

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

Héctor A. Rivadeneira Jurado,¹ Elias A. Rivadeneira Jurado,¹ Daniel Espinoza Freire,¹ Andrés F. Samaniego,¹ Ezequiel Lukin,² Fernando Bidolegui,² Sebastián Pereira¹

¹Servicio de Ortopedia y Traumatología, Hospital Sirio-Libanés, Ciudad Autónoma de Buenos Aires, Argentina

²Servicio de Ortopedia y Traumatología, Sanatorio Otamendi Miroli, Ciudad Autónoma de Buenos Aires, Argentina

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 casos 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 casos de fractura de platillo tibial, que no se habían cargado previamente en la plataforma para analizar la interpretación del ChatGPT-4o basada en el *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 epifisiodiafisaria fueron identificados, con precisión, en 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

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

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

Recibido el 5-9-2025. Aceptado luego de la evaluación el 21-11-2025 • Dr. HÉCTOR A. RIVADENEIRA JURADO • 1bhrribadeneirajurado@gmail.com  <https://orcid.org/0009-0008-6397-9718>

Cómo citar este artículo: Rivadeneira Jurado HA, Rivadeneira Jurado EA, Espinoza Freire D, Samaniego AF, Lukin E, Bidolegui F, Pereira S. 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. *Rev Asoc Argent Ortop Traumatol* 2025;90(6):556-560. <https://doi.org/10.15417/issn.1852-7434.2025.90.6.2224>



Esta Revista está bajo una Licencia Creative Commons Atribución-NoComercial-Compartir Obras Derivadas Igual 4.0 Internacional. (CC-BY-NC-SA 4.0).

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

INTRODUCCIÓN

La inteligencia artificial (IA) se presentó formalmente en 1956.¹ Con el pasar de los años, se fueron desarrollando programas informáticos para ser utilizados en distintos campos de desarrollo, inclusive la ortopedia y traumatología. No obstante, las plataformas actuales necesitan de un *prompt* o conjunto de información adecuado para poder obtener el resultado apropiado.²

En traumatología y ortopedia, las fracturas de platillo tibial representan un desafío diagnóstico y terapéutico. La clasificación de Schatzker, creada en 1979, para las fracturas de platillo tibial, se utiliza mucho en traumatología al igual que el modelo de evaluación tridimensional basado en tomografía computarizada (TC) para definir mejor el compromiso anatómico de los cuadrantes del platillo tibial, que dio origen a la clasificación de Schatzker-Kfuri, desarrollada en 2018.³ Esta clasificación permite diferenciar entre fracturas unicondilares, bicondilares y con disociación epifisodiafisaria, y ha mejorado la planificación quirúrgica.

El objetivo de este estudio fue comparar la capacidad del modelo de lenguaje multimodal ChatGPT-4o para clasificar fracturas de platillo tibial utilizando tanto radiografías como imágenes de TC, con la de estudios de reportes de casos publicados por un panel de expertos.

MATERIALES Y MÉTODOS

Se llevó a cabo un estudio retrospectivo, observacional para comparar la interpretación del observador experto y la del ChatGPT-4o. Para la creación del *prompt*, se incluyeron 45 reportes de casos publicados en distintas bases de datos, como PubMed, Elsevier, Scielo, que contaban con radiografías anteroposterior y de perfil de rodilla, y TC de rodilla con cortes axial, coronal y sagital. Se excluyeron todos los reportes de casos que tenían cortes de TC incompletos y los que no contaban con las radiografías completas. Por ende, los 45 reportes de casos publicados por expertos con radiografías y TC se usaron para mejorar la precisión de la interpretación del ChatGPT-4o. Las imágenes recopiladas fueron organizadas antes de cargarlas en la plataforma, en el siguiente orden: radiografía anteroposterior de rodilla, de perfil de rodilla, cortes axial, coronal y sagital del platillo tibial (Figura).

