The Knee 54 (2025) 81–89

Contents lists available at ScienceDirect

The Knee

journal homepage: www.elsevier.com/locate/thekne

Deep learning for tibial plateau fracture detection and classification ☆

N. van der Gaast [a,b,*], P. Bagave [c], N. Assink [d], S. Broos [a], R.L. Jaarsma [a], M.J.R. Edwards [b], E. Hermans [b], F.F.A. IJpma [d], A.Y. Ding [c], J.N. Doornberg [a,d], J.H.F. Oosterhoff [c,d], on behalf of the Machine Learning Consortium [1]

[a] Department of Orthopaedics & Trauma Surgery, Flinders Medical Centre and Flinders University, Adelaide, SA, Australia
[b] Department of Trauma Surgery, Radboud University Medical Center, Radboud University Nijmegen, the Netherlands
[c] Department of Engineering Systems and Services, Faculty of Technology Policy and Management, Delft University of Technology, Delft, the Netherlands
[d] Department of Orthopaedic and Trauma Surgery, University Medical Center Groningen, University of Groningen, the Netherlands

ARTICLE INFO

Article history:
Received 18 July 2024
Revised 17 January 2025
Accepted 4 February 2025

Keywords:
Tibial plateau fractures
Deep learning
Artificial intelligence

ABSTRACT

Background: Deep learning (DL) has been shown to be successful in interpreting radiographs and aiding in fracture detection and classification. However, no study has aimed to develop a computer vision model for tibia plateau fractures using the Schatzker classification. Therefore, this study aims to develop a deep learning model for (1) detection of tibial plateau fractures and (2) classification according to the Schatzker classification.
Methods: A multicenter approach was performed for the collection of radiographs of patients with tibia plateau fractures. Both anteroposterior and lateral images were uploaded into an annotation software and manually labelled and annotated. The dataset was balanced for optimizing model development and split into a training set and a test set. We trained two convolutional neural networks (GoogleNet and ResNet) for the detection and classification of tibia plateau fractures following the Schatzker classification.
Results: A total of 1506 knee radiographs from 753 patients, including 368 tibial plateau fractures and 385 healthy knees, were used to create the algorithm. The GoogleNet algorithm demonstrated high sensitivity (92.7%) but intermediate accuracy (70.4%) and positive predictive value (64.4%) in detecting tibial plateau fractures, indicating reliable detection of fractured cases. It exhibited limited success in accurately classifying fractures according to the Schatzker system, achieving an accuracy of only 34.6% and a sensitivity of 32.1%.
Conclusion: This study shows that detection of tibial plateau fractures is a task that a DL algorithm can grasp; further refinement is necessary to enhance their accuracy in fracture classification. Computer vision models might improve using different classification systems, as the current Schatzker classification suffers from a low interobserver agreement on conventional radiographs.

☆ This article is part of a special issue entitled: 'AI and new technologies' published in The Knee.
* Corresponding author at: Flinders Medical Centre/Flinders University, Flinders Drive, Bedford Park, Adelaide, Australia.
E-mail address: Nynke.vandergaast@radboudumc.nl (N. van der Gaast).
[1] Machine Learning Consortium: Minh-Son To, Bhavin Jadav, Jasper Prijs, Vincent Stirler, Katharina Denk, Reinier Spek.

https://doi.org/10.1016/j.knee.2025.02.001
0968-0160/© 2025 The Author(s). Published by Elsevier B.V.

1. Introduction

Deep learning (DL) has emerged as a powerful tool and its additional value in the medical field has been investigated widely. DL algorithms are designed to extract patterns and knowledge from extensive datasets, allowing the calculation of probabilities when presented with new, analogous data. They have had successful application in the medical field, such as the detection of diseases such as skin, breast or lung cancer or osteoarthritis on knee X-rays [1–4]. Clinical decision support models for orthopaedic trauma, such as a clinical prediction tool for scaphoid fractures after wrist trauma, has been proven effective in reducing the necessity for additional imaging [5] and they can achieve performance on par with humans in detecting fractures, such as wrist, hip, and ankle fractures [6,7]. Moreover, a systematic review by Langerhuizen et al. [8] examined the use of artificial intelligence in orthopaedic trauma, including 10 relevant articles. The review revealed that AI models were almost as accurate as humans in fracture detection, with AI models even surpassing human performance in detecting hip and proximal humerus fractures [6,8–10].

Tibial plateau fractures are mostly caused by trauma and account for approximately 1–2% of all fractures in adults [11]. The identification of a tibial plateau fracture is initially based on lateral and anteroposterior (AP) radiographic views. The most commonly used classification system is the Schatzker classification [12], which relies on these radiological views. Nowadays, the standard care for patients with tibial plateau fractures includes a computed tomography (CT) scan in addition to radiographs. One could argue that the addition of a CT scan might enhance the understanding of fracture lines and improve the accuracy and interobserver agreement of this classification. However, despite the incorporation of two-dimensional (2D) CT images and three-dimensional (3D) reconstructions of fractures, the interobserver agreement among surgeons for the Schatzker classification has seen little improvement over time, hardly surpassing a moderate interobserver agreement with a kappa value between 0.36 and 0.62 [13–18].

