

Evaluation of the Schatzker-Kfuri Classification of Tibial Plateau Fractures Using Radiographs and Computed Tomography: Comparison Between an Expert Observer and the ChatGPT-4o Model

Héctor A. Rivadeneira Jurado,* Elías A. Rivadeneira Jurado,* Daniel Espinoza Freire,* Andrés F. Samaniego,* Ezequiel Lulkin,* Fernando Bidolegui,** Sebastián Pereira*

*Orthopedics and Traumatology Service, Hospital Sirio-Libanés, Autonomous City of Buenos Aires, Argentina

**Orthopedics and Traumatology Service, Sanatorio Otamendi Miroli, Autonomous City of Buenos Aires, Argentina

ABSTRACT

Introduction: Artificial intelligence was formally introduced in 1956, and since then, platforms trained on large datasets have been developed to generate increasingly accurate outputs. The Kfuri-Schatzker classification of tibial plateau fractures enables more precise analysis, particularly when CT imaging is integrated. This study compared the diagnostic accuracy of the ChatGPT-4o model with that of expert evaluators. **Materials and Methods:** A retrospective observational study was conducted to compare the interpretations of an expert observer with those generated by ChatGPT-4o. A dataset of 45 expert-published case reports including radiographs and CT scans from databases such as PubMed, Elsevier, and SciELO was used to refine the prompt guiding ChatGPT-4o's analysis. Six additional case reports of tibial plateau fractures, none previously provided to the model, were selected for evaluation. ChatGPT-4o analyzed each case and proposed a classification according to the Schatzker-Kfuri system. Its responses were compared with the expert diagnoses reported in the literature. **Results:** ChatGPT-4o correctly classified all the cases analyzed. In bicondylar fractures, the model accurately identified components of subsidence, shear (split) pattern, and epiphyseal-diaphyseal dissociation. Cohen's kappa coefficient was 1.00, indicating perfect agreement.

Conclusion: The ChatGPT-4o model demonstrated high diagnostic accuracy in classifying tibial plateau fractures using the Schatzker-Kfuri system, achieving agreement comparable to that of an expert evaluator.

Keywords: Artificial intelligence; tibial plateau; Schatzker-Kfuri classification.

Level of Evidence: III

Evaluación de la clasificación de las fracturas de platillo tibial según Schatzker-Kfuri utilizando radiografías y tomografía. Comparación entre el observador experto y el modelo ChatGPT-4o

RESUMEN

Introducción: La inteligencia artificial fue presentada formalmente en 1956, luego, se crearon plataformas con un conjunto de información para obtener el resultado apropiado. La clasificación de fracturas de platillo tibial de Kfuri y Schatzker permite hacer un análisis más preciso, especialmente al integrar cortes tomográficos. En este estudio, se comparó la capacidad diagnóstica del modelo ChatGPT-4o con la evaluación del panel de expertos. **Materiales y Métodos:** Estudio retrospectivo, observacional para comparar la interpretación del observador experto y la del ChatGPT-4o. Se recopilaron 45 reportes de caso publicados por expertos con radiografías y tomografías, en distintas bases de datos, como PubMed, Elsevier, SciELO, que se usaron para mejorar el análisis del ChatGPT-4o. Se seleccionaron 6 reportes de caso de fractura de platillo tibial, que no se habían cargado previamente en la plataforma para analizar la interpretación del ChatGPT-4o en base al Prompt creado antes. El modelo ChatGPT-4o analizó cada uno de los casos y propuso una clasificación basada en el sistema de Schatzker-Kfuri. Las respuestas fueron contrastadas con la información obtenida de reportes de casos. **Resultados:** El ChatGPT-4o clasificó correctamente los casos analizados. Los componentes de hundimiento, trazo de cizallamiento (*split*) y disociación epifisodiafisaria fueron identificados, con precisión, en

Received on September 5th, 2025. Accepted after evaluation on November 21st, 2025 • Dr. HÉCTOR A. RIVADENEIRA JURADO • 1bhrivadeneirajurado@gmail.com  <https://orcid.org/0009-0008-6397-9718>

How to cite this article: Rivadeneira Jurado HA, Rivadeneira Jurado EA, Espinoza Freire DE, Samaniego AF, Lulkin E, Bidolegui F, Pereira S. Evaluation of the Schatzker-Kfuri Classification of Tibial Plateau Fractures Using Radiographs and Computed Tomography: Comparison Between an Expert Observer and the ChatGPT-4o Model. *Rev Asoc Argent Ortop Traumatol* 2025;90(6):556-560. <https://doi.org/10.15417/issn.1852-7434.2025.90.6.2224>

los casos bicondilares. Asimismo, se utilizaron medidas de concordancia kappa de Cohen: 1.00, lo cual se interpreta como concordancia perfecta. **Conclusión:** El ChatGPT-4o tuvo una alta capacidad diagnóstica en la clasificación de fracturas de platillo tibial según Schatzker-Kfuri, equiparable a la de un experto.

