OPEN ACCESS

Journal of Orthopedic Research and Therapy

Mishra AS, et al. J Orthop Res Ther 11: 1429.

www.doi.org/10.29011/2575-8241.001429

www.gavinpublishers.com

GAVIN PUBLISHERS

Research Article

Artificial Intelligence to Evaluate Radiographic Outcome following Hallux Valgus Surgery

Arya Suman Mishra[1]*, Hannah Meacher[1], Muhammad Murtaza Khan[1], Zaina Gaddoura[2], Matthew Hough[2], Abu Saeed[1], Catriona Heaver[1], Chris Marquis[1], Andrew Bing[1], Simon Hill[1], Nilesh Makwana[1], Theocharis Kyriacou[2], Paul Cool[1,2]

[1]RJAH Orthopaedic Hospital, Oswestry, UK

[2]Keele University, UK

***Corresponding author:** Arya Suman Mishra, Robert Jones and Agnes Hunt Orthopaedic Hospital Twympath Lane, Gobowen, Oswestry, SY10 7AG, UK

Citation: Mishra AS, Meacher H, Khan MM, Gaddoura Z, Hough M, et al. (2026) Artificial Intelligence to Evaluate Radiographic Outcome following Hallux Valgus Surgery. J Orthop Res Ther 11: 1429. https://doi.org/10.29011/2575-8241.001429

Received Date: September 10, 2026; **Accepted Date:** September 22, 2026; **Published Date:** September 25, 2026

Abstract

Purpose: Angular measurements in Hallux valgus are commonly used, but the methods described vary, particularly for post-operative films. This study aims to assess this variation, evaluate a new artificial-intelligence-based computer method, and determine whether it can serve as an objective outcome measure. **Methods:** Consecutive patients who had surgery for hallux valgus over six months were included. We collected information on age, gender, and type of surgery. Hallux valgus (HV), intermetatarsal (IM) and distal metatarsal articular (DMA) angles were measured on pre- and post-operative radiographs by seven observers of varying degrees of experience. We also read the images into the computer, and after users identified five bony landmarks, we segmented them with Segment Anything (SAM)® software. We then used the first eigenvector to define the long axis of each bone and calculated angles. **Results:** In both pre-operative and post-operative radiographs, the computer method showed less variation than human observers for all angles. For post-operative radiographs, the ICC was rated excellent for the HV angle and good for the IM angle. The computer method was considerably faster than human observers. **Conclusion:** The described computer method is reproducible with less variation and bias than human observers. In addition, it can be used on post-operative radiographs as an objective outcome measure.

Keywords: Hallux valgus, Angle measurement, Observer error, Artificial intelligence, Outcome measure.

Introduction

Angular measurements taken before corrective surgery for hallux valgus can be helpful in decision-making (Table 1) [1]. Many different methods have been described and the American Association of Orthopaedic Surgeons (AAOS) has published guidelines recommending a standard method [2-7]. The published literature on the reliability of this method in pre-operative radiographs shows reasonable consistency between observers for the hallux valgus (HV) angle and intermetatarsal (IM) angle [3,8]. But in clinical practice, the observer variation is felt to be larger. Besides, these angles have been shown to change with weight-bearing and metatarsal rotation, making comparison more difficult [3,9]. Computer models using artificial intelligence and neural networks have also been developed, but they have been trained on relatively few human observations, lack a clear definition of the central axis of the bone and use pre-operative radiographs [10].

Deformity	Hallux valgus angle (degrees)	Intermetatarsal angle (degrees)
Normal	< 15	< 9
Mild	>= 15 and < 20	>= 9 and < 11
Moderate	>= 20 and < 40	>= 11 and < 16
Severe	>= 40	>= 16

Table 1: Severity grading of Hallux Valgus based on HV and IM angles.

Ideally, the same method should be applied to the post-operative radiographs if these measures are to serve as an objective outcome measure. But it is difficult to use the same method in post-operative films due to the wide variety of osteotomy sites and configurations. Measurements are particularly complicated when a scarf osteotomy has been used to correct the deformity [11,12]. Very few studies have been conducted to test the reliability of the described methods in post-operative radiographs, none of which use the scarf osteotomy [13,14].

