ORIGINAL RESEARCH

Inter- and intra-observer reliability of injury diagnosis for peri-knee fractures: A comparison between two- and three-dimensional CT imaging

Hui Zhang[1], Patrick Manda[2], Brave Kadoko Nyirenda[2], Blessed Kondowe[2], Wenjing Wang[3], Jin Shang[1]*

1. Department of Medical Imaging, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, 710061, China
2. Radiology Department, Mzuzu Central Hospital, Mzuzu, Malawi
3. Department of Surgery Intensive Care Unit, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, 710061, China

***Corresponding Author:** Jin Shang; E-mail: shangjin01@qq.com

Abstract

Objective

This study aimed to assess whether three-dimensional (3D) CT imaging improves the inter- and intra-observer reliability of peri-knee fracture classifications, compared to two-dimensional (2D) CT imaging.

Methods

A retrospective analysis was conducted on 23 patients with peri-knee fractures, using both 2D and 3D-CT scans. Three radiologists classified distal femur, patella, and tibial plateau fractures according to Arbeitsgemeinschaft für Osteosynthesefragen/Orthopaedic Trauma Association (AO/OTA) and Schatzker systems. Reliability was measured using Cohen's kappa, with evaluations conducted at two separate intervals to assess intra- and inter-observer consistency.

Results

The intra-observer reliability for 2D-CT was substantial for distal femur ($\varkappa = 0.737$, IQR 0.615–0.788) and tibial plateau ($\varkappa = 0.732$, IQR 0.615–0.819) fractures, improving slightly with 3D-CT ($\varkappa = 0.775$, IQR 0.658–0.869; $\varkappa = 0.768$, IQR 0.628–0.882 respectively). Patella fracture classification showed almost perfect reliability ($\varkappa = 0.823$, IQR 0.707–0.882) with 2D-CT, further improving with 3D-CT ($\varkappa = 0.865$, IQR 0.764–0.951). However, inter-observer reliability showed no significant improvement with the addition of 3D-CT across all fracture types.

Conclusion

While 3D-CT marginally enhances intra-observer reliability for peri-knee fractures, the difference in inter-observer reliability compared to 2D-CT was not statistically significant.

Key words: Three-dimensional; Two-dimensional; Peri-knee fractures; Observer reliability

Introduction

Peri-knee fractures, including fractures of the distal femur, tibial plateau, and patella, have become common injuries resulting from road traffic accidents in Malawi in recent years[1]. These injuries are often complex, involving articular surfaces, and place a significant economic burden on both patients and the healthcare system in Malawi[2-5]. Preoperative planning for peri-knee fractures is critical, as the choice of surgical incision must be carefully determined based on the fracture type. Therefore, radiologists must provide orthopedic surgeons with detailed descriptions of peri-knee fractures, particularly those involving the distal femur, tibial plateau, and patella, as each type of fracture requires specific management considerations[1].

A knee trauma series of X-rays remains the most appropriate primary imaging modality for screening patients with knee pain, swelling, or deformity[6,7]. Radiographic features such as lipohemarthrosis, widening of the joint space, asymmetry, or incongruity may suggest intra-articular injuries[6]. However, certain fracture patterns may be easily missed on standard X-rays. A dedicated non-contrast CT scan of the knee is recommended for assessing intra-articular extension; CT should also be performed in patients with lipohemarthrosis identified on lateral decubitus X-rays, as well as in those who are unable to bear weight and are clinically suspected of having fractures that are not visible on X-rays[6].

Several studies have evaluated the inter-observer and intra-observer reliability of tibial plateau fracture classifications using two-dimensional (2D) CT and three-dimensional (3D) CT[8-10]. However, tibial plateau fractures are not isolated injuries within the context of peri-knee fractures and are often accompanied by fractures of the distal femur and patella. Currently, no studies have assessed the impact of 2D and 3D CT imaging on the inter-observer and intra-observer reliability of classification systems for peri-knee fractures. Therefore, this study aimed to explore whether 3D reconstructed CT images improve the inter-observer and intra-observer reliability.

Materials and methods

Study Methodology and Radiology

In this retrospective study, we selected 23 consecutive patients with peri-knee fractures who were treated at Mzuzu Central Hospital (MCH) following traffic accidents between November 2023 and August 2024. This study was approved by Mzuzu University Research Ethics Committee

(MZUNIREC) (Approval Number: MZUNIREC/DOR/24/153), and consent in written form was waived.

