Liu et al [22] developed a deep learning-based fully automatic cartilage damage detection system using segmentation and classification convolutional neural networks (CNNs). FritzBenjamin et al[23] compared the accuracy of a fully automatic DCNN with radiologists in identifying medial and lateral meniscus tears in MRII. The results showed that meniscus tear detection based on DCNN can be performed fully automatically with similar specificity. Shin H's study [24] showed that the CNN model can be used to diagnose the presence of meniscal tears and distinguish the types of meniscus tears. Tang et al [25] also proposed a fully automatic CNN-based knee segmentation system for the rapid and accurate evaluation of knee images.

All these developments suggest that artificial intelligence seems to be a breakthrough in solving the problem of femoral trochlea diagnosis. We propose an artificial intelligence system to label and detect the key points of knee MRI to assist in the diagnosis of femoral trochlear dysplasia. Our assumption is that the deep learning method will provide diagnostic performance similar to that of clinical radiologists and achieve higher intraobserver consistency to detect FTD.

MATERIALS AND METHODS

Description of the dataset

We collected 202 cases of femoral trochlear dysplasia examined in the First Hospital of Jilin University from January 2019 to August 2022. The inclusion criteria were that there was at least one history of patellar dislocation and the MRI suggested the presence of trochlear dysplasia. The exclusion criteria were as follows: osteoarthritis, rheumatoid arthritis, ipsilateral knee surgery history and MRI with image quality problems. Additionally, according to the above inclusion and exclusion criteria, 252 patients with other knee joint problems, such as mild soft tissue injury or joint effusion, were selected

for the same period of knee MRI. Sex and age were matched with the trochlear dysplasia group. The diagnosis of trochlear dysplasia was made by senior radiologists and senior surgeons with more than 10 years of working experience, and all differences were resolved before training and testing, as confirmed by MRI images. Finally, 464 cases of knee MRI were selected, including femoral trochlea dysplasia (n = 202) and normal trochlea (n = 252), 263 of whom were female and 201 of whom were male. The MRI images used in this study were 1.5T and were obtained from a Siemens digital radiography facility. Then, after preprocessing, the image dataset was randomly divided into a training set, validation set and test set. Ethical approval was provided by the Medical Ethics Committee.

Data preparation

The sample was marked by three doctors, including a senior surgeon, an intermediate surgeon and a junior surgeon. Trochlear depth and asymmetry of the facet length were measured 3 cm above the femorotibial joint space, and lateral trochlear inclination was measured on the first craniocaudal image that depicted the complete cartilaginous trochlea [10, 11]. All the samples in the training set and the validation set were marked by senior surgeons, and two surgeons marked and judged the samples in the test set (Figure 1).

Training the deep CNN and framework

In the current diagnosis process, doctors need to manually mark the anatomical key points of the MRII to measure the anatomical parameters and complete femoral trochlear dysplasia diagnosis through measurement results. However, the manual annotation process is time-consuming, and the diagnosis consistency varies. Therefore, a key point detection method is proposed in this paper to realize the automatic diagnosis of diseases.

Since different key points have different feature information, to achieve accurate automatic positioning of key points, this paper adopts the heatmap regression method

to detect the key points network. During training, the numerical coordinate labels are first converted into labels in the form of heatmaps according to a Gaussian distribution function. In the heatmap label, the pixel value closer to the key point is closer to 1. The model can fully learn the regional information between different key points by using the label in the form of a heatmap. The key point detection method using heatmap regression can be regarded as a pixel-level regression task. Therefore, a U-shaped network structure is designed in this paper to learn the mapping between the input MRII and the heatmap. The whole network is divided into two structures: encoder and decoder. The encoder is composed of five network layers through continuous convolution and downsampling operations.

