Artificial Intelligence to Diagnose Tibial Plateau Fractures: An Intelligent Assistant for Orthopedic Physicians

Peng-ran LIU[1], Jia-yao ZHANG[1], Ming-di XUE[1], Yu-yu DUAN[2], Jia-lang HU[3], Song-xiang LIU[1], Yi XIE[1], Hong-lin WANG[1], Jun-wen WANG[3#], Tong-tong HUO[1#], Zhe-wei YE[1#]

[1]*Department of Orthopedics, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China*
[2]*Hubei University of Chinese Medicine, Wuhan 430070, China*
[3]*Department of Orthopedics, Wuhan Pu'ai Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China*

[Abstract] Objective: To explore a new artificial intelligence (AI)-aided method to assist the clinical diagnosis of tibial plateau fractures (TPFs) and further measure its validity and feasibility. Methods: A total of 542 X-rays of TPFs were collected as a reference database. An AI algorithm (RetinaNet) was trained to analyze and detect TPF on the X-rays. The ability of the AI algorithm was determined by indexes such as detection accuracy and time taken for analysis. The algorithm performance was also compared with orthopedic physicians. Results: The AI algorithm showed a detection accuracy of 0.91 for the identification of TPF, which was similar to the performance of orthopedic physicians (0.92±0.03). The average time spent for analysis of the AI was 0.56 s, which was 16 times faster than human performance (8.44±3.26 s). Conclusion: The AI algorithm is a valid and efficient method for the clinical diagnosis of TPF. It can be a useful assistant for orthopedic physicians, which largely promotes clinical workflow and further guarantees the health and security of patients.

Key words: artificial intelligence; tibial plateau; fracture; diagnosis

Tibial plateau fracture (TPF) is a fracture of proximal tibia in the knee joint. It is a severe articular injury with a broad damage-spectrum to the locomotor system, which usually accompanies poor clinical effect and limited articular function[1, 2]. As the pivotal location of force conduction among the lower extremity, the proximal tibia could be damaged by a compression fracture, split fracture, bone defect or other structural injuries during an excessive violent load[3, 4]. Currently, A conventional X-ray remains the primary diagnostic method to detect TPF in the orthopedics department, which processes convenient, rapid, and simple characteristics compared with other imaging-related examinations. Generally, image reading ability is a basic clinical skill that has to be mastered by qualified orthopedic physicians, which concerns the accurate diagnosis of TPF and effective treatment for patients. However, under the pressure of overloaded clinical work caused by surging patient demand as well as insufficient medical resources, it is prone to induce the risk of missed diagnoses and misdiagnoses, especially when the TPF fractures are not obvious on an X-ray, such as a minor fracture, non-displaced fracture, or occult fracture. There have been studies to indicate that under severe and urgent conditions, diagnostic errors often occurred to influence a correct reading of the radiograph, and involved uncooperative patients, inadequate clinical history, time critical decisions, and simultaneous clinical work, during which the missed image-based diagnosis and misdiagnosis could even exceed 40%, causing a great threat to the health and safety of patients[5–7]. Thus, exploring an accurate and safe auxiliary method to quickly detect TPF has been of great interest to orthopedic physicians.

Artificial intelligence (AI), the interdisciplinary study of computer technology, mathematic, cybernetics as well as determinism, is a new technology in the 21st century, which has led to an earth-shaking transformation of the world's operational model. The conception of AI is studying and researching human intelligence (HI) and making imitation computers based on an intelligent algorithm to simulate HI and even surpass it[8]. With the appearance of machine

Peng-ran LIU, E-mail: lprlprlprwd@163.com
#Corresponding authors, Zhe-wei YE, E-mail: yezhewei@hust.edu.cn; Tong-tong HUO, E-mail: D202081631@hust.edu.cn; Jun-wen WANG, E-mail: wjw0730@163.com

learning (ML), deep learning (DL) and convolutional neural networks (CNNs), the primary techniques of AI that are favorably suited to capture feature items and learning, AI has transformed into a performed method in image analysis. It has also gradually formed several functional applications such as (1) computer vision, (2) speech recognition, (3) natural language recognition, (4) decision planning and (5) big data analysis, which have also been applied to traditional industries including the medical field. In past studies, AI has been successfully used to assist in tumor detection of pathological sections[9], gastrointestinal disease detection using capsule endoscopy[10], thyroid disease detection by ultrasound[11], pulmonary nodules and cancer detection with computerized tomography (CT), and also some orthopedic diseases with imaging-based examination such as scoliosis[12], osteoarthritis[13], and meniscus and cruciate ligament injuries[14]. All of these applications have shown satisfying results to improve detection accuracy and reduce the clinical workload. However, to the best of our knowledge, there are no studies involving AI-assisted TPF detection.