All patients underwent non-contrast knee CT scans using a Neusoft 16-slice spiral CT scanner. The CT scan settings included an appropriate field of view (FOV), a tube voltage of 120 kV, a tube current of 225 mA, and a rotation time of 0.6 s/r. The original data were acquired with a slice thickness of 16 mm × 0.625 mm. Continuous scans were performed with a slice thickness of 5 mm, with no gaps, at a pitch of 0.9. The raw data were transferred to the AVW workstation for post-processing, where both bone algorithm reconstruction and moderately smoothed soft tissue algorithm reconstruction were performed, with a reconstruction slice thickness of 1 mm and an interslice spacing of 0.625 mm. Bone window settings included a window level of 300 - 700 HU and a window width of 1500 - 3000 HU, while soft tissue window settings used a window level of 35 - 45 HU and a window width of 300 - 400 HU. Three-dimensional (3D) images were reconstructed at arbitrary rotation angles as needed.

Three radiologists with varying levels of experience served as observers. They were tasked with classifying distal femur and patella fractures according to the AO/OTA (Arbeitsgemeinschaft für Osteosynthesefragen/Orthopaedic Trauma Association) comprehensive classification system, and tibial plateau fractures according to the Schatzker classification system[11,12]. During each evaluation session, charts of the classification systems and relevant descriptions from the original publications were provided. The images were assessed in a blinded and randomized manner, with all identifying labels on the images obscured to minimize observer bias. Two rounds of evaluations were conducted: first, the 2D-CT images were assessed, followed by a second evaluation two weeks later in which both 2D and 3D images were assessed in combination.

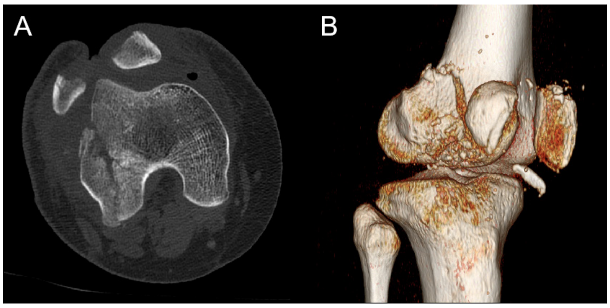

Figure 1 The axial CT image (A) shows a right intracondylar distal femur fracture (Arbeitsgemeinschaft für Osteosynthesefragen–Orthopedic Trauma Association [AO/OTA] type B1) in a 23-year-old woman who was involved in a motor vehicle collision. The three-dimensional CT image of the right knee (B) shows a severely comminuted stellate fracture of the patella (AO/OTA type C3)

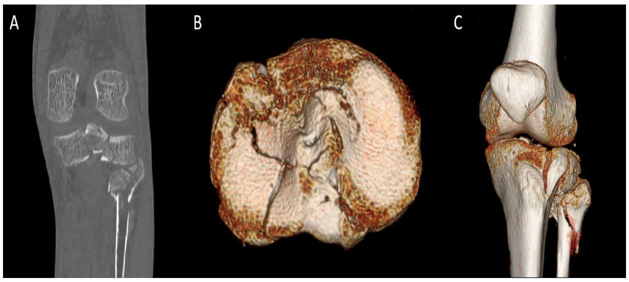

Figure 2 Left bicondylar fractures (Schatzker type V) in a 47-year-old man who was struck as a pedestrian by a moving vehicle. Coronal (A) and 3D (B) CT images show tibial spine involvement; the 3D-CT image (C) also shows a comminuted fracture of the fibular head

The use of 3D images alone was not analyzed, as they are typically not used without the availability of 2D images. After a two-week interval, the evaluations were repeated in a new random sequence to assess intra-observer reliability.

Statistical analysis

Cohen's kappa values were calculated to determine the reliability of an observer in classifying fractures on two separate occasions (intra-observer reliability) or by different observers on the same occasion (inter-observer reliability)[13,14]. The kappa value is a chance-corrected measure of agreement, comparing the observed level of agreement with the level expected by chance alone. Levels of agreement between and within observers were categorized according to the criteria for clinical diagnosis agreement described by Landis and Koch: $\varkappa < 0$, poor; 0.0 to 0.2, slight; 0.21 to 0.4, fair; 0.41 to 0.6, moderate; 0.61 to 0.8, substantial; and 0.81 to 1.0, almost perfect[15]. We used the method described by Doornberg et al. to determine the reliability between the upper and lower boundaries of the interquartile ranges (IQR) for intra-observer reliability and the 95% confidence interval for inter-observer reliability[16]. The smaller the gap between these boundaries, the more significant the statistical argument.