Good reliability has also been demonstrated using semi-automated angle measurements in weight-bearing Computed Tomography scans. These methods hold promise for the future when this modality may become more widely available and utilised routinely [15].

The aims of this study are to:

1. Assess the variation in angles measured by human observers in patients with hallux valgus surgery

2. Evaluate a computer model that calculates the angles, with the central axis of each bone defined as the first eigenvector of the segmented shape.

3. Compare computer and human measurements for reproducibility in preoperative and postoperative radiographs.

Methods

Sample size calculation was performed with G*Power (version 3.1.9.5) [16]. This indicated a minimum of 40 patients were required to detect a difference of 5 degrees between human and computer measurements (effect size of 0.6) using two-tailed paired Wilcoxon test with 80% power and a significance level of 5%. Therefore, all patients who underwent surgical correction for hallux valgus within a six-month period (between September 2023 and April 2024) were included to obtain sufficient data. Patients who had procedures other than an osteotomy (Keller's procedure or fusion) were excluded. There were no patients lost to follow up.

The HV, IM and distal metatarsal articular (DMA) angles were measured on preoperative, intraoperative and postoperative radiographs. All pre- and postoperative radiographs were taken in a weight-bearing position, with the postoperative images taken at least five weeks following surgery. Seven observers (four specialist registrars, two medical students and one consultant) measured the angles on-screen using the built-in angle measurement function of our Picture Archiving and Communication System (PACS; Synapse® version 4.4.4, Fujifilm® healthcare, USA, 2023). Each observer measured the angles on two separate occasions; at least two weeks apart.

A semi-automated computer model (see appendix for packages) was used, built on Python version 3.10 (Python Software Foundation, https://www.python.org/). Radiographs were imported with the pydicom package (https://pydicom.github.io/, see appendix) and converted to arrays. Segment Anything (SAM, Meta AI, Toronto, https://segment-anything.com/) software was used to segment the bones (Figure 1B) [17]. Five points were selected on the image to facilitate segmentation: the proximal phalanx, the first metatarsal, the second metatarsal, the medial cuneiform and the intermediate cuneiform (Figure 1A). Using the coordinates of the segmented contours, the first eigen vectors of the proximal phalanx, first metatarsal and second metatarsal were used to define the central axes (Figure 1C). The distal metatarsal articular surface was defined using 75 pixels of the metatarsal head contour on either side of the most distal point of the first metatarsal segmentation along the first eigenvector. The tangent to this surface was taken as the distal first metatarsal joint line. The HV, IM and DMA angles were then calculated from the central axes and distal first metatarsal joint line. Implanted metalwork can interfere with the segmentation process. Consequently, the metalwork was 'removed' with software using

thresholding (pixel value over 200) and restored with the simple-lama-inpainting algorithm (Figure 2) before segmentation [18,19].



Figure 1: Weight bearing radiographs of left foot used for segmentation with Segment Anything (SAM) software12 (a) User selected proximal phalanx, 1st and 2nd metatarsal, medial and intermediate cuneiforms (b) Segmentation with Segment Anything (SAM) software (c) Central axes found from the 1st eigenvector of the segmentations of the proximal phalanx and 1st and 2nd metatarsals. Hallux valgus angle 34.6 degrees, intermetatarsal angle: 14.7 degrees.