The feature map channel size of each layer is continuously increased from 64 to 1024, and two cascaded 3x3 convolutions are included in each layer. Each convolutional model is composed of a 3x3 convolution, batch normalization (BN) and the LeakyReLU activation function. The downsampling operation between different layers uses the maximum pooling method to retain the features learned by the previous layer. The decoder is also composed of a five-layer network.

However, in the encoder, the feature map learned by the encoder is restored to the original input scale by using an upsampling operation. In the upsampling process, the learned feature maps of the encoder and the decoder are splicedin the direction along the channel by means of skip cascading. The reason for this approach is that the low-level features of the image, such as the line, texture, and shape of the femur, are usually learned in the first few layers of the encoder. As the network continues to deepen, the network learns more abstract high-level features, such as anatomical information in femoral images. A more accurate heatmap is predicted by fusing low-level features with high-level features using a skip cascade. In a convolutional neural network, convolution usually learns the local feature information. Therefore, the model needs to focus on the global context information in the image to learn the discriminative information around the key points. This paper introduces a nonlocal neural network designed based on the attention mechanism of the previous structure and captures the global contextual

information in the image through this structure. Since a nonlocal network usually requires large computational complexity, it is placed between the encoder and decoder. Because the feature map learned in the last layer of the encoder has the smallest scale, learning high-level features is enhanced by the model. In the last layer of the entire network, the final heatmap is calculated using 1x1 convolutions. The number of channels in the heatmap is consistent with the number of key points. By obtaining the position of the extreme value of the heatmap on each channel, the final key point coordinates are obtained.

During training, the sigma size of the heatmap is set to 3. The loss function adopts the squared error loss function and uses Adam as the optimizer for training. The total number of iterations for training is 25,000, while the initial learning rate is set to 0.001 and the batch size is set to 4. The model is trained in the training set, and the model with the smallest error in the validation set is selected as the final test model (Figure 2).

Image processing

Because of the brightness difference in the collected patient images, histogram equalization is used to process the image brightness. Additionally, to prevent overfitting in the training process, the image is enhanced by SimpyITK, which includes random flipping, random scaling and random rotation. All images are sampled to 256×256 pixels. The pixel space size of the image in the dataset is distributed between 0.546 mm/pixels and 0.625 mm/pixels.

Statistical analysis:

SPSS 26.0 was used for all statistical analyses. Frequency and percentage were used to describe categorical variables, mean and standard deviation were used to describe continuous variables conforming to a normal distribution, and median and quartile were used to describe continuous variables not conforming to a normal distribution. Differences between groups were compared using the t test, Mann–Whitney U test, or chi-square test.

Using the paired sample T test, one-way ANOVA or Kendall's W test to compare the measurement results, time spent and measurement errors between the AI model and clinicians, P < 0.05 was considered to be statistically significant.

The sensitivity, specificity, positive predictive value, negative predictive value and accuracy of the AI model, junior doctors and intermediate doctors' diagnoses of abnormal trochlea were calculated. The area under the ROC curve (AUC) was calculated to compare the diagnostic efficacy of the AI model, junior doctors and intermediate doctors for abnormal trochlea. Additionally, the intraclass correlation coefficient (ICC) was calculated using the two-way mixed effects model to compare the absolute consistency of the diagnostic values of abnormal trochlea calculated by the AI model, junior doctors and intermediate doctors with those of senior doctors.

The kappa consistency test was used to compare the diagnostic consistency of the AI model, junior doctors, and intermediate doctors with that of senior doctors and the test-retest reliability of the AI model, junior doctors, and intermediate doctors after 2 wk.

RESULTS

General information

The age and sex of the participants were collected. Participants ranged in age from 16 to 68 years old. As shown in Table 1, there were 370 people in the training set, with a median age of 39 years, and 217 women (58.6%); there were 94 people in the test set, with a median age of 38.2 years, and 46 women (48.9%). There was no statistically significant difference in age or sex between the participants in the training set and the test set (P = 0.646, P = 0.090).