Therefore, the present study explored whether AI could be used for TPF detection on X-ray images and compared the performance of the AI to that of veteran orthopedic physicians, which aimed to further verify the feasibility and ability of AI-aided diagnosis. These results may aid in the development of a novel method for the clinical diagnosis and treatment of TPF.

1 MATERIALS AND METHODS

1.1 Database and Study Design

According to the difficulty of data acquisition, the research was performed as a multi-center study among five Chinese triple-A grade hospitals, including the Wuhan Union Hospital, Wuhan Puai Hospital, Wuhan Puren Hospital, Xiangya Changde Hospital, and Northern Jiangsu People's Hospital. After data collection between August 2020 and August 2021, a total of 542 anterior knee joint X-rays of patients with TPF were acquired and included in the final database. Next, the X-rays in the database were converted from Digital Imaging and Communications in Medicine (DICOM) files to Joint Photographic Experts Group (JPEG) files with a matrix size of 1080×1333 pixels by Photoshop 20.0 (Adobe Corp., USA). Using the random-number-table function in Excel (Microsoft Corp., USA), the 542 JPEG files were numbered and randomly divided into two datasets: the training dataset (including 458 files, for algorithm learning and training) and the test dataset (including 84 files, for algorithm validation). The ratio of the two datasets was nearly 17:3.

With the training dataset, a type of DL-CNN recognition algorithm was setup and trained to learn the appearance of TPF on knee joint X-rays. After the training, the algorithm could automatically recognize and label the suspicious area of TPF on knee joint X-rays in the test dataset, which could be considered a detection assistant. Finally, the performance of the CNN algorithm, including accuracy and time spent on analysis, were compared with a panel of five veteran orthopedic physicians in the Orthopedics Department of Wuhan Union Hospital. To protect patient privacy, all identifying information on the X-rays were anonymized and omitted. The study was approved by the Ethics Committee of Wuhan Union Hospital.

1.2 Training the Algorithm

1.2.1 RetinaNet Algorithm Firstly, a DL-CNN recognition algorithm [RetinaNet, primarily proposed by Kaiming He in 2018 (structure of the original RetinaNet shown in fig. 1)] was designed and improved. RetinaNet can extract image features with the Residual Network (ResNet), fuse the context information with the Feature Pyramid Network (FPN), and respectively predict the classification and location of objects with two CNNs. Compared to other frequently used algorithms (such as Faster R-CNN), RetinaNet has obvious detection advantages in terms of speed and accuracy.
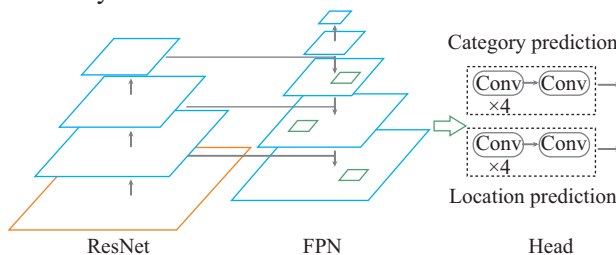

Fig. 1 The structure of RetinaNet

Subsequently, the original RetinaNet algorithm was further improved by applying the structure of MobileNet Version 2. For instance, the ResNet was replaced by Depthwise Separable Convolution in the feature extraction process. Additionally, a triple FPN was applied to expand the size of high-level images by upsampling, which was subsequently added to the low-level image to increase the semantic information and the accuracy of small targets in the low-level image. The structure of the improved RetinaNet was shown in fig. 2.
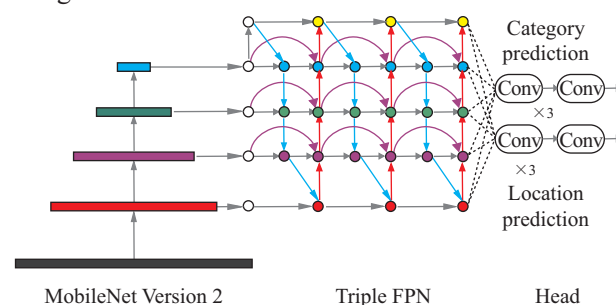