Palabras clave: Inteligencia artificial; platillo tibial; clasificación de Schatzker-Kfuri.

Nivel de Evidencia: III I

INTRODUCTION

Artificial intelligence (AI) was formally introduced in 1956.¹ Over the years, increasingly sophisticated computer programs have been developed for use in various fields, including orthopedics and traumatology. However, current platforms require an appropriate prompt or set of information to produce accurate outputs.²

In traumatology and orthopedics, tibial plateau fractures represent a significant diagnostic and therapeutic challenge. The Schatzker classification, created in 1979 and widely used in orthopedic practice, categorizes fractures of the tibial plateau. More recently, a three-dimensional evaluation model based on computed tomography (CT) was developed to better define the anatomical involvement of plateau quadrants, giving rise to the Schatzker-Kfuri classification in 2018.³ This system allows differentiation between unicondylar, bicondylar, and epiphyseal-diaphyseal dissociation fractures and has improved surgical planning.

The objective of this study was to compare the ability of the multimodal language model ChatGPT-4o to classify tibial plateau fractures using both radiographs and CT images with that of expert case reports published in the literature.

MATERIALS AND METHODS

A retrospective, observational study was conducted to compare the interpretation of an expert observer with that of ChatGPT-4o. To create the prompt, 45 case reports published in databases such as PubMed, Elsevier, and SciELO were included. These reports contained anteroposterior and lateral knee radiographs and axial, coronal, and sagittal CT scans of the knee. Case reports with incomplete CT series or without complete radiographs were excluded. Thus, the 45 expert-validated case reports with radiographs and CT scans were used to improve the accuracy of ChatGPT-4o's interpretive performance. Prior to uploading, the images were organized in the following order: anteroposterior knee radiograph, lateral radiograph, and axial, coronal, and sagittal CT slices of the tibial plateau (Figure).

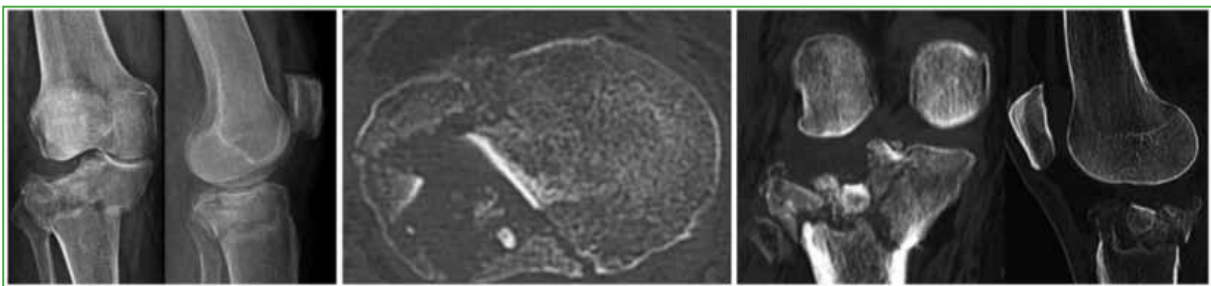


Figure. Sequence of radiographic and CT images used for interpretation by the platform.

Additional content was incorporated into the prompt, including anatomical descriptions, basic traumatology concepts, examples of split or shear fractures, depression patterns, combined fracture mechanisms, and fractures with epiphyseal-diaphyseal extension.

Descriptive information and corresponding illustrations were progressively uploaded until the AI model prompt was complete. Subsequently, 45 case reports in DICOM format were uploaded to further refine the model's interpretive capability. Finally, six expert case reports not previously included were used for the evaluation phase. Each of the six reports represented a different fracture pattern included in the classification system.

ChatGPT-4o sequentially analyzed each image set and proposed a classification based on the Schatzker–Kfuri system. The proposed classification was recorded and compared with the reference classification documented in the original case reports. A classification was considered correct when it matched the expert-reported classification exactly.