Diagnostic efficiency and accuracy

As shown in Table 2, the average time for the AI model to make a diagnosis through two levels of MRI images was only 0.14 s, which was significantly shorter than that of intermediate doctors and junior doctors. Compared with senior doctors, junior doctors, intermediate doctors and AI models had statistically significant differences in the measurement errors of the three parameters.

AI model diagnosis effect

As shown in Table 3, taking the diagnosis results of senior doctors as a reference, the sensitivity, specificity, positive predictive value and negative predictive value of the AI model ranged from 0.74-0.96. The accuracy of the AI model was 0.88, and the AUC was 0.88. The ability of the AI model to distinguish abnormal trochlea at all MRI levels was higher than that of intermediate doctors and junior doctors. In the diagnosis of abnormal trochlea, the kappa value of the AI model was 0.76, which was highly consistent with that of senior doctors. It was better than that of intermediate doctors (0.60) and junior doctors (0.58), which had a moderate degree of consistency compared with senior doctors.

As shown in Table 4, comparing the measurement results of the diagnostic parameters between the AI model and the senior doctor, the paired sample t test results showed that there was no significant difference in the lateral trochlear inclination and the depth of the trochlear groove ($P = 0.424$), but there was a significant difference in the facet ratio of medial to lateral ($P = 0.024$).

As shown in Table 5, the measurement results of the junior doctor, intermediate doctor and AI models on diagnostic parameters were compared. The results showed that the measurements of the trochlear groove depth and lateral trochlear inclination of the AI model were in good agreement with those of senior doctors (ICC=0.91, 95%CI, 0.86-0.94, $p < 0.001$, ICC=0.91 95%CI, 0.87-0.94, $P < 0.001$), while the measurement of the facet ratio of medial to lateral was generally consistent with that of senior doctors (ICC=0.71, 95%CI, 0.59-0.80, $p < 0.001$). The measurement of the facet ratio of medial to lateral between junior doctors and intermediate doctors was less consistent with that of senior doctors (ICC=0.57, 95%CI, 0.32-0.73, $p < 0.001$; ICC= 0.41 95%CI, 0.16-0.60, $P < 0.001$). The measurement results of the three parameters of the AI model were better than those of junior and intermediate doctors.

Retest reliability

As shown in Table 6, the diagnosis of abnormal trochlea by the AI model was consistent at all MRI levels, showing very high test-retest reliability. For intermediate doctors and

junior doctors, the diagnosis had a high consistency, and the kappa value ranged from 0.76 to 0.78.

As shown in Table 7, the measurement results of the AI model for the three parameters were completely consistent (ICC = 1.00, 95%CI, 1.00 to 1.00, P < 0.001; ICC = 1.00, 95%CI, 1.00 to 1.00, P < 0.001; ICC = 1.00, 95%CI, 1.00 to 1.00, P < 0.001). The retest reliability of the measurement of the depth of the trochlear groove and lateral trochlear inclination were good between junior and intermediate doctors, with ICC values ranging from 0.89 to 0.93, but the retest reliability of the facet ratio of medial to lateral was moderate (ICC = 0.71, 95%CI, 0.60-0.80, P < 0.001; ICC = 0.62, 95%CI, 0.47-0.73, P < 0.001).

DISCUSSION

FTD is a progressive disease that leads to patellar instability and even patellofemoral arthritis. There are many methods to treat FTD, such as trochleoplasty [26,27]. However, trochleoplasty is not suitable for all patients with FTD. For a large number of patients without patellar instability or mild FTD (type A), conservative treatment or simple MPFL is the best choice [28,29]. Therefore, early FTD diagnosis and determination of its severity are critical for identifying patients who require observation or treatment. The diagnosis of early FTD is challenging. There are different measurement parameters to describe FTD in the related literature, including X-ray, CT and MRI. However, standard lateral radiographs of the knee joint are not common in clinical work. With the continuous development of medical imaging technology, MRI can clearly show the anatomy of the articular cartilage surface, ligaments and muscles can better show the morphology of the femoral trochlea, and has more advantages in evaluating FTD [30,31]. Therefore, our model is based on MRII, not X-ray[32-34]. Among various quantitative radiology methods to characterize the femoral trochlea, trochlear depth, asymmetry of the facet length and lateral trochlear inclination are considered the best prediction parameters, with extremely high sensitivity and specificity [35-37].