Fig. 2 The structure of the improved RetinaNet

1.2.2 Data Labeling   The training dataset (458 files with TPF) was labeled by two senior orthopedic physicians with more than 10 years of experience with the labeling software LabelImg (https://github.com/tzutalin/LabelImg). The fracture line of each file was labeled carefully as a region of interest (ROI) for training the improved-RetinaNet. Then, the process of data enhancement of the training dataset was implemented by the algorithm, including image rollover, rotation, cropping and blurring, which doubled the original training dataset (916 files with TPF).

1.2.3 Training of Improved RetinaNet   The improved-RetinaNet was trained with the labeled and enhanced training dataset using the Adam optimizer. The training parameters were set as follows: batch size, 2; dropout, 8; initial learning rate, le-3; and learning rate, 0.1.

1.3 Performance Assessment of the Algorithm

The detection ability of TPF in X-ray images of the trained and improved RetinaNet was assessed using the test database, which was considered the final algorithm performance. Each of the 84 files with TPF in the test database was diagnosed by the trained and improved RetinaNet to automatically label the suspicious fracture line. The output results would be judged by the two senior orthopedic physicians with more than 10 years working experience. The potential divergence was arbitrated by the radiological report of the 84 X-rays from the imaging department.

1.4 Performance Assessment of Orthopedic Physicians

To assess the diagnostic performance of orthopedic physicians on the clinical front line, there was also a panel set up of five volunteer orthopedic physicians from the Orthopedics Department of Wuhan Union Hospital, who had more than five years of working experience in image reading and orthopedics. These five orthopedic physicians were independent of this study and did not participate in any processes of this project. The panel of orthopedic physicians were set to individually diagnose the "blind" (without any reminder or radiological report) test dataset as TPF or non-TPF and they were not given a time limit. The diagnostic results of each physician were collected and judged by the same two senior orthopedic physicians to determine their accuracy, which was considered with the time taken for diagnosis as the performance of the orthopedic physicians.

1.5 Comparison of the Performance of the Algorithm and Orthopedic Physicians

To compare the performance of the improved RetinaNet and orthopedic physicians, the accuracy and time spent on analysis of the algorithm and orthopedic physicians were collected. The indexes were evaluated using the Student's *t*-test, and the procedure of the entire study was illustrated in fig. 3.

1.6 Statistical Analysis

The data from this study were presented as the mean±standard deviation (SD) or percentage, and the statistical analysis was performed using GraphPad Prism 7.0 software (GraphPad Corp., USA). The significance between the algorithm group and orthopedic physician group was evaluated using the Student's *t*-test. $P<0.05$ was considered to indicate statistical significance.

2 RESULTS

2.1 Performance of the Algorithm

2.1.1 Evaluation of the Algorithm Calculation

The final loss-curve-graph and P-S-curve-graph of the improved RetinaNet model were produced using Python script (https://www.Python.org) and are shown in fig. 4. The loss was composed of three parts: regression loss, classification loss, and total loss. As shown in fig. 4A and B, the loss curve of the training and test datasets decreased rapidly and then flattened out, which indicated that the model could converge well. To quantify the detection performance of the model, the P-S curve was drawn to show the diagnostic capability of the algorithm (fig. 4C). The P-S curve
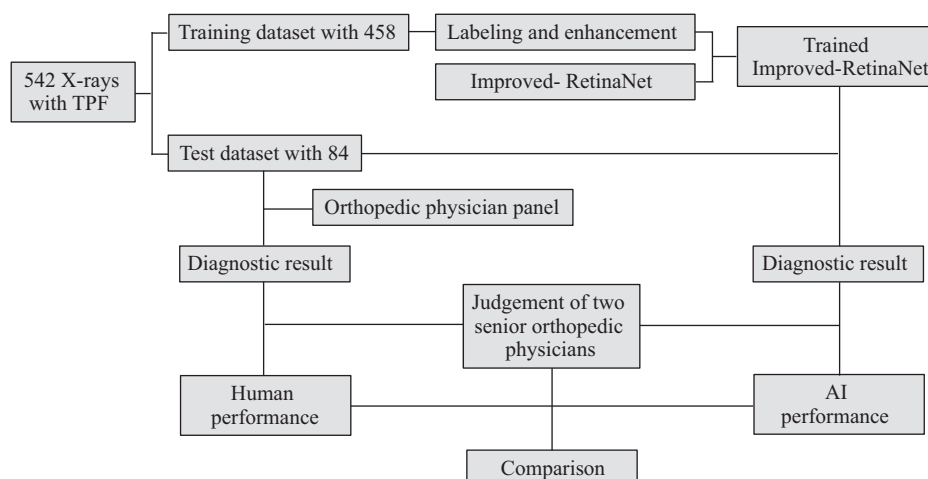