Recognition of different fracture characteristics guides surgical decision-making by informing different treatment strategies for each specific fracture characteristic [19]. A well-performing DL algorithm could potentially assist surgeons in recognizing these fracture characteristics and, consequently, support surgical decision-making in the future [8,20]. The study by Liu et al. [21] partly answered this question and generated an algorithm for the recognition of tibial plateau fractures that achieved an accuracy of 91%, which was comparable to orthopaedic surgeons, with an accuracy of 92%. The study of Lind et al. [22] created a ResNet-based neural network that classifies proximal tibia fractures according to the AO/OTA classification [23], with an area under the curve varying between 0.72 and 0.99. To the best of our knowledge, algorithms that can detect and classify tibial plateau fractures according to the Schatzker classification are still lacking. Therefore, this study aims to develop a DL model for (1) detection of tibial plateau fractures and (2) automated classification according to the Schatzker classification. The hypothesis is that fracture detection will be successful, but automated classification will be difficult and less reliable because of the low interobserver agreement due to subjectivity in the classification system itself.

2. Material and methods

2.1. Study design

This is a retrospective diagnostic imaging study that was performed with imaging from two hospitals from different countries. The procedures used in this study adhere to the tenets of the Declaration of Helsinki and ethics approval was obtained from each participating hospital. The imaging data for this study was collected from two hospitals: Flinders Medical Centre, Australia, and the University Medical Center Groningen, the Netherlands. In each centre, approval from the ethical board was gained. (CALHN Reference Number: 13991 and Research Register number: 202200114). This study was reported following the clinical AI research (CAIR) checklist [24] (Supplementary Table S2).

2.2. Tibial plateau fracture data

Inclusion criteria were patients that presented with a tibial plateau fracture, had available injury radiographs of sufficient quality (e.g., including the entire injury with both lateral and AP view) and received a CT of the knee for adequately classifying the fractures. Exclusion criteria were age <18 years, isolated fractures of the tibial eminence, minor avulsion fractures, previous ipsilateral surgery of the lower leg or congenital bone disorders.

The South Australian Medical Imaging (SAMI) database was used to identify patients who were treated for a tibial plateau fracture between January 2010 and December 2020. First, a search was performed to find radiographs of the knee accompanied by a radiology report and a CT scan. The anonymized radiology reports were used to find tibial plateau fractures using natural language processing. In total, over 90,000 studies were analysed, with approximately 2000 radiology reports containing the terms ‘tibial plateau fracture’, ‘plateau fracture’ ‘fracture’ and/or ‘Schatzker’. After manual selection, a total of 229 tibial plateau fractures with adequate imaging methods were extracted, of which 16 fractures were excluded based on our exclusion criteria. A second dataset, containing tibial plateau fracture data from an existing dataset from a different study [25,26] from the University Medical Center Groningen (UMCG) was used. This dataset contained patients who suffered from a tibial plateau fracture between January 2003 and December 2019 from one specific hospital in the Netherlands. This dataset contained a total of 186 tibial plateau fractures, of which 31 fractures were excluded based on our exclusion criteria.

In total, 369 tibial plateau fracture patients were included (Figure 1). The radiographs were retrieved in DICOM format and were converted into a PNG file in the pre-processing stage and different views were combined using Visual Studio Code (V 1.89.1, Microsoft, WA, USA). The final combined PNG files were uploaded to a safe digital environment called Labelbox (Labelbox, San Fransisco, CA, USA) for labelling and annotation of the fractures.

2.3. Classification of tibial plateau fractures

For the classification of tibial plateau fractures, the Schatzker classification [12] was used. Each fracture was classified based on the available radiographs and CT scans by one orthopaedic trauma research fellow and one trauma surgeon (N.G. and E.H./B.J.). Any disagreement was dissolved by involving a third, experienced trauma surgeon (R.J./M.J.E.).

2.4. Non-fractured knees

A matched control group of healthy knees is crucial when developing and evaluating a DL algorithm. When searching for the tibial plateau fracture data (n = 2000), many radiology reports stated ‘no fracture’. Radiographs of non-fractured knees were included from the SAMI database when they showed a non-fractured knee of a patient >18 years old without any old or other acute fractures (e.g., fibula, femur or patella). If there was inconclusive evidence regarding the presence or absence of a fracture, the patient was also excluded. In total, 385 radiographs that stated ‘no fracture’ in the radiology report with sufficient quality of X-rays were included. All radiographs were provided in DICOM format. The radiographs of each patient were then converted to a PNG file and different views were combined using Visual Studio Code (V 1.89.1, Microsoft, WA, USA). The final combined PNG files were uploaded to Labelbox and labelled as ‘no fracture’.

