Identifying Volar Locking Plates on Plain Radiographs: Can Artificial Intelligence Models 'Beat' Clinicians?

Allen Albert [1], Alex Nicholls [2], Duncan Avis [2], Medhat Zekry [1], Matay Arsan [1], Priyanshu Saha [1], Adam Stoneham [1]

1. Trauma and Orthopaedics, Homerton University Hospital, London, GBR 2. Trauma and Orthopaedics, Hampshire Hospitals NHS Foundation Trust, Hampshire, GBR

Corresponding author: Allen Albert, allen.jasper.durai@gmail.com

Abstract

Background

Hand surgeons are frequently required to identify volar locking plates on plain radiographs. This is important, for example, when planning what equipment is required for revision, implant removal or periprosthetic fractures, but it can be challenging, especially if surgery took place in another hospital or even country. Artificial intelligence (AI) clearly has potential in medical image recognition, but its role in orthopaedic implant identification currently remains uncertain. This study compared the performance of an openly available AI model, ChatGPT 5, with that of experienced hand consultants.

Methods

Fifty-two radiographs of distal radius plates from 10 major implant manufacturers were obtained from open-access sources. An AI programme (ChatGPT 5) and five hand consultants independently identified the manufacturer for each radiograph. Accuracy was calculated for each rater. Pairwise comparisons between AI and each consultant were assessed using McNemar's test, and a standard logistic regression with clustered standard errors was fitted to compare AI with consultants as a group.

Results

ChatGPT5 correctly identified just 3 of 52 radiographs (5.8% accuracy). Consultant accuracies ranged between 13.5 and 46.2% (mean 30.8% ± 11.1). McNemar's test showed that Consultants 1, 3, 4 and 5 significantly outperformed the AI (p < 0.01), while Consultant 2 did not (p = 0.289). In a standard logistic regression with clustered standard errors, the human cohort had 7.26 times higher odds of correct identification compared with the AI (OR 7.26, 95% CI 2.27-23.18, p < 0.001).

Conclusion

Identifying volar locking distal radius plates from plain radiographs remains difficult, even for experienced surgeons and even the 'best' consultant identified under 50% correctly. That notwithstanding, humans were on average seven times better than ChatGPT 5 which only identified just over 5% correctly. While current non-specialised AI tools are not suitable for implant identification currently, dedicated AI models trained on curated orthopaedic datasets may hold promise for future clinical use.

Introduction

Distal radius fractures rank amongst the commonest adult orthopaedic injuries; the incidence has been steadily rising over the last 20 years, and this trend is expected to continue [1]. Treatment can be nonoperative or operative depending on many factors including the degree of displacement and the patients' age and comorbidities [2], while significant variation exists in clinical practice [3]. Although good results are achievable with percutaneous wire fixation [4,5], there has been a steady rise in the use of internal fixation in North America [6,7] and Europe [8,9] and volar locking plates have risen from around 20% of surgeries in 1997 to over 90% in 2018 [10].

Hardware removal following open reduction and internal fixation is controversial with wide variation in practice between different surgeons in different countries [11]. Although there is little support for routine removal in asymptomatic patients, justifiable indications include pain, stiffness, or irritation and the procedure is successful and straightforward in many cases [12,13]. Other indications for removal might include revision or in the instance of periprosthetic fractures [14].

To remove the plate and screws, it is necessary to know what hardware was used during the index procedure and, in particular, what implant-specific equipment (e.g. screwdriver) will be required. This can be challenging, especially if the surgery took place in another institution or country, in emergency settings, or if the operative notes are not available. Very little evidence exists regarding surgeons' ability to determine the make and model of distal radius plates based on plain radiographs alone. Artificial Intelligence (AI) may already have a role in detecting distal radius fractures [15,16]. We postulate that it will also have a growing role in helping surgeons in this situation, thereby potentially reducing operative time and complications.

Materials And Methods

Fifty-two AP radiographs of in-situ distal radius plates were obtained from open-access online sources to obviate issues surrounding copyright or patient confidentiality. The images included 10 major international manufacturers, with at least five radiographs per brand to ensure balanced representation. Each radiograph was independently verified against manufacturer catalogues and online implant databases to confirm brand identity. The included manufacturers and their parent companies are listed in Table 1.