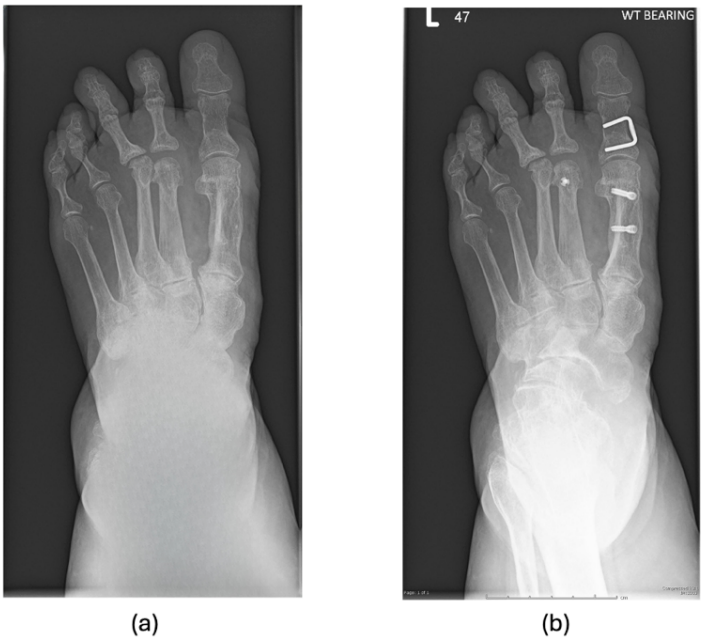


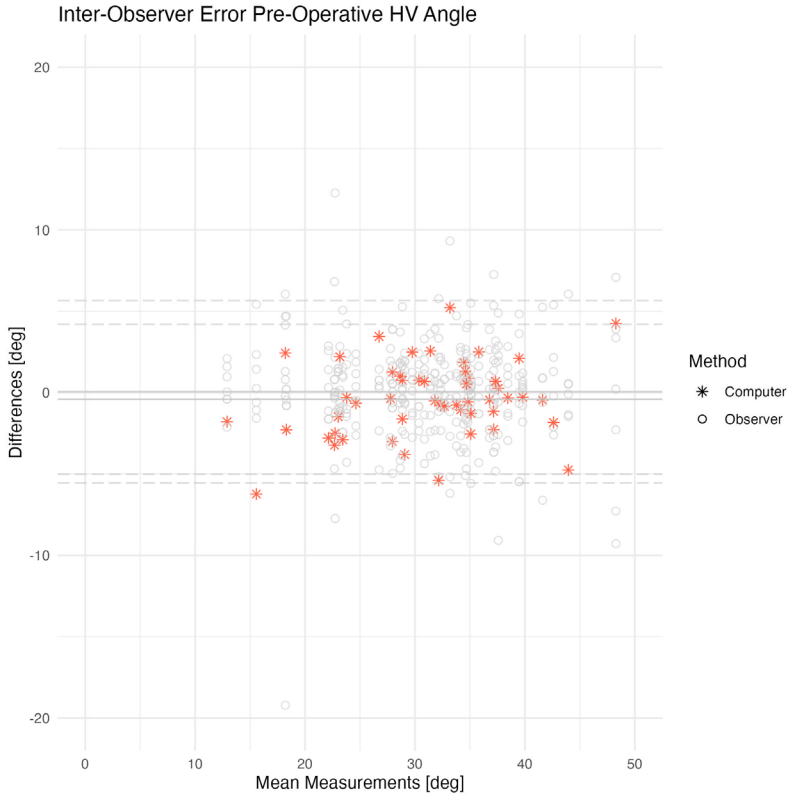
Figure 2: Metalwork needs to be removed with software as it interferes with segmentation. (a) Weight bearing radiograph with metalwork in situ following surgery for hallux valgus (b) Software removal of metalwork by thresholding pixel values over 200 (white) and 'inpainted' with the simple-lama-inpainting algorithm [13,14].

Statistical analysis was performed with R statistical software version 4.4.3 (R Core Team) and the ggplot2 package was used to produce the plots [20,21]. Numerical variables could not be modelled with a normal distribution and were summarised with the median and interquartile range. Categorical variables were expressed as numbers or proportions. Bland-Altman plots were used to compare measurements for each angle and visualise observer error [22]. For intra-observer error, the mean of the two values recorded by the same observer was used as the true value of the measurement. For inter-observer error, this was the mean of all values recorded by all observers for the measurement. The intraclass correlation coefficient (ICC) was calculated with the irr package [23]. An ICC of less than 0.5 was regarded as poor, between 0.5 and 0.75 as moderate, between 0.75 and 0.9 as good, and more than 0.9 as excellent. Differences between groups were compared using the Wilcoxon signed-rank test at a 5% significance level. The article was structured in accordance with the STROBE checklist for cohort studies [24].