2.5. Labelling

All combined radiographs were uploaded on to a safe digital environment that enables the addition of labels and annotations (Labelbox, San Francisco, CA, USA). All fractures were labelled according to the Schatzker classification: Schatzker 1 t/m 6, ‘no fracture’ or ‘exclusion’. As the lateral and AP radiographs were combined, the image positions were also labelled in Labelbox: Left AP-right Lateral or Right AP-Left Lateral. All labels were completed by the first author (Figure 2).

2.6. Annotations

In Labelbox, the tibia and fibula were manually annotated for each tibial plateau fracture as described above. All fracture lines where then drawn by a fracture pencil to guide the algorithm towards the fracture (Figure 3).

2.7. Algorithm development

The annotated and labelled fractures were extracted from Labelbox using Python (V 3.12, Wilmington, DE, USA). For the fracture detection model, 368 fractures and 385 non-fractures were used. The training set contained 327 fractures and 345

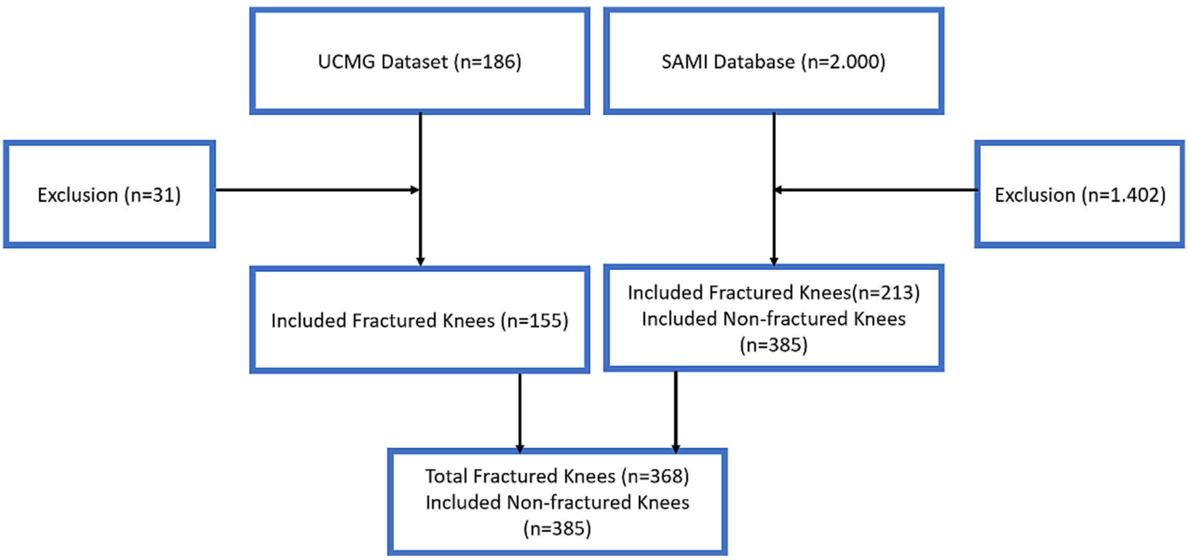

Figure 1. Flow chart of patient inclusion. SAMI, South Australian Medical Imaging database; UCMG, University Medical Center Groningen.

Figure 2. Labels.

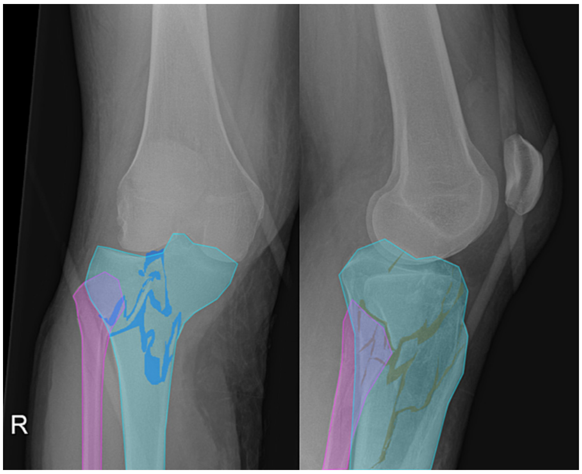

Figure 3. Annotations.

non-fractures and the test set contained 41 fractures and 40 non-fractures. GoogleNet and ResNet architecture were used for fracture detection. GoogleNet architecture is a deep convolutional neural network, that is a variant of the Inception Network, developed by researchers at Google, and includes multiple inception modules stacked together, leading to a network with adequate depth and accuracy [27]. ResNet introduces a residual learning framework that simplifies training for deeper neural networks by reformulating layers to learn residual functions relative to their inputs. This approach improves optimization and allows deeper networks to achieve higher accuracy [28].