RESULTS

All six cases were correctly classified by the model. The following fracture patterns were accurately identified:

Pure depression (type III)

Lateral split/shear (type I)

Bicondylar fracture without dissociation (type V)

Epiphysiodiaphyseal dissociation (type VI)

Medial column involvement (type IV)

Depression with lateral split (type II)

A summary of the comparison is shown in the [Table](#).

The analysis of the six expert-described case reports compared with the ChatGPT-4o interpretation demonstrated complete agreement based on both radiographic and three-dimensional assessment. Importantly, the study yielded a Cohen's kappa coefficient of 1.00, which is interpreted as perfect agreement.

Table. Results of expert case reports compared with the Schatzker-Kfuri classification generated by ChatGPT-4o.

Case	Schatzker classification	Schatzker-Kfuri classification	ChatGPT-4o classification	Affected quadrants	Concordance
1	IV	IV - PM	Schatzker IV	PM	Correct
2	II	II - AL	Schatzker II	AL	Correct
3	V	V - AL + PM	Schatzker V	AL + PM	Correct
4	I	I - AL (pure <i>split</i>)	Schatzker I	AL (pure <i>split</i>)	Correct
5	VI	VI - AL + PM + (epiphyseal-diaphyseal dissociation)	Schatzker VI	AL + PM + epiphyseal-diaphyseal dissociation	Correct
6	III	III - AL (pure depression)	Schatzker III	AL (pure depression)	Correct

AL = anterolateral; PM = posteromedial.

DISCUSSION

The results of this study are consistent with recent publications demonstrating the growing potential of artificial intelligence in the diagnosis of articular fractures. In particular, studies by Mohammadi et al.⁴ and Van der Gaast et al.⁵ have shown that models such as ChatGPT-4o can achieve accuracy levels comparable to those of specialist radiologists in interpreting radiographs. Similar findings have been reported in resource-limited settings, where the addition of three-dimensional CT reconstruction significantly improved diagnostic interpretation, as described by Markhardt et al.⁶ Furthermore, recent reviews on AI in orthopedic surgery emphasize the need for comparative studies against expert clinicians to establish the validity of AI-based image interpretation. Gyftopoulos et al.,⁷ and Kuo et al.⁸ evaluated the predictive performance of deep-learning models for classifying tibial plateau fractures

and demonstrated promising applicability in real clinical scenarios. Additional contributions from Giordano et al.,⁹ Singh Sidhu et al.,¹⁰ Cai et al.,¹¹ Liu et al.,¹² Martinez and Cayon,¹³ and De Cicco et al.¹⁴ offer complementary evidence regarding surgical approaches, associated fracture patterns, and functional prognosis that could eventually be incorporated into automated models for classification and therapeutic planning.

Kuo et al.⁸ also reported that AI performance exhibits a sensitivity and specificity approximately 3% lower than that of physicians, although the differences were not statistically significant. Likewise, Alenazi et al.¹⁵ highlight that AI can be a valuable adjunct to clinical judgment, particularly in environments with limited specialist availability.

The Schatzker–Kfuri classification represents a greater interpretative challenge than traditional radiographic systems, as it incorporates tomographic and three-dimensional information. Despite this, the model in our study was able to identify fracture patterns in each quadrant with high precision and to correctly recognize the presence or absence of metaphyseal-diaphyseal dissociation.

Overall, our findings demonstrate that, when provided with appropriate visual guidance and a structured analytical approach, multimodal language models can serve as useful adjuncts in orthopedic education and diagnostic support in traumatology.

CONCLUSIONS

The ChatGPT-4o model correctly classified all six cases of tibial plateau fractures according to the three-dimensional Schatzker-Kfuri classification, achieving complete agreement with an expert observer. This paves the way for the use of AI as a tool for clinical decision support, particularly in training or diagnostic validation settings.

Conflicts of interest: The authors declare no conflicts of interest.