However, the manual measurement of femoral trochlea parameters is tedious, time-consuming, and easily produces great variability. The development of artificial intelligence in imaging-related fields provides a breakthrough for solving problems of the femoral trochlea. Artificial intelligence has been previously applied to the diagnosis and treatment of femoral trochlea diseases. CerveriPietro et al[38] combined a statistical shape model (SSM) with a stacked sparse autoencoder (SSPA) network to represent the individual morphology of the trochlea through a set of parameters and used these parameters to distinguish between different degrees of abnormalities and calculate the SSM of normal and dysplasia trochlea regions. Some scholars[39] also proposed a reproducible measurement method of the 3D femoral model, quantified the knee parameters of the distal femur, and used an artificial neural network to predict the parameter values describing the geometry of the normal trochlear groove to simulate the surgical procedure of femoral trochleoplasty. However, limited by the computer level at that time, the further application of artificial intelligence was not studied. However, there is no doubt that the research of Cerveri Pietro and Jin[38, 39] provided evidence that artificial intelligence can be applied to the diagnosis and treatment of FTD.

In this study, we propose an aided diagnosis algorithm framework based on deep learning technology, which shows that deep learning achieves high performance in distinguishing normal femoral trochlea from trochlea dysplasia, which is superior to junior doctors and intermediate doctors and similar to that of senior doctors (ICC=0.91, 95%CI, 0.87 to 0.94, $p < 0.001$). For the diagnosis of trochlea dysplasia, the accuracy of the AI model was 0.88, and the AUC value was 0.88. The range of sensitivity, specificity, positive predictive value and negative predictive value was 0.74-0.96, which was higher than that of junior doctors and intermediate doctors. In terms of intragroup consistency and intergroup consistency, the AI model was also superior to junior doctors and intermediate doctors. However, in terms of diagnosis time, the average time of the AI model was only 0.14 s, which was significantly shorter than that of junior doctors and intermediate doctors.

To locate the key points accurately and automatically, the thermal map regression method is used to detect the key points. An algorithm based on deep learning is developed and used to automatically measure the femoral trochlear parameters in axial MRI images. The system can accurately, efficiently and stably evaluate trochlear groove depth and lateral trochlear inclination. Its performance is superior to junior doctors and intermediate doctors and similar to that of senior doctors. The good performance of the system shows the clinical potential of artificial intelligence for assisting surgeons in the tedious process of accurately measuring femoral trochlear parameters. It is good to help young doctors with less experience detect femoral trochlear dysplasia at an early stage.

The feasibility of using deep learning to help doctors diagnose diseases on MRI of the knee joint has been well verified. The high efficiency, accuracy and repeatability of artificial intelligence can solve the problem of the large demand for diagnosis and measurement of the femoral trochlea. In addition to its widespread recognition, MRI does not involve cutting the selected sample image, thus avoiding the problem that the accuracy rate may be improved due to image cutting[23].

In addition, it was found that the measurement of lateral trochlear inclination and depth of the trochlear groove were better than the measurement of the facet ratio of medial to lateral, both in terms of intragroup consistency and in terms of the error of disease diagnosis. The reason for this is that although MRI can better observe the shape of the femoral trochlea compared with X-ray, it cannot describe the shape of the trochlear surface by only marking key points, which leads to large measurement errors and poor consistency within and between groups. Therefore, when using MRI to diagnose FTD, the selection of parameters is more inclined to measure the lateral trochlear inclination and the depth of the trochlear groove.