2.8. Data augmentation was used to optimize the algorithm's results

For the fracture classification model, it was aimed to use 40 samples of each Schatzker classification subtype. However, for Schatzker 1 fractures, there were only 37 samples, which was accepted. For Schatzker 3, there were only 17 samples. Therefore, the images were duplicated in order to balance the numbers. Lastly, 40 non-fractures were added. Ninety per cent of the data was used for the training set and 10% was used for the test set. A GoogleNet architecture without pretrained weights was used first. The weights were then retrained for the task of classifying seven different classes (six subtypes of Schatzker and non-fractures).

2.9. Outcomes measures

To evaluate the performance of the model the following diagnostic performance metrics were calculated using the accuracy, sensitivity, specificity, precision (PPV), negative predictive value (NPV) and F1 score. Moreover, a precision-recall (PR) curve and receiver operating characteristic (ROC) curve were drawn (Supplementary file S1).

3. Results

3.1. Baseline characteristics

The algorithm was trained on a total of 753 AP and 753 lateral knee radiographs, including 368 tibial plateau fractures and 385 healthy knees. The mean age and the sex of the included patients could not be retrieved, as this was a dataset with only radiographs accessible. The division according to the Schatzker classification was as follows: 37 × Schatzker 1 (10.1%), 146 × Schatzker 2 (39.7%), 17 × Schatzker 3 (4.6%), 46 × Schatzker 4 (12.5%), 65 × Schatzker 5 (17.7%) and 57 × Schatzker 6 (15.5%).

3.2. Fracture detection

The sensitivity of the GoogleNet algorithm reached 92.7%, accuracy reached 70.4%, positive predictive value reached 64.4%, and the specificity reached 47.5%. Nineteen of 81 samples of the test set data were classified as a fracture, where no fracture was addressed by the expert opinion (i.e., false positive) The sensitivity of the ResNet algorithm reached 82.9%, accuracy reached 65.4%, positive predictive value reached 61.8%, and the specificity reached 47.5%. (Table 1, Figures 4 and 5). Moreover, the ROC curve and PR curve for both GoogleNet and ResNet are shown in Figures 6 and 7, respectively.

3.3. Fracture classification

The sensitivity of the GoogleNet classification model reached an accuracy of only 32.1%, indicating that the model is not accurately classifying Schatzker classifications on conventional radiographs for patients with tibia plateau fractures (Figure 5).

4. Discussion

This study aimed to develop a DL model for detecting tibial plateau fractures and for automated classification using the Schatzker classification. This study reveals that the detection of fractures by a GoogleNet algorithm was the most accurate compared to a ResNet algorithm, with a high sensitivity (92.7% vs. 82.9%) and an intermediate accuracy (70.4% vs. 65.4%). The classification of tibial plateau fractures was less successful, with a sensitivity of <50%.

4.1. Interpretation of results

Regarding fracture detection, the DL algorithm's high precision suggests its potential as a screening tool, aiding clinicians in promptly identifying fractures. However, its intermediate accuracy and sensitivity underscores the need of further refinement to improve the algorithm. In contrast, the DL algorithm exhibited limited success in classifying tibial plateau fractures according to the Schatzker classification. With an accuracy of less than 50%, the algorithm's performance in accurately categorizing fractures into specific subtypes was suboptimal. This difficulty in classifying tibial plateau fractures aligns with challenges faced by surgeons due to inconsistent interpretations and only moderate inter- and intra-observer reliability

Table 1
Performance metrics.

	Fracture detection (Googlenet)	Fracture detection (Resnet)
Accuracy	70.4%	65.4%
Sensitivity (recall)	92.7%	82.9%
Specificity	47.5%	47.5%
PPV (precision)	64.4%	61.8%
NPV	86.4%	73.1%
F1-score	0.76	0.71

Model performance metrics are explained in Supplementary Table S1. NPV, negative predictive value; PPV, positive predictive value.

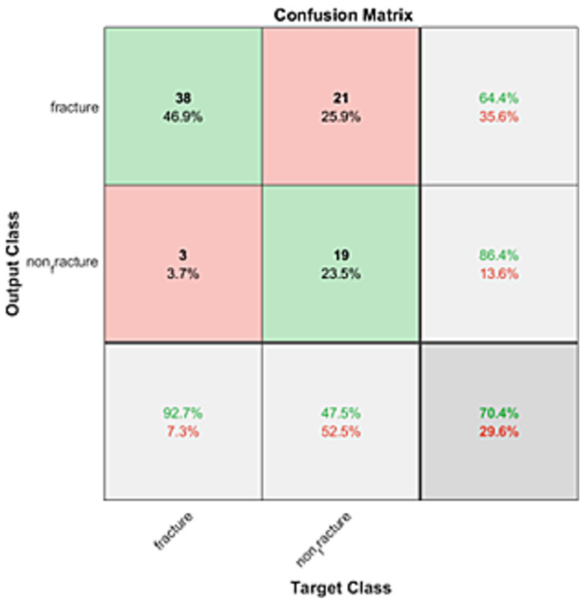

Figure 4. Confusion matrix fracture detection.