For the analyses, SPSS 22 (IBM, Armonk, New York) and Microsoft Excel 2016 (Microsoft Corporation, Redmond, Washington) were used.

Results

Classification Systems (Tables 1 and 2)

Distal Femur: AO/OTA Classification

When observers used 2D-CT, the intra-observer reliability for classification based on the AO/OTA subtypes (A.1-3, B.1-3, or C.1-3) was substantial (average $\varkappa2D = 0.737$, IQR 0.615–0.788). After the addition of 3D reconstruction, the reliability slightly improved, reaching $\varkappa3D = 0.775$ (IQR 0.658–0.869).

However, the addition of 3D-CT did not enhance the inter-observer reliability of the AO/OTA classification system for distal femur fractures. For classification by subtype (A.1-3, B.1-3, or C.1-3), the inter-observer reliability using 2D-CT was substantial ($\varkappa2D = 0.637$, 95% CI: 0.581–0.691, $P < 0.0001$), and there was no change after the addition of 3D reconstruction, as rated by the Landis and Koch classification ($\varkappa3D = 0.646$, 95% CI: 0.578–0.712, $P < 0.0001$).

Patella: AO/OTA Classification

When observers used 2D-CT for classification based on the AO/OTA subtypes (A, B, or C.1-3), the average intra-observer reliability was almost perfect (mean $\varkappa2D = 0.823$, IQR 0.707–0.882). After the addition of 3D reconstructions, the reliability slightly improved to $\varkappa3D = 0.865$ (IQR 0.764–0.951).

However, the addition of 3D-CT did not improve the inter-observer reliability of the AO/OTA classification system for patellar fractures. For classification based on subtypes (A, B, or C.1-3), the inter-observer reliability using 2D-CT was also almost perfect ($\varkappa_2D = 0.836$, 95% CI: 0.781–0.991, $P < 0.0001$), and there was no change following the addition of 3D reconstructions according to the Landis and Koch classification rating ($\varkappa3D = 0.845$, 95% CI: 0.778–0.912, $P < 0.0001$).

Tibial Plateau: Schatzker Classification

When using 2D-CT, the intra-observer reliability based on

Table 1 Intra-observer reliability of 2D and 3D computed tomography for peri-knee fractures classification

	2D imaging round 1 (kappa)	Category	3D imaging round 2 (kappa)	Category
Distal femur: AO/OTA classification				
1	0.875	Almost perfect	0.835	Almost perfect
2	0.657	Substantial	0.771	Substantial
3	0.679	Substantial	0.719	Substantial
Average	0.737	Substantial	0.775	Substantial
Patella: AO/OTA classification				
1	0.936	Almost perfect	0.955	Almost perfect
2	0.728	Substantial	0.783	Substantial
3	0.805	Almost perfect	0.857	Almost perfect
Average	0.823	Almost perfect	0.865	Almost perfect
Tibial plateau: Schatzker classification				
1	0.794	Substantial	0.823	Almost perfect
2	0.538	Moderate	0.622	Substantial
3	0.864	Almost perfect	0.859	Almost perfect
Average	0.732	Substantial	0.768	Substantial
Peri-knee fractures classification				
Average	0.764	Substantial	0.803	Almost perfect

Table 2 Inter-observer reliability of 2D and 3D computed tomography for peri-knee fractures classification

Classification	2D imaging round 1 (kappa)	95% Confidence interval	Category	3D imaging round 2 (kappa)	95% Confidence interval	Category	Significance of change in kappa value
Distal femur (AO/OTA)	0.637	0.581-0.691	Substantial	0.646	0.578-0.712	Substantial	NS
Patella (AO/OTA)	0.836	0.781-0.991	Almost perfect	0.845	0.778-0.912	Almost perfect	NS
Tibial plateau (Schatzker)	0.645	0.594-0.696	Substantial	0.696	0.638-0.754	Substantial	NS
Peri-knee fractures							
Average	0.706		Substantial	0.729		Substantial	

the Schatzker classification was substantial, with an average kappa coefficient of 0.732 (IQR 0.615–0.819). After the addition of 3D reconstruction, the intra-observer reliability remained substantial ($\varkappa 3D = 0.768$, IQR 0.628–0.882).