Manufacturer	Parent company/Corporate group, Headquarters
Acumed	Colson Medical / Marmon Holdings (Berkshire Hathaway), USA
Arthrex	Arthrex, Inc. (independent), USA
DePuy Synthes	Johnson & Johnson MedTech, USA
KLS Martin	KLS Martin SE & Co. KG, Germany
Skeletal Dynamics	Skeletal Dynamics LLC (independent), USA
Newclip Xpert	Newclip Technics, France
Smith & Nephew	Smith & Nephew plc, UK
Stryker	Stryker Corporation, USA
Medartis	Medartis AG, Switzerland
Zimmer Biomet	Zimmer Biomet Holdings, Inc., USA

TABLE 1: Distal radius plate manufacturers, parent companies, and countries of origin

An openly available artificial intelligence model, ChatGPT 5 (paid version), OpenAI, USA was asked to correctly identify the manufacturer of each distal radius plate. The same set of radiographs (in a random order) was also distributed to five orthopaedic hand consultants from two different hospitals in England, who were independently asked to identify the manufacturer of the plates.

All responses, both from ChatGPT and from the consultants, were recorded and scored using a binary coding system: 1 for a correct identification and 0 for an incorrect identification. Statistical analysis was performed using R (version 4.5.1), with a significance threshold set at $p < 0.05$. Accuracy was calculated for each rater. For pairwise comparisons between each consultant and AI, McNemar's test was performed to report the matched odds ratio. To compare AI performance with all consultants combined, a standard logistic regression with clustered standard errors was performed to report odds ratios (ORs) and 95% confidence intervals (CIs).

Results

Five consultant hand surgeons took part in this study with a total of 34.5 years of consultant experience (range 1 - 12, mean 6.9, Standard Deviation 3.9).

Across the 52 radiographs, the AI programme correctly identified three implants (5.8% accuracy) while the five hand consultants (Consultants 1, 2, 3, 4 and 5) achieved accuracies of 38.5% (20/52), 13.5% (7/52), 46.2% (24/52), 28.8% (15/52), and 26.9% (14/52), respectively. The mean consultant accuracy was 30.8% (SD 11.1; range 13.5 - 46.2%) (Table 2).

2025 Albert et al. Cureus 17(12): e98715. DOI 10.7759/cureus.98715

2 of 6

Rater	Correct / (n)	Accuracy (%)
AI programme	3/52	5.80%
Consultant 1	20/52	38.50%
Consultant 2	7/52	13.50%
Consultant 3	24/52	46.20%
Consultant 4	15/52	28.80%
Consultant 5	14/52	26.90%
Consultants (mean ± SD)	–	30.8% ± 11.1
Range (min–max)	–	13.5 – 46.2%

TABLE 2: Accuracy of AI model and five independent hand consultants in identifying volar locking distal radius plate manufacturer (n = 52 radiographs)

Accuracy was calculated as the proportion of correctly identified plates. Descriptive statistics for human raters (Consultants) include the mean ± standard deviation (SD) and the observed range (minimum–maximum).

Pairwise comparisons using McNemar's χ^2 test demonstrated that four of the consultants significantly outperformed the AI (Table 3).

Comparison	Both Correct	AI Correct/Consultant Wrong	AI Wrong/Consultant Correct	Both Wrong	Matched OR	$\chi^2(1)$	p-value
AI vs Consultant 1	3	0	17	32	∞	15.06	<0.001
AI vs Consultant 2	1	2	6	43	3	1.13	0.289
AI vs Consultant 3	2	1	22	27	22	17.39	<0.001
AI vs Consultant 4	0	3	15	34	5	6.72	<0.01
AI vs Consultant 5	2	1	12	37	12	7.69	<0.01

TABLE 3: Pairwise comparison of AI vs individual consultants using McNemar's χ^2 test

Cross-tabulation of correct and incorrect responses for the AI programme versus each consultant across 52 radiographs. "Both Correct" and "Both Wrong" indicate cases where the AI and the consultant produced identical outcomes. "AI Correct / Consultant Wrong" and "AI Wrong / Consultant Correct" show discordant cases and these were used to create contingency tables for McNemar's analysis. The matched odds ratio (OR) quantifies the relative odds of correct identification by the consultant compared with the AI. Statistical significance was evaluated with McNemar's χ^2 test with continuity correction (df=1) and it was used to compare correct identification rates between the AI and each consultant. Both the χ^2 statistic and p-value are shown. Statistical significance was defined as $p < 0.05$ (corresponding to $\chi^2 \geq 3.84$).

When all human consultants were considered together, a standard logistic regression with clustered standard errors accounting for repeated measures within radiographs showed that humans had 7.26 times higher odds of correct identification compared with the AI programme (OR 7.26, 95% CI 2.27 - 23.18, Wald z 2.97, $p < 0.001$) (Table 4).