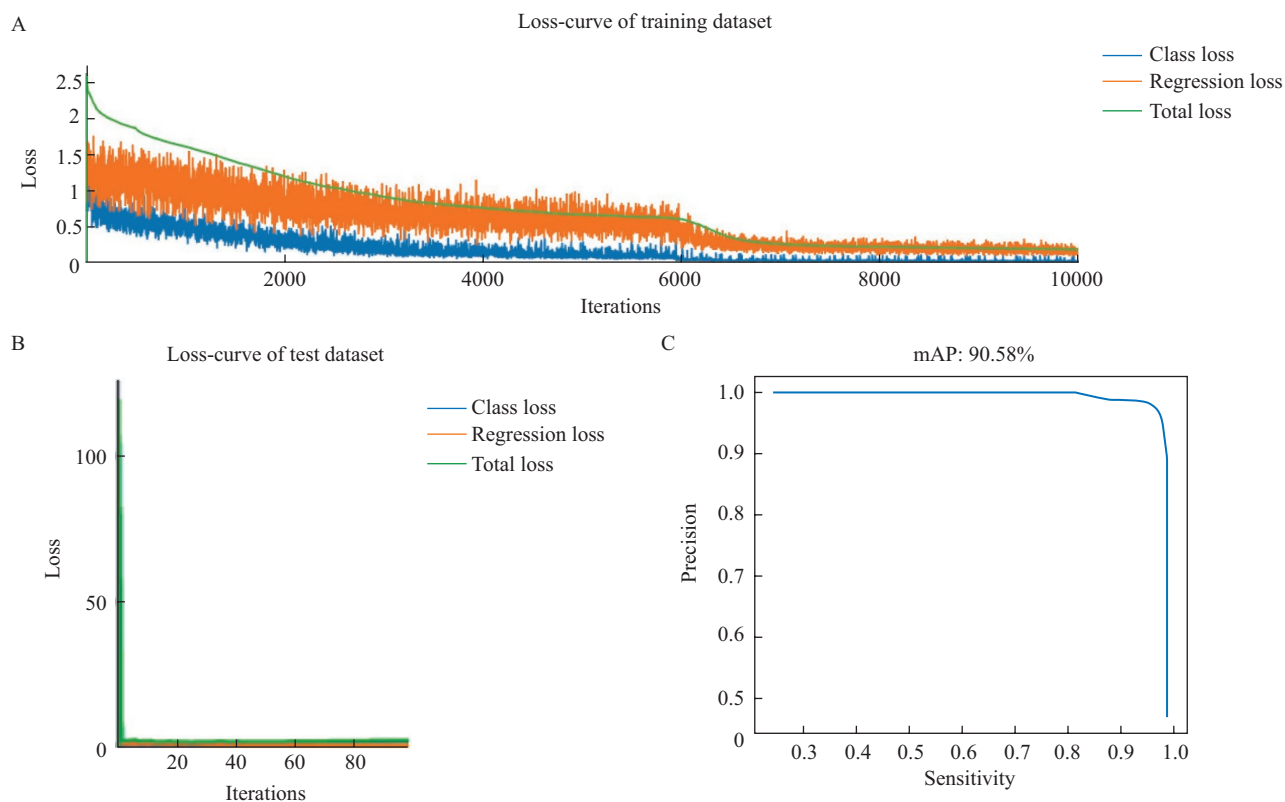**Fig. 3** Flow diagram for the entire study

Fig. 4 Loss-curve-graphs of improved RetinaNete
A: Loss-curve-graphs of training dataset. B: Loss-curve-graphs of test dataset. C: P-S-curve-graph of improved RetinaNet, with parameters as follows: score threshold, 0.5; mean average precision (mAP), 0.9058; average precision (AP), 0.963; autoregression (AR), 0.907; F1 score, 0.934.

was created by changing the threshold value, in which the abscissa represented sensitivity, and the ordinate represented precision. The detection performance of the algorithm could then be intuitively demonstrated by calculating the area under the curve.