Results

51 patients underwent surgery for correction of Hallux Valgus during the study period. Three patients were excluded because they had undergone first metatarsophalangeal joint fusion (two patients) and Keller's procedure (one patient) for correction of their deformity, leaving 48 patients who were included in the study. There were 46 female and two male patients. All patients had unilateral procedures, with the right side being operated in 44%. The median age was 58 years (IQR 51-64 years). The most common metatarsal osteotomy used was a scarf osteotomy (40 patients), followed by a chevron osteotomy (eight patients). Among these, concomitant Akin procedures were performed in 27 (68%) and four (50%) patients, respectively. The postoperative weight-bearing radiographs were obtained at a median 46 days (IQR: 38-54) after the surgical procedure.

The inter-observer and intra-observer variation in preoperative and postoperative radiographs for HV, IM and DMA angles is presented in Figures 3, 4 and 5, respectively. Table 2 presents the ICCs for inter-observer and intra-observer error on preoperative, intraoperative and postoperative radiographs.



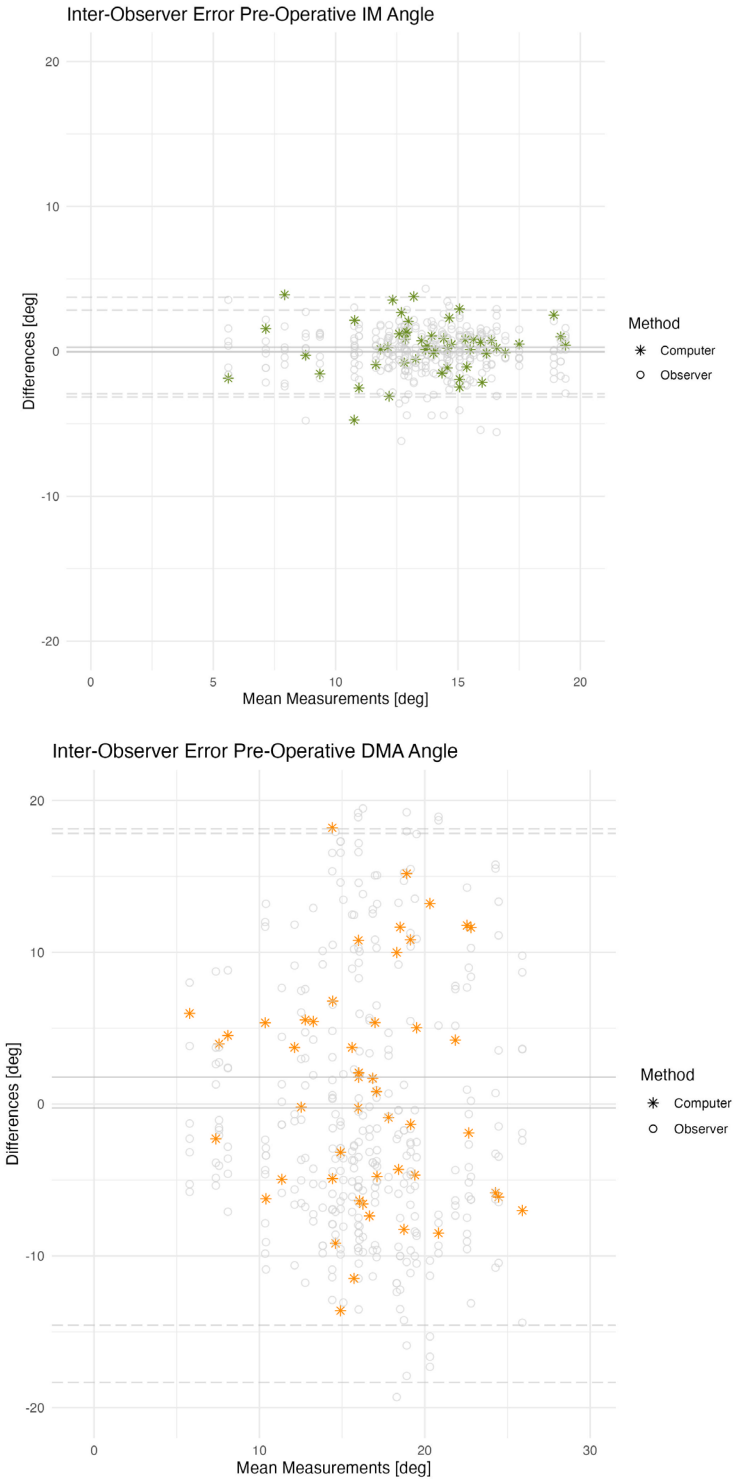
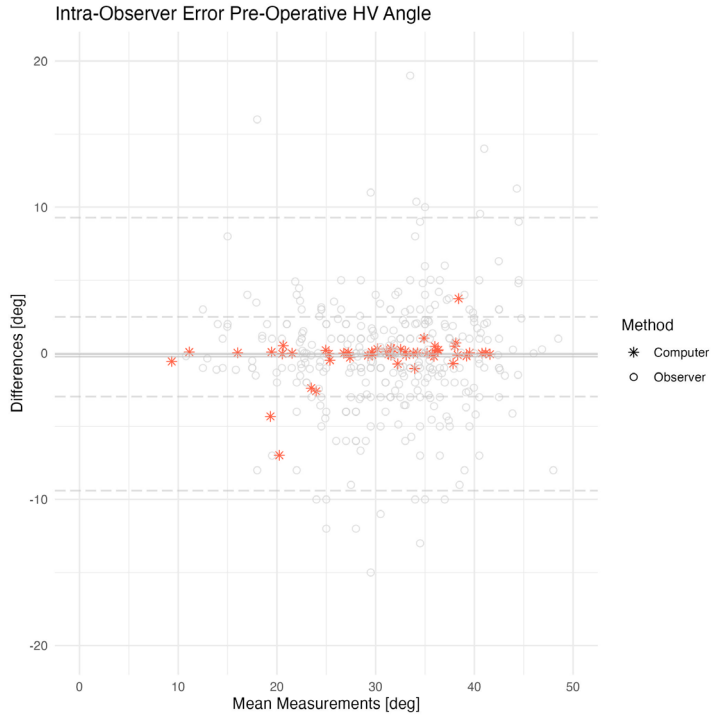
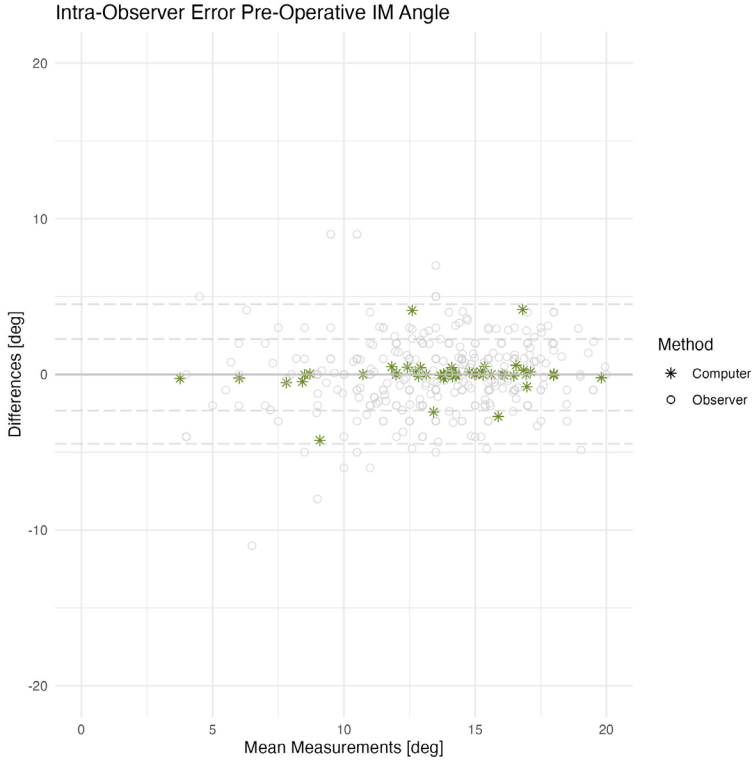


Figure 3: Bland-Altman plots of inter-observer error in the measurement of angle in pre-operative radiographs.