E. Rivadeneira ORCID ID: <https://orcid.org/0009-0006-5784-5700>

D. E. Freire ORCID ID: <https://orcid.org/0009-0000-9882-6027>

A. F. Samaniego ORCID ID: <https://orcid.org/0000-0002-6616-6471>

E. Lulkin ORCID ID: <https://orcid.org/0000-0002-4119-0483>

F. Bidolegui ORCID ID: <https://orcid.org/0000-0002-0502-2300>

S. Pereira ORCID ID: <https://orcid.org/0000-0001-9475-3158>

REFERENCES

1. Lhotská L. Umělá inteligence v medicíně a zdravotnictví: Příležitost a/nebo hrozba? *Čas Lék Čes* 2023;162(7-8):275-8. Available at: <https://www.prolekare.cz/casopisy/casopis-lekaru-ceskych/2023-7-8-1/umela-inteligence-v-medicine-a-zdravotnictvi-prilezitost-a-nebo-hrozba-136669>
2. Mucci T. La historia de la inteligencia artificial. *IBM Think* 2019 [citado 2025 nov 21]. Available at: <https://www.ibm.com/es-es/think/topics/history-of-artificial-intelligence>.
3. Kfuri M, Schatzker J. Revisiting the Schatzker classification of tibial plateau fractures. *Injury* 2018;49(12):2252-63. <https://doi.org/10.1016/j.injury.2018.07.010>
4. Mohammadi M, Parviz S, Parvaz P, Pirmoradi MM, Afzalimoghaddam M, Mirfazaelian H. Diagnostic performance of ChatGPT in tibial plateau fracture in knee X-ray. *Emerg Radiol* 2025;32(1):59-64. <https://doi.org/10.1007/s10140-024-02298-y>
5. Van der Gaast N, Bagave P, Assink N, Broos S, Jaarsma RL, Edwards MJR, et al. Deep learning for tibial plateau fracture detection and classification. *Knee* 2025;54:81-9. <https://doi.org/10.1016/j.knee.2025.02.001>
6. Markhardt B, Gross JM, Monu J. Schatzker classification of tibial plateau fractures: Use of CT and MR imaging improves assessment. *Radiographics* 2009;29(2):585-97. <https://doi.org/10.1148/rg.292085078>
7. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Cantarelli Rodrigues T, Recht MP. Artificial intelligence in musculoskeletal imaging: current status and future directions. *AJR Am J Roentgenol* 2019;213(3):506-13. <https://doi.org/10.2214/AJR.19.21117>
8. Kuo R, Harrison C, Curran T, Jones B, Freethy A, Cussons D, et al. Artificial intelligence in fracture detection: A systematic review and meta-Analysis. *Radiology* 2022;304(1):50-62. <https://doi.org/10.1148/radiol.211785>

9. Giordano V, Schatzker J, Kfuri M. The 'Hoop' plate for posterior bicondylar shear tibial plateau fractures: Description of a new surgical technique. *J Knee Surg* 2022;35(2):123-9. <https://doi.org/10.1055/s-0036-1593366>
10. Singh Sidhu GA, Hind J, Ashwood N, Kaur H, Bridgwater H, Rajagopalan S. A systematic review of current approaches to tibial plateau, *Cureus* 2022;14(7):e27183. <https://doi.org/10.7759/cureus.27183>
11. Cai D, Zhou Y, He W, Yuan J, Liu C, Li R, et al. Automatic segmentation of knee CT images of tibial plateau fractures based on three-dimensional U-Net: assisting junior physicians with Schatzker classification. *Eur J Radiol* 2024;178:111605. <https://doi.org/10.1016/j.ejrad.2024.111605>
12. Liu Y, Fang R, Tu B, Zhu Z, Zhang C, Ning R. Correlation of preoperative CT imaging shift parameters of the lateral plateau with lateral meniscal injury in Schatzker IV-C tibial plateau fractures. *BMC Musculoskelet Disord* 2023;24(1):793. <https://doi.org/10.1186/s12891-023-06924-7>
13. Martinez A, Cayon M. Fracturas del platillo tibial posterior. *Revista Colombiana de Cirugía Ortopédica y Traumatología* 1999;13(1):37-1. Disponible en: <https://sccot.org/pdf/RevistaDigital/1999/Vol13N1/37-41.pdf>
14. De Cicco F, Verbner J, Abrego M, Taype D, Carabelli G, Barla J, et al. Soporte circunferencial posterior en fracturas de platillo tibial. *Rev Asoc Argent Ortop Traumatol* 2021;86(2):219-27. <https://doi.org/10.15417/issn.1852-7434.2021.86.2.1018>
15. Alenazi HK, Alahmari RA, Mubarak Hassan Al Faraj A, Nasser Almurkan M, Saleh Al Hashel IM, Al Hagwi AI, et al. The future of artificial intelligence in X-ray radiography: Enhancing healthcare and workflow efficiency. *J Int Crisis Risk Commun Res* 2024;7(53):51-3. <https://doi.org/10.63278/jicrcr.vi.708>