2.1.2 Algorithm Performance   After training the algorithm, the test process of the algorithm automatically labeled the suspicious TPF area on the X-ray with a rectangle (fig. 5). According to the personal judgments by two senior orthopedic physicians, there were seven diagnostic errors (missed diagnosis or misdiagnosis) in the 84 files, and the accuracy of the algorithm was 0.91. The time spent on the entire test process was 47 s, which gave an average of 0.56 s per image.

2.2 Performance of the Orthopedic Physicians

The manual diagnostic results from the panel of orthopedic physicians were collected to calculate the accuracy, as well as time spent, and these data were shown in table 1.

2.3 The Difference Between the Algorithm and Orthopedic Physicians

The performance of the algorithm and orthopedic physicians were compared, and the results were shown in table 2 and fig. 6. The results show an accuracy of 0.91 for the algorithm and 0.92±0.03 for the orthopedic physicians, and this difference in accuracy between the

two groups was not statistically significant (P=0.86). However, the average time spent by the algorithm was 16 times faster than human performance (0.56 s versus 8.44±3.26 s), which indicated a statistically significant difference between the two groups (P<0.001).

3 DISCUSSION

TPF is a complex fracture in the knee joint caused by high-impact events. The clinical diagnosis of suspected TPF is often insufficient due to an unclear X-ray presentation, such as a minor fracture, non-displaced fracture, or occult-fracture, which can lead to missed diagnoses or misdiagnoses especially in emergency situations[15-17]. A missed diagnosis or misdiagnosis in the diagnostic process can result in severely negative clinical outcomes for patients, including delayed treatment and poor recovery. Therefore, a valid solution to assist orthopedic physicians in avoiding incorrect diagnoses is required to improve patient outcomes.

In this study, AI technology was used to construct an improved-RetinaNet algorithm that could assist orthopedic physicians to recognize TPF on knee joint X-rays, and it showed satisfying performance, with an accuracy of 0.91 and average time spent of 0.56 s. In comparison to human performance (accuracy of

Table 1 Performance of the orthopedic physicians

Performance	Physician 1	Physician 2	Physician 3	Physician 4	Physician 5
Correct/Incorrect	76/8	74/10	80/4	79/5	77/7
Accuracy	0.90	0.88	0.95	0.94	0.91
Entire time spent	697.2 s	495.6 s	1050.0 s	394.8 s	907.2 s
Average time spent	8.3 s	5.9 s	12.5 s	4.7 s	10.8 s

Table 2 Comparison between the algorithm and orthopedic physicians

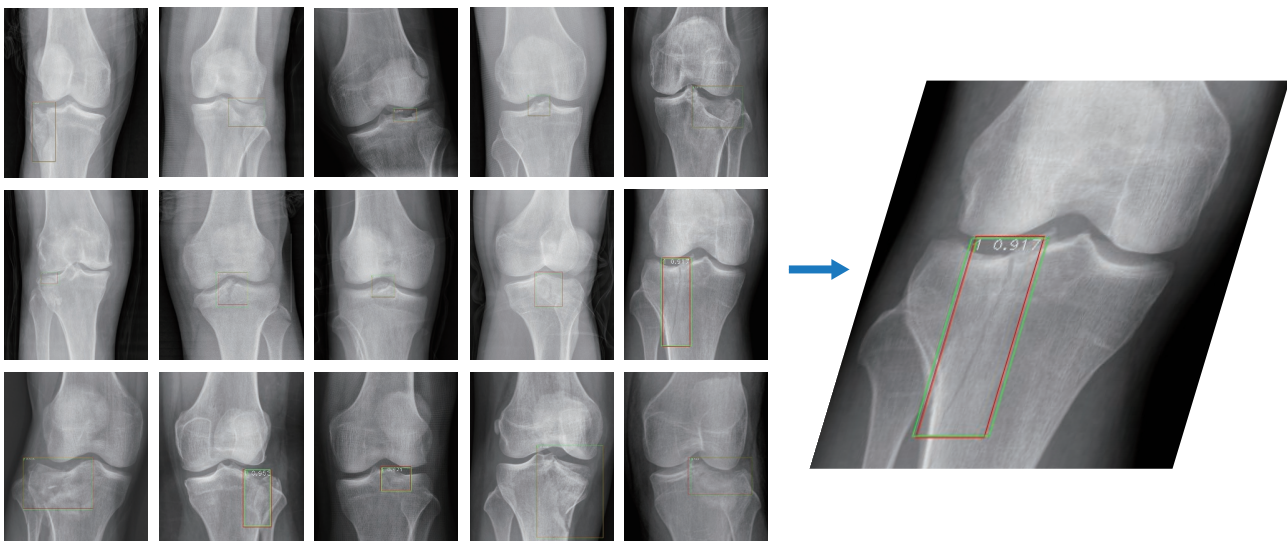| Performance	Algorithm	Orthopedic physicians	T value	P value
Correct/Incorrect	77/7	/	/	/
Accuracy	0.91	0.92±0.03	0.19	0.86
Average time spent	0.56 s	8.44±3.26 s	5.45	<0.001