Intra-Observer Error Pre-Operative HV Angle
Differences [deg]
Mean Measurements [deg]
Method
Computer
Observer


Intra-Observer Error Pre-Operative IM Angle
Differences [deg]
Mean Measurements [deg]
Method
Computer
Observer

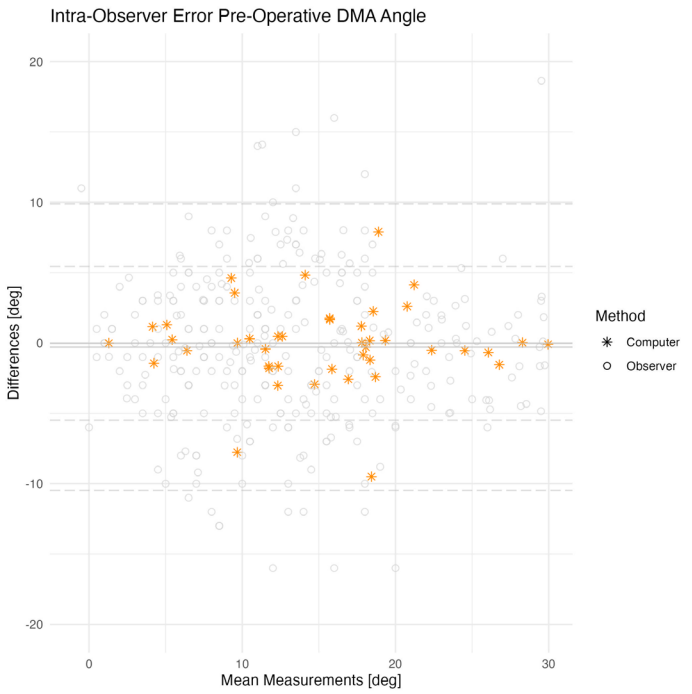


Figure 4: Bland-Altman plots for intra-observer error in the measurement of angles in pre-operative radiographs.

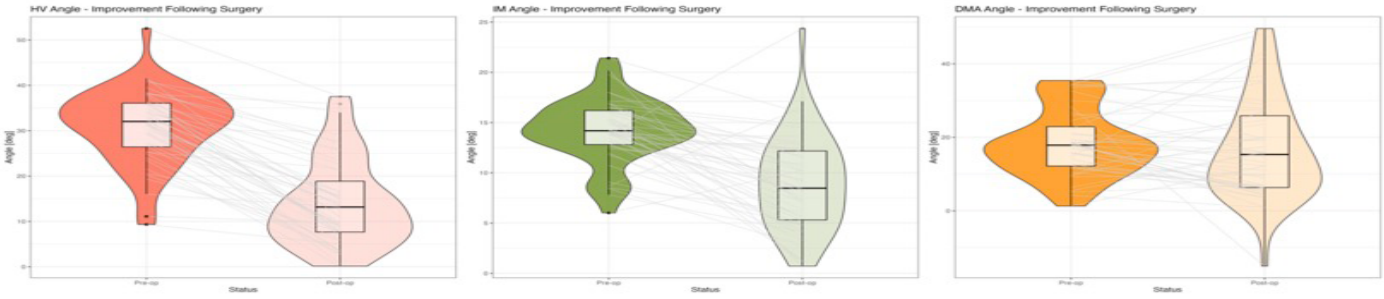


Figure 5: Improvement in angles following surgery, measured using the computer method.

	Inter-observer ICC	Intra-observer ICC - Observers	Intra-observer ICC - Computer
Pre-operative			
HV angle	0.856 (0.793- 0.907)	0.828 (0.793- 0.858)	**0.986** (0.975- 0.992)
IM angle	0.747 (0.653-0.829)	0.767 (0.721- 0.807)	**0.945** (0.905- 0.968)
DMA angle	0.138 (0.054-0.255)	0.879 (0.853- 0.901)	**0.956** (0.924- 0.975)
Intra-operative			
HV angle	0.485 (0.316- 0.656)	0.751 (0.702- 0.793)	0.618 (0.355- 0.791)
IM angle	0.189 (0.087- 0.339)	0.679 (0.612- 0.736)	0.712 (0.493- 0.846)
DMA angle	0.015 (-0.021- 0.082)	0.627 (0.559- 0.687)	0.488 (0.175- 0.711)
Post-operative			
HV angle	0.551 (0.436- 0.672)	0.373 (0.281- 0.459)	**0.924** (0.87- 0.956)