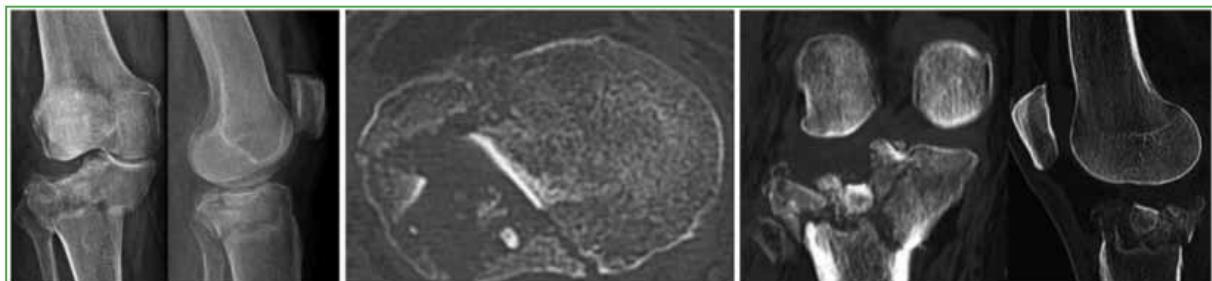


Figura. Secuencia de imágenes radiográficas y tomográficas para la interpretación por parte de la plataforma.

Además, se agregó información para crear el *prompt* con contenido de anatomía, conceptos básicos de traumatología, ejemplos sobre cizallamiento o *split*, hundimiento, combinación de patrones de fractura y ejemplos de fracturas con extensión epifisodiafisaria.

En la plataforma, se cargó información descriptiva asociada a ilustraciones progresivamente hasta completar el *prompt*, del modelo de la IA, luego se cargaron 45 reportes de casos en formato DICOM a fin de mejorar la interpretación de la plataforma. Por último, se usaron 6 reportes de casos publicados por expertos que no se habían cargado antes en la plataforma para su interpretación. Cada uno de los 6 reportes de casos contenía un ejemplo de los distintos trazos de fractura que se evalúan en la clasificación.

El modelo ChatGPT-4o analizó, de forma secuencial, cada imagen y propuso una clasificación basada en el sistema de Schatzker-Kfuri. Se registró la clasificación propuesta y se comparó con la clasificación oficial del caso. Se consideró como “correcta” aquella coincidencia completa con la clasificación mencionada en los reportes de casos.

RESULTADOS

Los 6 casos fueron correctamente clasificados por el modelo. Se identificaron correctamente los patrones de:

- Hundimiento puro (tipo III)
- Cizallamiento lateral (tipo I)
- Fractura bicondilar sin disociación (tipo V)
- Disociación epifisodiasisaria (tipo VI)
- Compromiso medial (tipo IV)
- Fracturas con hundimiento más *split* lateral (tipo II)

En la [Tabla](#), se resume la comparación.

El análisis de 6 reportes de casos previamente descritos por parte del panel de expertos comparado con el análisis del ChatGPT-4o muestra concordancia según el análisis radiográfico junto con el análisis tridimensional. Cabe mencionar que el estudio presenta un coeficiente kappa de Cohen: 1.00, lo cual se interpreta como una concordancia perfecta.

Tabla. Resultado de la comparación entre el observador experto y el ChatGPT-4o

Caso	Clasificación de Schatzker	Clasificación de Schatzker-Kfuri	Clasificación del ChatGPT-4o	Cuadrantes afectados	Concordancia
1	IV	IV - PM	Schatzker IV	PM	Correcto
2	II	II - AL	Schatzker II	AL	Correcto
3	V	V - AL + PM	Schatzker V	AL + PM	Correcto
4	I	I - AL (<i>split</i> puro)	Schatzker I	AL (<i>split</i> puro)	Correcto
5	VI	VI - AL + PM + (disociación epifisodiasisaria)	Schatzker VI	AL + PM + disociación epifisodiasisaria	Correcto
6	III	III - AL (hundimiento puro)	Schatzker III	AL (hundimiento puro)	Correcto

AL = anterolateral; PM = posteromedial.