Fig. 5 A selection of the output X-rays from the test dataset
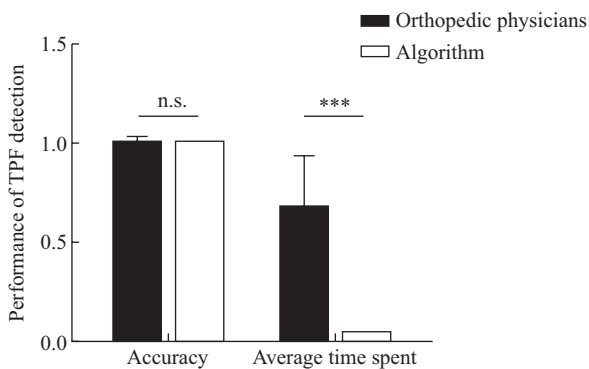The suspicious fractures were labeled with a rectangle by the trained and improved RetinaNet algorithm.

Fig. 6 The comparison between the algorithm and orthopedic physicians
n.s.: not significant. ***$P<0.001$

0.92±0.03), the AI algorithm was found to exhibit a similar accuracy of TPF recognition. However, due to the nonemergency and time-free constraints on the diagnosis process for the clinical panel in this study, which does not simulate the real circumstances in the emergency department and time-limited situations, the authors believe that the performance of the AI algorithm in this study would be even better in the real clinical environment. As for the time spent on the

diagnosis of TPF in the present study, the AI algorithm greatly reduced the time (16 times faster) compared with orthopedic physicians, which undoubtedly could be a powerful auxiliary method and relieve the burden of TPF rapid-diagnosis in the clinic.

Various studies have already testified the ability and feasibility of AI models for the detection of different clinical diseases involving X-ray images, which have shown similar results to the current study. For instance, in pulmonary disease detection, Hyunsuk designed an AI model trained by 5485 chest X-rays, and the sensitivity and specificity of lung nodule recognition reached 86.2% and 85.0%, respectively, which further arrived at 75.0% and 83.3% for lung cancer detection. In Hyunsuk's research, the AI model performed even better than the professional radiologist, and could also be a favorable assistant to reduce excessive workload and save medical resources[18]. In the detection of pneumonia, Wang set up a CNN algorithm to learn from 1647 chest images of confirmed patients and 800 from normal patients. After the training process, the sensitivity and specificity of the model for pneumonia detection reached 92.3% and 85.1%, respectively. The AI-diagnostic average time