IM angle	0.439 (0.294- 0.589)	0.755 (0.7- 0.8)	0.892 (0.817- 0.938)
DMA angle	0.204 (0.117- 0.325)	0.784 (0.741- 0.821)	0.763 (0.599- 0.862)

Table 2: Intraclass Correlation Coefficient for inter-observer and intra-observer variation in measurements. Upper and lower 95% confidence limits are also presented. ICCs rated excellent are in bold.

Computer and human measurements were compared with the paired Wilcoxon signed rank test. This confirms there is no significant difference in the estimated HV and IM angles between human observers and the computer. Bland-Altman plots also confirm there is no bias in the measurements (Figures 3 and 4). However, there is significantly less variation in the pre-operative and post-operative computer measurements of the HV and IM angle compared to human observations ($p<0.001$). The variation in the intra-operative measurement of the hallux valgus angle using the computer method was not significantly different to human observations ($p=0.296$), but the IM angle was ($p=0.014$).

Table 3 presents the median and interquartile range for measurements using the computer method. Figure 5 presents a plot showing improvement in angles measured by the computer method with surgical correction. As expected, there was a significant improvement in HV and IM angles postoperatively ($p<0.001$). However there was no significant difference in the DMA angle ($p=0.374$). There was no significant difference between intraoperative and postoperative measurements (p-values for HV angle= 0.366, IM angle= 0.918, DMA angle= 0.168).

The time taken for a set of measurements using the computer method was 23 seconds, compared to a median of 3 minutes for human observers (IQR: 1.6–4.4 minutes).

	HV angle	**IM angle**	**DMA angle**
Pre-operative	31.8 (26.3- 37.3)	14.1 (12.6- 15.6)	17.8 (11.8- 23.8)
Intra-operative	11.5 (6.5- 16.5)	8.8 (4.8- 12.8)	16.1 (4.1- 28.1)
Post-operative	13.2 (6.2- 20.2)	8.5 (5- 12)	15.3 (5.3- 25.3)

Table 3: Median and IQR of different angles measured using the computer method. All angles are in degrees.

Discussion

Our results (Table 2) show that the inter-observer ICC was good for the HV angle and moderate for the IM angle in pre-operative radiographs. The DMA angle measurements demonstrate poor inter-observer agreement but good to excellent intra-observer agreement. For all three angles (HV, IM and DMA) the computer method had an excellent ICC for intra-observer error with the smallest confidence intervals, that was significantly better than human observations. In postoperative radiographs, the computer method was significantly more reproducible in measuring the HV and IM angles than human observers, with excellent intra-observer ICCs. The intra-observer ICC for the DMAA angle was, however, lower than that of human observers, yet still considered good. A lack of bias with the computer method was also evident, as the values recorded by it were not significantly different from those of human observers for both pre- and postoperative radiographs.

Measurements of intra-operative radiographs are more challenging as they are non-weight-bearing and prone to positional changes. This is confirmed by human and computer observations that showed poor association. However, the post- and intra-operative computer measurements were not significantly different and it is felt they can still be of some use, but should be interpreted with caution.

It is challenging to apply a single method of angle measurement to all osteotomy sites and geometries that can be used for correction, leading to further variability in postoperative films. This is reflected in the AOFAS guidelines, which recommend using two different methods in postoperative radiographs. Only two studies have evaluated the reliability of these various methods in the postoperative scenario, and neither of them utilised the scarf osteotomy [13,14]. The method described in this paper uses a clear definition that is reproducible and can be applied to all types of osteotomies. Consequently, this method can be used as an objective outcome measure following surgery.

Other computer models based on deep learning convolutional neural networks use lines or points defined by human observers to train the network. Results are biased by human opinion and lack a definition for the central axes to calculate angles. [10] Again, their use in post-operative radiographs has not been evaluated. In contrast, our computer method defines the central axis as the first

eigenvector of the segmented bone. This mathematical definition eliminates variation, allowing it to be used as an objective outcome measure.