This study demonstrates that the deep learning algorithm can be used to assist in the diagnosis of FTD, but this does not mean that it is ready for immediate clinical practice. We evaluated the diagnostic performance of the CNN based on a single axial MRI image of the knee to maintain the project simplicity, which may not actually fully reflect the clinically relevant situation because the assessment of femoral trochlear

development will involve at least two levels of MRI imaging. Finally, limited by the resolution of the 1.5T MRI axial image, the diagnostic performance of the CNN is reduced to a certain extent. Using higher resolution images can improve diagnostic accuracy.

However, there are still some deficiencies that are unavoidable in the initial stage of the research. In the future, we hope to conduct further research based on the existing data and research results, such as how to classify FTD to guide the treatment of different types.

CONCLUSION

In conclusion, this paper adopts the heatmap regression method base on deep learning to build an AI model. All values of AI model were superior to junior doctors and intermediate doctors and similar to senior doctors.Therefore, the diagnosis of FTD on knee MRI can be aided by AI and can be achieved with a high level of accuracy.AI have great potential to become a useful tool for the assisted diagnosis of orthopaedic diseases.Its greatest significance is to assist young front-line clinicians with less experience to complete the diagnosis of the disease faster and more accurately.

ARTICLE HIGHLIGHTS

Research background

Femoral trochlear dysplasia (FTD) is an important risk factor for patellar instability, with an incidence of 96% in patients with recurrent patellar instability.Magnetic resonance imaging(MRI) has become the preferred method for evaluating FTD.However, tedious and repeated measurement is essential to using these qualitative and quantitative parameters to diagnose FTD, and it easily produces considerable differences in intragroup consistency and intergroup consistency.

Research motivation

Whether artificial intelligence can be used to assist in the diagnosis of femoral trochlear dysplasia remains unclear.

Research objectives
Research objectives was to propose an artificial intelligence(AI) system to label and detect the key points of knee MRI to assist in diagnosing FTD quickly and accurately.

Research methods
We searched knee MRI cases , including femoral trochlear dysplasia and normal femoral trochlea,All the samples marked by doctors were divided into three sets,including the training set, the validation set and the test set.The performance of AI model to diagnose FTD was improved through continuous training and learning.

Research results
All values(The accuracy, sensitivity, specificity,etc) were superior to junior doctors and intermediate doctors and similar to senior doctors. In terms of intragroup consistency and intergroup consistency, the AI model was also superior to junior doctors and intermediate doctors. However, diagnostic time was much lower than that of junior doctors and intermediate doctors.

Research conclusions
AI has great potential in the assisted diagnosis of orthopedic diseases.Its greatest significance is to assist young front-line clinicians with less experience to complete the diagnosis of the disease faster and more accurately.

Research perspectives
In the future, we hope to conduct further research based on the existing data and research results, such as how to classify FTD to guide the treatment of different types.

19 words — < 1%

8 mdpi-res.com
Internet
17 words — < 1%

9 opg.optica.org
Internet
17 words — < 1%

10 ped-rheum.biomedcentral.com
Internet
16 words — < 1%

11 reproductive-health-journal.biomedcentral.com
Internet
15 words — < 1%

12 Maoxi Zhong, Xiaozhu Zeng, Jinmin Xue, Yubing Yue, Yuxi Zhu. "PD-1/PD-L1 Inhibitors vs. Chemotherapy For Previously Treated Advanced Gastroesophageal Cancer: A Meta-Analysis of Randomized Controlled Trials", Research Square Platform LLC, 2021
Crossref Posted Content
14 words — < 1%

13 mdanderson.influuent.utsystem.edu
Internet
12 words — < 1%

14 pdfs.semanticscholar.org
Internet
12 words — < 1%

EXCLUDE QUOTES   ON   EXCLUDE SOURCES   < 12 WORDS
EXCLUDE BIBLIOGRAPHY   OFF   EXCLUDE MATCHES   < 12 WORDS