spent was 0.55 min, which was 15 min less compared to human performance, providing great help for rapid clinical diagnosis[19]. Parallel studies have been conducted for the detection of orthopedic diseases. Similar to TPF in this study, the distal radius fracture (DRF) is a common fracture in orthopedics, whose rapid diagnosis is also confronted with difficulty. Because of this, Gan set up a CNN algorithm based on 2340 DRF patients and assessed the diagnostic ability for the detection of DRF on X-ray images. In their research, Gan indicated that the network exhibited a similar performance to the orthopedic physicians and could be feasible in clinical application as an auxiliary method under extended conditions[20]. Also concerning the detection of DRF, Lindsey developed a CNN model using a larger database that consisted of 135 409 cases of DRF patients. After the DL process by analyzing the database, the accuracy of the model was higher than the diagnostic ability of 18 senior professional orthopedic physicians. The sensitivity and specificity increased by 10.7% and 9.4% on average, which effectively reduced the risk of missed diagnosis and misdiagnosis in DRF detection and provided substantial improvements to patient care[21]. Similarly, for the detection of humeral supracondylar fractures (HSFs), Choi collected data of HSF patients who visited the orthopedic department within five years and designed an AI network. After training the algorithm, it was able to achieve efficient results with high sensitivity for HSF recognition, which plays an important role in rapid clinical diagnosis[22]. In the detection of proximal humeral fractures (PHFs), Chung analyzed four classifications of PHF (346 greater tuberosity fractures, 514 surgical neck fractures, 269 three-part fractures, and 247 four-part fractures) using an AI model. The final result proposed an effective algorithm for the detection of PHF, and Chung further noted that the AI was able to improve the Neer classification, which brought profound significance for clinical PHF diagnosis and treatment[23]. Besides fracture detection, AI technology has also played an inspiring role in the diagnosis of other orthopedic diseases, such as scoliosis, arthritis, bone tumors, and meniscus and ligament injuries[24–26]. Taken together, these previous studies have confirmed the validity of AI-aided methods for clinical diagnosis.

In the current study, an AI algorithm was applied for the detection of TPF. The results indicated that the TPF diagnostic effects were greatly facilitated, and the accuracy even reached the level of human performance under non-emergent circumstances. Therefore, we believe the AI algorithm can be fully competent as an assistant of orthopedic physicians for the detection of TPF by X-ray. With the support of AI technology, the clinical workload could be reduced, medical resources could be saved, and more importantly, the health and safety of patients could be

better guaranteed to a greater extent. However, the current study also has some deficiencies. (1) Despite the data-augmentation to expand the training test, the database in this study was small, which could influence the final performance of the AI algorithm. In future research, the scale of database should be enlarged to improve the performance of the algorithm. (2) The X-ray images collected in this study only consisted of anteroposterior film, which might not be optimal for the automated diagnosis of TPF. Moreover, the normal knee joint X-ray was not included in this study. As such, the index of sensitivity and specificity could not be accessed. In future research, lateral X-rays and X-ray images of normal knee joints should be included. (3) The algorithm in this study still has room for improvement to achieve better performance. Finally, (4) the automatic diagnosis in this study merely stayed at the fracture line recognition, which did not include fracture classification. The classification is a necessary process of fracture treatment and provides a crucial reference to develop a surgical plan. In future research, fracture classification recognition will be added to the algorithm. Additionally, the algorithm achieved a satisfying performance but lacks further confirmation. If a miscalculation occurs during real clinical application, the medical responsibility still cannot be identified due to incomplete relevant laws. However, this should not be the reason to limit the development of AI-aided diagnostic tools, as new medical models still have a bright future.

In conclusion, the AI-based diagnostic algorithm is a valid and efficient method for the diagnosis of TPF. It can serve as a useful intelligent assistant for orthopedic physicians, which can streamline clinical workflow and help to guarantee the health and security of patients.

Conflict of Interest Statement

The authors declare that there are no conflicts of interest relevant to this article.

REFERENCES

1   Schulak DJ, Gunn DR. Fractures of tibial plateaus. A review of the literature. Clin Orthop Relat Res, 1975(109):166-177
2   Cho JW, Kim J, Cho WT, *et al*. Approaches and fixation of the posterolateral fracture fragment in tibial plateau fractures: a review with an emphasis on rim plating via modified anterolateral approach. Int Orthop, 2017,41(9): 1887-1897
3   Xie X, Zhan Y, Wang Y, *et al*. Comparative Analysis of Mechanism-Associated 3-Dimensional Tibial Plateau

Fracture Patterns. J Bone Joint Surg Am, 2020,102(5): 410-418

4 Hofmann A, Gorbulev S, Guehring T, *et al*. Autologous Iliac Bone Graft Compared with Biphasic Hydroxyapatite and Calcium Sulfate Cement for the Treatment of Bone Defects in Tibial Plateau Fractures: A Prospective, Randomized, Open-Label, Multicenter Study. J Bone Joint Surg Am, 2020,102(3):179-193