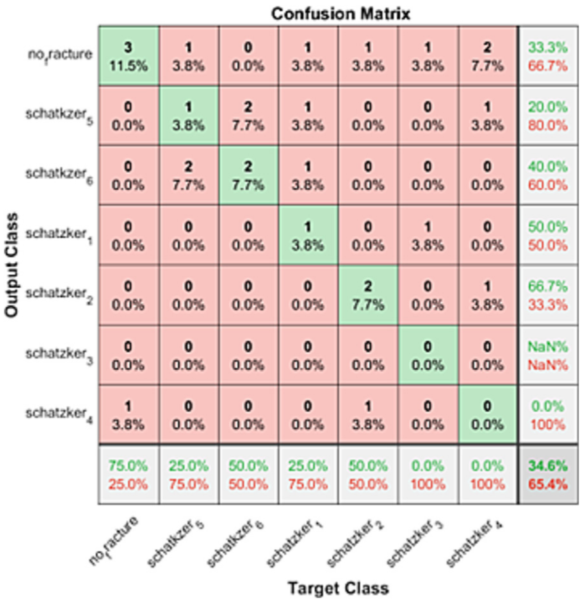

Figure 5. Confusion matrix fracture classification.

on the most commonly used classification systems [15,29–32]. The Schatzker classification is the most commonly used classification system around the world, however, it encounters controversies when looking at the consistency in classification between surgeons. The interobserver agreement for this classification system is widely ranging from a kappa value between 0.36 and 0.62 based on the sole use of radiographs [13,17,18,29,33–37]. Radiographs, the traditional imaging method, potentially lack detail necessary for accurate classification due to overlapping characteristics of the Schatzker classification, making them difficult to distinguish. Despite the advancements in imaging technology, the use of 2D and/or 3D CT does not lead to more consistency and still could cause possible misinterpretations [29,38]. As surgeons are already struggling with the nuances of the subtypes of the Schatzker classification, our study shows that a DL algorithm also struggles with this task.

One could argue to propose a new classification method, however new classification systems based on 2D and 3D imaging methods have already been proposed in recent years; i.e., Luo's three-column concept [39], the 10-segment classification by

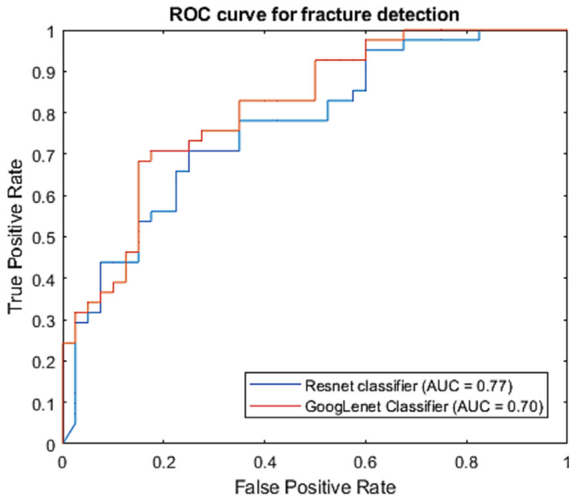

Figure 6. Precision-recall (PR) curve. PPV, positive predictive value.

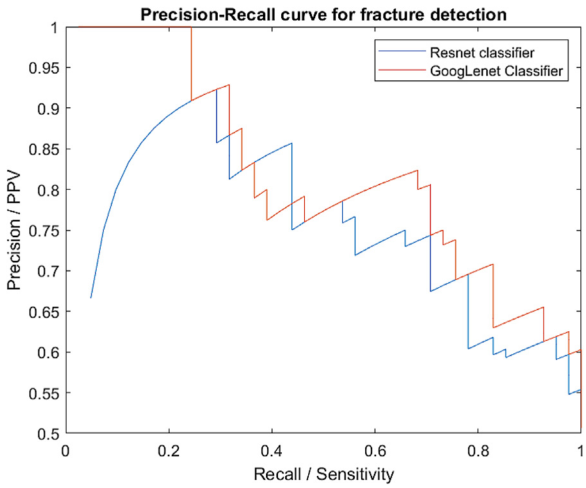

Figure 7. Receiver operating characteristic (ROC) curve.

Krause et al. [40] and the two-column classification by Anwar et al. [41]. However, these classification systems are not widely adopted in clinical practice. Therefore, as a future perspective it might be interesting to explore DL algorithms' potential to propose consistent fracture classifications.