DISCUSIÓN

Los resultados de este estudio concuerdan con los de publicaciones recientes que muestran el potencial de la IA en el diagnóstico de las fracturas articulares. En particular, en estudios, como el de Mohammadi y cols.,⁴ y Van der Gaast y cols.,⁵ se ha señalado que modelos, como ChatGPT-4o, pueden alcanzar niveles comparables a los de los especialistas en la interpretación de radiografías. Esto también se ha destacado en escenarios de limitación de recursos, donde el uso de TC tridimensional ha mejorado la interpretación, como lo describen Markhardt y cols.⁶ Además, revisiones recientes sobre la IA en cirugía ortopédica subrayan la necesidad de desarrollar estudios comparativos con especialistas para identificar la validez de la IA en la interpretación de imágenes, tal como argumentan Gyftopoulos y cols.,⁷ y Kuo y cols.⁸ En ambos estudios, se evaluó el rendimiento predictivo de modelos de

aprendizaje profundo para la clasificación de fracturas de platillo tibial, respaldando el valor de la IA en escenarios clínicos reales. Asimismo, investigaciones, como las de Giordano y cols.,⁹ Singh Sidhu y cols.,¹⁰ Cai y cols.,¹¹ Liu y cols.,¹² Martinez y Cayon,¹³ y De Cicco y cols.¹⁴ aportan evidencia complementaria sobre el abordaje quirúrgico, los patrones asociados y el pronóstico funcional que pueden ser integrados a modelos automatizados de clasificación y planificación terapéutica.

Por otro lado, Kuo y cols.⁸ indicaron que la IA tiene una sensibilidad y una especificidad un 3% menores que la interpretación de los médicos, pero con resultados no significativos. Asimismo, Alenazi y cols.¹⁵ destacan que la IA puede ser útil acompañada del juicio médico para ahorrar tiempo en sectores de escasos recursos.

La clasificación de Schatzker-Kfuri, al incorporar información tomográfica y tridimensional, plantea un reto mayor que la clasificación radiográfica tradicional. Sin embargo, el modelo fue capaz de identificar, con precisión, los trazos en cada cuadrante y reconocer la presencia o ausencia de disociación metafisodiasfisaria.

En nuestro estudio, se demostró que, con una adecuada guía visual y el análisis sistemático, los modelos de lenguaje con capacidad multimodal pueden ser herramientas útiles en la educación médica y la asistencia diagnóstica en traumatología.

CONCLUSIONES

El modelo ChatGPT-4o clasificó correctamente los 6 casos de fractura de platillo tibial según la clasificación tridimensional de Schatzker-Kfuri, con una concordancia total con un observador experto. Esto abre el camino al uso de la IA en el apoyo a la decisión clínica, particularmente en escenarios de entrenamiento o validación diagnóstica.

Conflicto de intereses: Los autores no declaran conflictos de intereses.

ORCID de E. A. Rivadeneira: <https://orcid.org/0009-0006-5784-5700>
 ORCID de D. Espinoza Freire: <https://orcid.org/0009-0000-9882-6027>
 ORCID de A. F. Samaniego: <https://orcid.org/0000-0002-6616-6471>

ORCID de E. Lukin: <https://orcid.org/0000-0002-4119-0483>
 ORCID de F. Bidolegui: <https://orcid.org/0000-0002-0502-2300>
 ORCID de S. Pereira: <https://orcid.org/0000-0001-9475-3158>

BIBLIOGRAFÍA

1. Lhotská L. Umělá inteligence v medicíně a zdravotnictví: Příležitost a/nebo hrozba? *Čas Lék Čes* 2023;162(7-8):275-8. Disponible en: <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]. Disponible en: <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/rgr.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>