For inter-observer reliability across all six Schatzker classification types (I-VI) assessed with 2D-CT, it was moderate ($\varkappa 2D = 0.645$, 95% CI: 0.594–0.696, $P < 0.0001$). With the addition of 3D images, the inter-observer reliability improved to $\varkappa 3D = 0.696$ (95% CI: 0.638–0.754, $P < 0.0001$). However, as the 95% confidence intervals overlapped, the difference was not considered statistically significant.

Discussion

The knee joint is a complex structure primarily involving internal and external rotation, anterior-posterior sliding, and flexion-extension movements. It serves crucial functions related to motion and weight-bearing. The upper and lower ends of the bones in the knee joint are composed of cancellous bone, and the surrounding soft tissue is limited[17]. These anatomical features make the joint susceptible to both direct and indirect trauma. With the increasing incidence of traffic accidents and a broader range of activities, high-energy injuries have become more frequent, leading to significant changes in trauma mechanisms. Therefore, for trauma patients, appropriate and scientifically sound auxiliary examinations upon hospital admission are essential to prevent misdiagnosis and missed diagnoses, ultimately improving the detection rate of fractures around the knee joint[18].

Currently, the imaging evaluation of fractures around the knee joint is primarily conducted using three commonly applied methods: plain X-rays, CT, and MRI. Plain X-rays are the first choice for assessing fractures around the knee joint. MSCT offers high-density resolution and can reveal soft tissue abnormalities that may not be detected on plain films. CT also clearly defines the location, boundaries, and extent of fractures[19]. With its high image resolution and lack of overlapping structures, CT allows for the visualization of intra-articular and joint capsule injuries that are difficult to detect on plain films. Contrast-enhanced scans provide valuable information regarding vascular status and blood supply, while 3D-CT enables a three-dimensional display of bone and joint trauma. This modality provides a multi-directional view, offering crucial guidance for clinicians in treatment planning, and has been widely adopted in clinical practice. In a study by Suero EM et al. on five tibial plateau fractures, the authors noted that preoperative 3D reconstruction was a valuable tool for fracture segmentation and was a key factor in preoperative planning[20]. Compared to CT, MRI provides a comprehensive evaluation of soft tissue, including ligaments, menisci, cartilage, and tendons. However, MRI is not routinely used and is not recommended before the fixation of joint fractures. Typically, MRI is appropriate for planning ligament reconstruction after the fracture has been stabilized[21].

Preoperative evaluation of fracture classification plays a crucial role in treatment planning. Important factors such as fracture displacement, the presence of free intra-articular bone fragments, and the degree of articular surface depression provide essential information for clinicians[8,22]. This information enables thorough preoperative preparation, reduces operative time, improves treatment outcomes and patient satisfaction, and lowers disability rates. Studies on inter-observer and intra-observer reliability of fracture classification systems have gained widespread acceptance, with the Schatzker and AO classification systems being the most commonly used[13]. The reliability of fracture classification and grading has evolved from conventional X-rays to 2D and 3D reconstructed CT scans, as well as MRI scans[16,23]. In a study by Hu et al. on 21 tibial plateau fractures, the authors reported that 3D-CT scans were more reliable than 2D CT scans[23]. However, the study did not specify whether the added value of 3D imaging was statistically significant, and the research was limited to the classification of tibial plateau fractures rather than peri-knee fractures. Doornberg et al., in their study of 45 patients, compared CT scans with both 2D and 3D reconstruction modes[16]. They concluded that there was no added value of 3D-CT after performing a 2D-CT scan. According to their findings, 3D-CT scans did not significantly improve the reliability of characterizing and classifying tibial plateau fractures.

Our study is the first to integrate a classification system for peri-knee fractures to evaluate intra-observer reproducibility and inter-observer reliability and to define statistically significant differences between 2D and 3D-CT images. We found that adding 3D-CT did not improve the average intra-observer reliability for the distal femur AO classification, patella AO classification, and tibial plateau Schatzker classification. However, it did increase the average intra-observer reliability for the overall peri-knee fracture classification system from substantial ($\varkappa 2D = 0.764$) to almost perfect ($\varkappa 3D = 0.803$) (Table 1). Adding 3D-CT did not alter the average inter-observer reliability for the distal femur AO classification, patella AO classification, tibial plateau Schatzker classification, or the peri-knee fracture classification system (Table 2). In other words, compared to 2D-CT, 3D-CT helps radiologists feel more confident in identifying peri-knee fracture classifications, thus providing more accurate preoperative planning for clinicians. This is because, on 2D images, it is difficult to track individual fragments from one image to another, and no single 2D scan can effectively depict the entire articular surface due to its division across different slices. In our clinical practice, we have found that interpreting 2D images can sometimes be confusing for orthopedic surgeons. 3D-CT images may be easier to interpret, as they provide a clearer view of the articular surface, fracture complexity, and the spatial relationship of the fragments[24,25].