The study's limitations include the relatively small dataset. Despite including 368 fractures, the distribution of subtypes resulted in a disproportionate representation, with Schatzker 2 fractures (n = 146) significantly outnumbering Schatzker 3 fractures (n = 17). This imbalance might be the cause of the inaccurate algorithm for fracture classification. Ideally, a more balanced representation of each fracture subtype might enhance the algorithm's performance.

Moreover, variations in the quality of radiographs across two different hospitals could have introduced additional challenges. Radiograph quality can vary due to factors such as differences in equipment and techniques and cast overprojections may have introduced biases into the dataset. However, these variations reflect the reality of clinical practice. Any algorithm intended for clinical use must be robust enough to handle such challenges.

Future studies could enhance the algorithm by optimizing it with larger datasets from different hospitals across the world and exploring alternative data augmentation techniques. A fully functional algorithm could further support clinicians by incorporating visualizations of fracture lines and treatment options, providing valuable insights and guidance for optimal tibial plateau fracture treatment. Moreover, the algorithm could be compared to clinicians' performance.

Lastly, the final algorithm has to be externally validated, which is essential for its reliability and generalizability beyond the dataset used in this study. External validation is crucial to ensure that the algorithm performs consistently across different settings and patient populations. Therefore, future studies should prioritize external validation to confirm the algorithm's utility and effectiveness in real-world clinical scenarios.

5. Conclusions

While DL algorithms show promise in fracture detection, further refinement is necessary to enhance their accuracy in fracture classification. While machine learning might present a theoretical solution to mitigate human biases and enhance fracture classification reliability, this study shows that this task is still too complex for an algorithm to accurately perform due to the complexity of fracture classification by surgeons as a ground truth. We acknowledge the need for more extensive and diverse data to train these algorithms effectively. By addressing the limitations and leveraging the strengths of both DL algorithms and clinical expertise, we can harness the full potential of AI-driven approaches to improve patient outcomes in orthopaedic trauma care.

CRediT authorship contribution statement

N. van der Gaast: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **P. Bagave:** Visualization, Validation, Methodology, Investigation, Formal analysis. **N. Assink:** Writing – review & editing, Data curation. **S. Broos:** Software, Data curation. **R.L. Jaarsma:** Writing – review & editing, Supervision, Resources, Project administration, Data curation, Conceptualization. **M.J.R. Edwards:** Writing – review & editing, Supervision. **E. Hermans:** Writing – review & editing, Supervision, Data curation, Conceptualization. **F.F.A. IJpma:** Writing – review & editing, Supervision. **A.Y. Ding:** Writing – review & editing, Supervision, Software, Methodology. **J.N. Doornberg:** Writing – review & editing, Supervision, Resources, Methodology, Data curation, Conceptualization. **J.H.F. Oosterhoff:** Writing – review & editing, Writing – original draft, Supervision, Software, Project administration, Methodology.

Ethics approval

Data collection for this study was collected Consent to participants Informed consent was not obtained from all individual participants included in the study since only anonymous data was used.

Funding

Nynke van der Gaast was supported by Radboud University Medical Center.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: ‘‘Nynke van der Gaast reports was provided by Radboud University Medical Center. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper”.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.knee.2025.02.001.

References

[1] Schalekamp S, van Ginneken B, Karssemeijer N, Schaefer-Prokop CM. Chest radiography: new technological developments and their applications. Semin Respir Crit Care Med 2014;35:3–16. doi: https://doi.org/10.1055/s-0033-1363447.

[2] Cruz-Roa AA, Arevalo Ovalle JE, Madabhushi A, González Osorio FA. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. Med Image Comput Comput Assist Interv 2013;16:403–10. doi: https://doi.org/10.1007/978-3-642-40763-5_50.

[3] Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NNC. *Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent*. Sci Rep 2017;7:46450. doi: https://doi.org/10.1038/srep46450.

[4] Norman B, Pedoia V, Noworolski A, Link TM Majumdar S. *Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs*. J Digit Imaging 2019;32:471–7. doi: https://doi.org/10.1007/s10278-018-0098-3.

[5] Bulstra AEJ. Machine Learning Consortium *A machine learning algorithm to estimate the probability of a true scaphoid fracture after wrist trauma*. J Hand Surg Am 2022;47:709–18. doi: https://doi.org/10.1016/j.jhsa.2022.02.023.

[6] Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A. *Artificial intelligence for analyzing orthopedic trauma radiographs*. Acta Orthop 2017;88:581–6. doi: https://doi.org/10.1080/17453674.2017.1344459.

[7] Liao Z, Liao K, Shen H, van Boxel MF, Prijs J, Jaarsma RL. CNN attention guidance for improved orthopedics radiographic fracture classification. IEEE J Biomed Health Inform 2022;26:3139–50. doi: https://doi.org/10.1109/JBHI.2022.3152267.