5 Guly HR. Diagnostic errors in an accident and emergency department. Emerg Med J, 2001,18(4):263-269

6 Berlin L. Defending the "missed" radiographic diagnosis. AJR Am J Roentgenol, 2001,176(2):317-322

7 Pinto A, Reginelli A, Pinto F, *et al*. Errors in imaging patients in the emergency setting. Br J Radiol, 2016, 89(1061):20150914

8 Myers TG, Ramkumar PN, Ricciardi BF, *et al*. Artificial Intelligence and Orthopaedics: An Introduction for Clinicians. J Bone Joint Surg Am, 2020,102(9):830-840

9 Wang S, Yang DM, Rong R, *et al*. Pathology Image Analysis Using Segmentation Deep Learning Algorithms. Am J Pathol, 2019,189(9):1686-1698

10 Ding Z, Shi H, Zhang H, *et al*. Gastroenterologist-Level Identification of Small-Bowel Diseases and Normal Variants by Capsule Endoscopy Using a Deep-Learning Model. Gastroenterology, 2019,157(4):1044-1054

11 Nguyen DT, Pham TD, Batchuluun G, *et al*. Artificial Intelligence-Based Thyroid Nodule Classification Using Information from Spatial and Frequency Domains. J Clin Med, 2019,8(11):1976

12 Chen K, Zhai X, Sun K, *et al*. A narrative review of machine learning as promising revolution in clinical practice of scoliosis. Ann Transl Med, 2021,9(1):67

13 Norman B, Pedoia V, Noworolski A, *et al*. Applying Densely Connected Convolutional Neural Networks for Staging Osteoarthritis Severity from Plain Radiographs. J Digit Imaging, 2019,32(3):471-477

14 Pedoia V, Norman B, Mehany SN, *et al*. 3D convolutional neural networks for detection and severity staging of meniscus and PFJ cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects. J Magn Reson Imaging, 2019,49(2):400-410

15 Yao X, Zhou K, Lv B, *et al*. 3D mapping and classification of tibial plateau fractures. Bone Joint Res, 2020,9(6):258-267

16 Samsami S, Patzold R, Winkler M, *et al*. The effect of coronal splits on the structural stability of bi-condylar tibial plateau fractures: a biomechanical investigation. Arch Orthop Trauma Surg, 2020,140(11):1719-1730

17 Wang Y, Wang J, Tang J, *et al*. Arthroscopy Assisted Reduction Percutaneous Internal Fixation versus Open Reduction Internal Fixation for Low Energy Tibia Plateau Fractures. Sci Rep, 2018,8(1):14068

18 Yoo H, Kim KH, Singh R, *et al*. Validation of a Deep Learning Algorithm for the Detection of Malignant Pulmonary Nodules in Chest Radiographs. JAMA Netw Open, 2020,3(9):e2017135

19 Wang M, Xia C, Huang L, *et al*. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. Lancet Digit Health, 2020,2(10):e506-e515

20 Gan K, Xu D, Lin Y, *et al*. Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. Acta Orthop, 2019,90(4):394-400

21 Lindsey R, Daluiski A, Chopra S, *et al*. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci USA, 2018,115(45):11591-11596

22 Choi JW, Cho YJ, Lee S, *et al*. Using a Dual-Input Convolutional Neural Network for Automated Detection of Pediatric Supracondylar Fracture on Conventional Radiography. Invest Radiol, 2020,55(2):101-110

23 Chung SW, Han SS, Lee JW, *et al*. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop, 2018,89(4): 468-473

24 Gorelik N, Gyftopoulos S. Applications of Artificial Intelligence in Musculoskeletal Imaging: From the Request to the Report. Can Assoc Radiol J, 2021,72(1): 45-59

25 Garwood ER, Tai R, Joshi G, *et al*. The Use of Artificial Intelligence in the Evaluation of Knee Pathology. Semin Musculoskelet Radiol, 2020,24(1):21-29

26 Watanabe K, Aoki Y, Matsumoto M. An Application of Artificial Intelligence to Diagnostic Imaging of Spine Disease: Estimating Spinal Alignment From Moire Images. Neurospine, 2019,16(4):697-702