Conclusion

In conclusion, the routine use of 3D-CT in addition to 2D-CT does improve the classification of peri-knee fractures based on absolute improvements in kappa values, but the difference between 2D-CT and 3D-CT imaging in classifying peri-knee fractures was not statistically significant.

Acknowledgements

This work was supported by the Natural Science Foundation of Shaanxi Province (No. 2023-JC-QN-0979 and No. 2023-JC-QN-0898).

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval

This study was approved by Mzuzu University Research Ethics Committee (MZUNIREC/DOR/24/153).

Data availability statement

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

References

1.Dreizin D, Edmond T, Zhang T, Sarkar N, Turan O, Nascone J. CT of Periarticular Adult Knee Fractures: Classification and Management Implications. Radiographics. 2024 Sep;44(9):e240014. doi: 10.1148/rg.240014.

2.Schade AT, Khatri C, Nwankwo H, Carlos W, Harrison WJ, Metcalfe AJ. The economic burden of open tibia fractures: A systematic review. Injury. 2021 Jun;52(6):1251-1259. doi: 10.1016/j.injury.2021.02.022.

3.Mwafulirwa K, Munthali R, Ghosten I, Schade A. Epidemiology of Open Tibia fractures presenting to a tertiary referral centre in Southern Malawi: a retrospective study. Malawi Med J. 2022 Jun;34(2):118-122. doi: 10.4314/mmj.v34i2.7.

4.Schade AT, Sabawo M, Nyamulani N, Mpanga CC, Ngoie LB, Metcalfe AJ et al. Functional outcomes and quality of life at 1-year follow-up after an open tibia fracture in Malawi: a multicentre, prospective cohort study. Lancet Glob Health. 2023 Oct;11(10):e1609-e1618. doi: 10.1016/S2214-109X(23)00346-7.

5.Schade AT, Nyamulani N, Banza LN, Metcalfe AJ, Leather A, Madan JJ et al. Protocol for a prospective cohort study of open tibia fractures in Malawi with a nested implementation of open fracture guidelines. Wellcome Open Res. 2021 Sep 13;6:228. doi: 10.12688/wellcomeopenres.17145.1.

6.Taljanovic MS, Chang EY, Ha AS, Bartolotta RJ, Bucknor M, Chen KC et al. Expert Panel on Musculoskeletal Imaging; ACR Appropriateness Criteria® Acute Trauma to the Knee. J Am Coll Radiol. 2020 May;17(5S):S12-S25. doi: 10.1016/j.jacr.2020.01.041.

7.Stiell IG, Wells GA, Hoag RH, Sivilotti ML, Cacciotti TF, Verbeek PR et al. Implementation of the Ottawa Knee Rule for the use of radiography in acute knee injuries. JAMA. 1997 Dec 17;278(23):2075-

9.

8.Patange Subba Rao SP, Lewis J, Haddad Z, Paringe V, Mohanty K. Three-column classification and Schatzker classification: a three- and two-dimensional computed tomography characterisation and analysis of tibial plateau fractures. Eur J Orthop Surg Traumatol. 2014 Oct;24(7):1263-70. doi: 10.1007/s00590-013-1308-9.

9.Brunner A, Horisberger M, Ulmar B, Hoffmann A, Babst R. Classification systems for tibial plateau fractures; does computed tomography scanning improve their reliability? Injury. 2010 Feb;41(2):173-8. doi: 10.1016/j.injury.2009.08.016. Epub 2009 Sep 9.

10.Lazaro LE, Wellman DS, Pardee NC, Gardner MJ, Toro JB, Macintyre NR 3rd, Helfet DL, Lorich DG. Effect of computerized tomography on classification and treatment plan for patellar fractures. J Orthop Trauma. 2013 Jun;27(6):336-44. doi: 10.1097/BOT.0b013e318270dfe7.