[8] Langerhuizen DWG, Janssen SJ, Mallee WH, van den Bekerom MPJ, Ring D, Kerkhoffs GMMJ. What are the applications and limitations of artificial intelligence for fracture detection and classification in orthopaedic trauma imaging? A systematic review. Clin Orthop Relat Res 2019;477:2482–91. doi: https://doi.org/10.1097/corr.0000000000000848.

[9] Chung SW, Han SS, Lee JW, Oh KS, Kim NR, Yoon JP. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop 2018;89:468–73. doi: https://doi.org/10.1080/17453674.2018.1453714.

[10] Urakawa T, Tanaka Y, Goto S, Matsuzawa H, Watanabe K, Endo N. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. Skeletal Radiol 2019;48:239–44. doi: https://doi.org/10.1007/s00256-018-3016-3.
[11] Court-Brown CM, Caesar B. *Epidemiology of adult fractures: a review*. Injury 2006;37:691–7. doi: https://doi.org/10.1016/j.injury.2006.04.130.
[12] Schatzker J, McBroom R, Bruce D. *The tibial plateau fracture. the Toronto experience 1968–1975*. Clin Orthop Relat Res 1979;138:94–104.
[13] te Stroet MAJ, Holla M, van Biert J, Kampen A. The value of a CT scan compared to plain radiographs for the classification and treatment plan in tibial plateau fractures. Emerg Radiol 2011;18:279–83. doi: https://doi.org/10.1007/s10140-010-0932-5.
[14] Doornberg JN, Rademakers MV, van den Bekerom MP, Kerkhoffs GM, Ahn J, Steller EPH. Two-dimensional and three-dimensional computed tomography for the classification and characterisation of tibial plateau fractures. Injury 2011;42:1416–25. doi: https://doi.org/10.1016/j.injury.2011.03.025.
[15] 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;49:473–90. doi: https://doi.org/10.1016/j.injury.2018.01.025.
[16] Brunner A, Honigmann P, Treumann T, Babst R. *The impact of stereo-visualisation of three-dimensional CT datasets on the inter- and intraobserver reliability of the AO/OTA and Neer classifications in the assessment of fractures of the proximal humerus*. J Bone Joint Surg Br 2009;91:766–71. doi: https://doi.org/10.1302/0301-620x.91b6.22109.
[17] Castiglia MT, Nogueira-Barbosa MH, Vieira Messias AM, Salim R, Fogagnolo F, Schatzker J. *The impact of computed tomography on decision making in tibial plateau fractures*. J Knee Surg 2018;31:1007–14. doi: https://doi.org/10.1055/s-0038-1627464.
[18] Chan PS, Klimkiewicz JJ, Luchetti WT, Esterhai JL, Kneeland JB, Dalinka MK. *Impact of CT scan on treatment plan and fracture classification of tibial plateau fractures*. J Orthop Trauma 1997;11:484–9. doi: https://doi.org/10.1097/00005131-199710000-00005.
[19] Molenaars RJ, Mellema JJ, Doornberg JN, Kloen P. *Tibial plateau fracture characteristics: computed tomography mapping of lateral, medial, and bicondylar fractures*. J Bone Joint Surg Am 2015;97:1512–20. doi: https://doi.org/10.2106/jbjs.N.00866.
[20] Fu T, Viswanathan V, Attia A, Zerbib-Attal E, Kosaraju V, Barger R. *Assessing the potential of a deep learning tool to improve fracture detection by radiologists and emergency physicians on extremity radiographs*. Acad Radiol 2024;31:1989–99. doi: https://doi.org/10.1016/j.acra.2023.10.042.
[21] Liu PR, Zhang JY, Xue MD, Duan YY, Hu JL, Liu SX. Artificial intelligence to diagnose tibial plateau fractures: An intelligent assistant for orthopedic physicians. Curr Med Sci 2021;41:1158–64. doi: https://doi.org/10.1007/s11596-021-2501-4.
[22] Lind A, Akbarian E, Olsson S, Nasell H, Sköldenberg O, Razavian AS. Artificial intelligence for the classification of fractures around the knee in adults according to the 2018 AO/OTA classification system. PLoS One 2021;16:. doi: https://doi.org/10.1371/journal.pone.0248809e0248809.
[23] Meinberg E, gel J, Roberts CS, Karam MD, Kellam FF. *Fracture and dislocation classification compendium—2018*. J Orthop Trauma 2018;32:S1–S. doi: https://doi.org/10.1097/bot.0000000000001063.
[24] Olczak J, Pavlopoulos J, Prijs J, IJpma FFA, Doornberg JN, Lundström C. *Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal*. Acta Orthop 2021;92:513–25. doi: https://doi.org/10.1080/17453674.2021.1918389.