Comparison	Odds Ratio	95% CI (Lower–Upper)	Wald z	p-value
All five consultants vs AI	7.26	2.27 – 23.18	2.97	0.000817

TABLE 4: Clustered logistic regression comparing AI vs human consultants

A standard logistic regression with clustered standard errors was fitted to compare AI and consultant accuracy. The Wald z-statistic, Odds ratio (OR) and 95% confidence intervals (CI) are presented. Significance threshold: $p < 0.05$ (corresponding to $|z| \geq 2.58$)

Discussion

This study compared the performance of an openly available AI model, ChatGPT 5, with that of experienced hand surgeons in identifying distal radius plate manufacturers from plain radiographs.

This task proved difficult for both humans and machines. The surgeons in this study, even with nearly 35 years of combined consultant experience, were only able to identify around half of the plates at best, reflecting differences in training or experience, overlap in design features across manufacturers and the inherent limitations of radiographs in distinguishing fine details or identifying features.

Nevertheless, clinicians consistently demonstrated superior recognition performance compared with AI. Across 52 radiographs, representing 10 major brands, all four consultants - even the 'worst' outperformed AI, which achieved an accuracy of just 5.8%. Although there was variability in human performance (13.5-46.2% correctly identified), when all consultants were considered together, a standard logistic regression with clustered standard errors showed that humans had over 7-fold higher odds of correct identification compared with the AI.

Although AI models can be trained for many aspects of medical image interpretation including hand and wrist fracture detection [17,18], the performance of an untrained model on this task is poor, underlying the limitations of general-purpose AI models when applied to specialised medical tasks [19]. Targeted training on datasets of orthopaedic implants therefore would be necessary to achieve clinically useful performance.

Limitations

Radiographs were only obtained in their AP projection and from open-access sources, so they may not fully reflect the range of quality or variation encountered in clinical practice. We only selected 10 manufacturers, and so inevitably many other smaller brands were excluded. 52 examples were selected (at least five of each). Our original power calculations did not anticipate that either group (but AI in particular) would identify so few plates correctly, and so this is probably an underestimation of the number required. Surgeons themselves often rely on a relatively small repertoire of implants determined by institutional preferences, training, and experience. Recruiting busy consultant hand surgeons is challenging; a larger number would not only have permitted us to calculate confidence intervals, but it would have potentially expanded the familiarity with a greater number of implants. Finally, the AI model evaluated was not specifically designed for implant recognition, which likely contributed to its low accuracy. This is a rapidly moving field, and our study only represents a snapshot in time; were it to be repeated in just 6 or 12 months, the results would likely be different.

Conclusions

Volar locking distal radius plates are commonly used in the treatment of fractures or following corrective osteotomies. Recognising the plate manufacturer from plain radiographs is important when planning hardware revision, removal or periprosthetic fractures. In this study, even experienced hand surgeons struggled to identify 10 commonly used plating systems (13-46% correct). Nevertheless, they remained considerably superior to the performance of ChatGPT 5 on the same task (< 6%). Well-trained, custom-built AI systems are required to help in orthopaedic hardware recognition and, in time, are expected to become a useful tool in this domain.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Allen Albert, Alex Nicholls, Duncan Avis, Medhat Zekry, Matay Arsan, Adam Stoneham

Acquisition, analysis, or interpretation of data: Allen Albert, Priyanshu Saha, Adam Stoneham

Drafting of the manuscript: Allen Albert

Critical review of the manuscript for important intellectual content: Allen Albert, Alex Nicholls, Duncan Avis, Medhat Zekry, Matay Arsan, Priyanshu Saha, Adam Stoneham

Supervision: Adam Stoneham

Disclosures

Human subjects: Informed consent for treatment and open access publication was obtained or waived by all participants in this study. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Intellectual property info:** We have written to a few manufacturers to include the x-ray of their product to this research paper. Should we receive permission before the acceptance of the paper, we may add some of the images to the manuscript. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