11.Müller ME, Nazarian S, Koch P, Schatzker J. The comprehensive classification of fractures of long bones. Springer Science & Business Media, 2012.

12.Schatzker J, McBroom R, Bruce D. The tibial plateau fracture. The Toronto experience 1968--1975. Clin Orthop Relat Res. 1979 Jan-Feb;(138):94-104.

13.Maripuri SN, Rao P, Manoj-Thomas A, Mohanty K. The classification systems for tibial plateau fractures: how reliable are they? Injury. 2008 Oct;39(10):1216-21. doi: 10.1016/j.injury.2008.01.023.

14. Wainwright AM, Williams JR, Carr AJ. Interobserver and intraobserver variation in classification systems for fractures of the distal humerus. J Bone Joint Surg Br. 2000 Jul;82(5):636-42. doi: 10.1302/0301-620x.82b5.9673.

15.Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 Mar;33(1):159-74.

16.Doornberg JN, Rademakers MV, van den Bekerom MP, Kerkhoffs GM, Ahn J, Steller EP et al. Two-dimensional and three-dimensional computed tomography for the classification and characterisation of tibial plateau fractures. Injury. 2011 Dec;42(12):1416-25. doi: 10.1016/j.injury.2011.03.025. Epub 2011 May 13.

17.Rudran B, Little C, Wiik A, Logishetty K. Tibial Plateau Fracture: Anatomy, Diagnosis and Management. Br J Hosp Med (Lond). 2020 Oct 2;81(10):1-9. doi: 10.12968/hmed.2020.0339. Epub 2020 Oct 30.

18.Wang Y, Luo C, Zhu Y, Zhai Q, Zhan Y, Qiu W, Xu Y. Updated Three-Column Concept in surgical treatment for tibial plateau fractures - A prospective cohort study of 287 patients. Injury. 2016 Jul;47(7):1488-96. doi: 10.1016/j.injury.2016.04.026. Epub 2016 May 2.

19.Lôbo CFT, Helito PVP, Bordalo-Rodrigues M, Helito CP. Computed tomography (CT), X-ray, and MRI evaluation of two anterolateral knee reconstruction techniques: lateral extra-articular tenodesis (LET) and the anterolateral ligament (ALL) reconstruction. Skeletal Radiol. 2020 Jul;49(7):1037-1049. doi: 10.1007/s00256-020-03402-2. Epub 2020 Mar 9.

20.Suero EM, Hüfner T, Stübig T, Krettek C, Citak M. Use of a virtual 3D software for planning of tibial plateau fracture reconstruction. Injury. 2010 Jun;41(6):589-91. doi: 10.1016/j.injury.2009.10.053. Epub 2009 Nov 25.

21. Ye QL, Li J, Hou JD, Cai L, Xiong X, Wang HJ. Comparison of multislice spiral CT and MRI in diagnosis of occult fracture of knee joint with meniscus and ligament injury. Zhongguo Gu Shang. 2022 Oct 25;35(10):967-70. Chinese. doi: 10.12200/j.issn.1003-0034.2022.10.011.

22.Zhu Y, Hu CF, Yang G, Cheng D, Luo CF. Inter-observer reliability assessment of the Schatzker, AO/OTA and three-column classification of tibial plateau fractures. J Trauma Manag Outcomes. 2013 Sep 11;7(1):7. doi: 10.1186/1752-2897-7-7.

23.Hu YL, Ye FG, Ji AY, Qiao GX, Liu HF. Three-dimensional computed tomography imaging increases the reliability of classification systems for tibial plateau fractures. Injury. 2009 Dec;40(12):1282-5. doi: 10.1016/j.injury.2009.02.015. Epub 2009 Jun 16.

24.Millar SC, Arnold JB, Thewlis D, Fraysse F, Solomon LB. A systematic literature review of tibial plateau fractures: What classifications are used and how reliable and useful are they? Injury. 2018 Mar;49(3):473-490. doi: 10.1016/j.injury.2018.01.025. Epub 2018 Jan 31.

25.Wicky S, Blaser PF, Blanc CH, Leyvraz PF, Schnyder P, Meuli RA. Comparison between standard radiography and spiral CT with 3D reconstruction in the evaluation, classification and management of tibial plateau fractures. Eur Radiol. 2000;10(8):1227-32. doi: 10.1007/s003300000326.