[25] Assink N, Vaartjes TP, Kramer CJSA, Bosma E, Nijveldt RJ, ten Brinke JG. what is the patient-reported outcome and complication incidence after operative versus nonoperative treatment of minimally displaced tibial plateau fractures? Clin Orthop Relat Res 2024;482:1744–52. doi: https://doi.org/10.1097/corr.0000000000003057.
[26] Vaartjes TP, Assink N, Nijveldt RJ, van Helden SH, Bosma E, El Moumni M. *Functional outcome after nonoperative management of tibial plateau fractures in skeletally mature patients: what sizes of gaps and stepoffs can be accepted?*. Clin Orthop Relat Res 2022;480:2288–95. doi: https://doi.org/10.1097/CORR.0000000000002266.
[27] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition.
[28] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR).
[29] Huitema JM, van der Gaast N, Jaarsma RL, Doornberg JN, Edwards MJR, Hermans E. *The effect of addition of 2DCT scans and 3DCT scans for the classification of tibial plateau fractures: a systematic review*. Eur J Trauma Emerg Surg 2024;50:71–9. doi: https://doi.org/10.1007/s00068-023-02344-3.
[30] Gahlot N, Kunal K, Elhence A, Meena U, Gupta A, Netji J. *Ten-segment classification has lowest inter/intra-observer reliability as compared to Schatzker, three-column and AO systems for tibial plateau fractures: a comparison based on surgeons' experience*. Arch Bone Jt Surg 2023;11:256–61. doi: https://doi.org/10.22038/abjs.2022.63442.3061.
[31] Besa P, Angulo M, Vial R, Vega R, Irribarra L, Lobos D. *The AO classification system for tibial plateau fractures: an independent inter and intraobserver agreement study*. Injury 2023;54:110741. doi: https://doi.org/10.1016/j.injury.2023.04.028.
[32] Koksal A, Abul K, Kergin D, Cimen O, Mert M, Keskin A. An evaluation of intra and inter observer reliability of the five used classification systems of tibial plateau fractures. Acta Orthop Belg 2022;88:805–10. doi: https://doi.org/10.52628/88.4.9845.
[33] 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;41:173–8. doi: https://doi.org/10.1016/j.injury.2009.08.016.
[34] Masouros PT, Mitrogiannis G, Antoniou G, Chatzidaki C, Kourtzis D, Garnavos C. Interobserver reliability of Schatzker, AO Foundation-Orthopaedic Trauma Association, and Luo classifications for tibial plateau fractures: does three-dimensional CT improve outcomes? Cureus 2022;14:. doi: https://doi.org/10.7759/cureus.22227e22227.
[35] de Lima Lopes C, da Rocha Candido Filho CA, de Lima E Silva TA, Kraue Gonçalves MC, de Oliveira RL, Gomes de Lima PR. Importance of radiological studies by means of computed tomography for managing fractures of the tibial plateau. Rev Bras Ortop 2014;49:593–601. doi: https://doi.org/10.1016/j.rboe.2014.10.010.
[36] Gicquel T, Nijihi N, Venduevre T, Teyssedou S, Gayet LE, Huten D. Tibial plateau fractures: reproducibility of three classifications (Schatzker, AO, Duparc) and a revised Duparc classification. Orthop Traumatol Surg Res 2013;99:805–16. doi: https://doi.org/10.1016/j.otsr.2013.06.007.
[37] Taşkesen A, Demirkale I, Okkaoglu MC, Özdemir M, Bilgili MG, Altay M. *Intraobserver and interobserver reliability assessment of tibial plateau fracture classification systems*. Eklem Hastalik Cerrahisi 2017;28:177–81. doi: https://doi.org/10.5606/ehc.2017.56816.
[38] Huitema JM, van der Gaast N, Brouwers L, Jaarsma RL, Doornberg JN, Edwards MJR. *Are 3D-printed models of tibial plateau fractures a useful addition to understanding fractures for junior surgeons?*. Clin Orthop Relat Res 2022;480:1170–7. doi: https://doi.org/10.1097/corr.0000000000002137.
[39] Luo CF, Sun H, Zhang B, Zeng BF. *Three-column fixation for complex tibial plateau fractures*. J Orthop Trauma 2010;24:683–92. doi: https://doi.org/10.1097/BOT.0b013e3181d436f3.
[40] Krause M, Preiss A, Müller G, Madert J, Fehske K, Neumann MV, et al. *Intra-articular tibial plateau fracture characteristics according to the "Ten segment classification*. Injury 2016;47:2551–7. doi: https://doi.org/10.1016/j.injury.2016.09.014.
[41] Anwar A, Zhang Y, Zhao Z, Gao Y, Sha L, Lv D. Two column classification of tibial plateau fractures; description, clinical application and reliability. Injury 2019;50:1247–55. doi: https://doi.org/10.1016/j.injury.2019.04.021.