References

1. Stirling ER, Johnson NA, Dias JJ: Epidemiology of distal radius fractures in a geographically defined adult population. J Hand Surg Eur Vol. 2018, 43:974-82. 10.1177/1753193418786378
2. Lichtman DM, Bindra RR, Boyer MI, et al.: Treatment of distal radius fractures . J Am Acad Orthop Surg. 2010, 18:180-9. 10.5435/00124635-201003000-00007
3. Walenkamp MM, Mulders MA, Goslings JC, Westert GP, Schep NW: Analysis of variation in the surgical treatment of patients with distal radial fractures in the Netherlands. J Hand Surg Eur Vol. 2017, 42:39-44. 10.1177/1753193416651577
4. Karantana A, Downing ND, Forward DP, et al.: Surgical treatment of distal radial fractures with a volar locking plate versus conventional percutaneous methods: a randomized controlled trial. J Bone Joint Surg Am. 2013, 95:1737-44. 10.2106/JBJS.L.00232
5. Costa ML, Achten J, Parsons NR, Rangan A, Griffin D, Tubeuf S, Lamb SE: Percutaneous fixation with Kirschner wires versus volar locking plate fixation in adults with dorsally displaced fracture of distal radius: randomised controlled trial. BMJ. 2014, 349:g4807. 10.1136/bmj.g4807
6. Azad A, Kang HP, Alluri RK, Vakhshori V, Kay HF, Ghiassi A: Epidemiological and treatment trends of distal radius fractures across multiple age groups. J Wrist Surg. 2019, 8:305-11. 10.1055/s-0039-1685205
7. Chung KC, Shauver MJ, Birkmeyer JD: Trends in the United States in the treatment of distal radial fractures in the elderly. J Bone Joint Surg Am. 2009, 91:1868-73. 10.2106/JBJS.H.01297
8. Mattila VM, Huttunen TT, Sillanpää P, Niemi S, Pihlajamäki H, Kannus P: Significant change in the surgical treatment of distal radius fractures: a nationwide study between 1998 and 2008 in Finland. J Trauma. 2011, 71:939-42; discussion 942-3. 10.1097/TA.0b013e3182231af9
9. Mellstrand-Navarro C, Pettersson HJ, Tornqvist H, Ponzer S: The operative treatment of fractures of the distal radius is increasing: results from a nationwide Swedish study. Bone Joint J. 2014, 96-B:963-9. 10.1302/0301-620X.96B7.33149
10. Viberg B, Tofte S, Rønnegaard AB, Jensen SS, Karimi D, Gundtoft PH: Changes in the incidence and treatment of distal radius fractures in adults - a 22-year nationwide register study of 276,145 fractures. Injury. 2023, 54:110802. 10.1016/j.injury.2023.05.033
11. Yamamoto M, Fujihara Y, Fujihara N, Hirata H: A systematic review of volar locking plate removal after distal radius fracture. Injury. 2017, 48:2650-6.
12. Kılıç E, Savaş MF: Does removal of volar locking plate affect patient functional outcomes? . ACH Med J. 2024, 3:90-5. 10.5505/achmedj.2024.73745
13. Lee JK, Lee Y, Kim C, Kim M, Han SH: Volar locking plate removal after distal radius fracture: a 10-year retrospective study. Arch Orthop Trauma Surg. 2021, 141:1711-9. 10.1007/s00402-020-03637-7
14. Kanji R, Nutt J, Stavropoulos S, Elmorsy A, Schneider HJ: Distal radius re-fracture with bending of implant and neurovascular compromise. J Clin Orthop Trauma. 2017, 8:S40-2. 10.1016/j.jcot.2017.08.009
15. Suzuki T, Maki S, Yamazaki T, et al.: Detecting distal radial fractures from wrist radiographs using a deep convolutional neural network with an accuracy comparable to hand orthopedic surgeons. J Digit Imaging. 2022, 35:39-46. 10.1007/s10278-021-00519-1
16. Oka K, Shiode R, Yoshii Y, Tanaka H, Iwahashi T, Murase T: Artificial intelligence to diagnosis distal radius fracture using biplane plain X-rays. J Orthop Surg Res. 2021, 16:694. 10.1186/s13018-021-02845-0
17. Kraus M, Anteby R, Konen E, Eshed I, Klang E: Artificial intelligence for X-ray scaphoid fracture detection: a systematic review and diagnostic test accuracy meta-analysis. Eur Radiol. 2024, 34:4341-51. 10.1007/s00330-023-10473-x
18. Wong CR, Zhu A, Baltzer HL: The accuracy of artificial intelligence models in hand/wrist fracture and

dislocation diagnosis: a systematic review and meta-analysis. JBJS Rev. 2024, 12:10.2106/JBJS.RVW.24.00106

19. Sarangi PK, Datta S, Panda BB, Panda S, Mondal H: Evaluating ChatGPT-4's performance in identifying radiological anatomy in FRCR Part 1 examination questions. Indian J Radiol Imaging. 2025, 35:287-94. 10.1055/s-0044-1792040

2025 Albert et al. Cureus 17(12): e98715. DOI 10.7759/cureus.98715

6 of 6