The improved consistency in measuring angles in preoperative and postoperative radiographs can enable the use of these angles as an objective outcome measure for improvement following surgery. The computer method is also much faster than human observers, and larger series can be evaluated more efficiently. It can also be easily replicated with minimal cost, as it utilises open-source software, easy-to-use open-source code, and does not require an extensive training dataset. All the resources necessary have been presented in the appendix.

One of the limitations of our computer method is the persistence of measurement variation. This is caused by differences in segmentation by the SAM algorithm, which results in slightly different bone contours being identified, in turn changing the measured eigenvector. The root cause is that the SAM algorithm is sensitive to variations in the grayscale levels of the identification points on the bones. In the future, removing human interaction and identifying segmentation points using artificial intelligence is likely to reduce or eliminate this variation. This will also help address the other limitations, which are a lack of full automation and the need for human input.

Conclusion

The computer method described is more reproducible and faster than human observers. Additionally, it can serve as an objective measure of outcome following hallux valgus surgery.

Declaration of conflicts of interest

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article

Ethical Guidelines

Ethical approval was not sought for the present study, as it is an anonymous retrospective review of radiographs which is part of a larger study of AI in the diagnosis of Orthopaedic conditions.

References

1. Ray JJ, Friedmann AJ, Hanselman AE, Vaida J, Dayton PD, et al. (2019) Hallux Valgus. Foot Ankle Orthop 4(2): 247301141983850.
2. Coughlin MJ, Saltzman CL, Nunley JA (2002) Angular Measurements in the Evaluation of Hallux Valgus Deformities: A Report of the Ad Hoc Committee of the American Orthopædic Foot and Ankle Society on Angular Measurements. Foot Ankle Int 23: 68-74.
3. Coughlin MJ, Freund E (2001) The Reliability of Angular Measurements in Hallux Valgus Deformities. Foot Ankle Int 22: 369-379.
4. Resch S, Ryd L, Stenström A, Johnsson K, Reynisson K, et al. (1995) Measuring Hallux Valgus: A Comparison of Conventional Radiography and Clinical Parameters with Regard to Measurement Accuracy. Foot Ankle Int 16: 267-270.
5. Richardson EG, Graves SC, McClure JT, Boone RT (1993) First Metatarsal Head-Shaft Angle: A Method of Determination. Foot Ankle 14(4): 181-185.
6. Seo J-H, Ahn J-Y, Boedijono D (2016) Point-Connecting Measurements of the Hallux Valgus Deformity: A New Measurement and Its Clinical Application. Yonsei Med J 57(3): 741-747.
7. Schneider W, Knahr K (1998) Metatarsophalangeal and intermetatarsal angle: different values and interpretation of postoperative results dependent on the technique of measurement. Foot & ankle international 19(8): 532-536.
8. Van Der Woude P, Keizer SB, Wever-Korevaar M, Thomassen BJW (2019) Intra- and Interobserver Agreement in Hallux Valgus Angle Measurements on Weightbearing and Non-Weightbearing Radiographs. J Foot Ankle Surg 58(4): 706-712.
9. Vittetoe DA, Saltzman CL, Krieg JC, Brown TD (1994) Validity and Reliability of the First Distal Metatarsal Articular Angle. Foot Ankle Int 15(10): 541-547.
10. Takeda R, Mizuhara H, Uchio A, Iidaka T, Makabe K, Kasai T, et al. (2024) Automatic estimation of hallux valgus angle using deep neural network with axis-based annotation. Skeletal Radiol 53(11): 2357-2366.
11. Miller JW (1974) Distal first metatarsal displacement osteotomy. Its place in the schema of bunion surgery. J Bone Joint Surg Am 56(5): 923-931.
12. Bock P, Lanz U, Kröner A, Grabmeier G, Engel A, et al. (2010) The Scarf Osteotomy: A Salvage Procedure for Recurrent Hallux Valgus in Selected Cases. Clin Orthop 468(8): 2177-87.
13. Shima H, Okuda R, Yasuda T, Jotoku T,Kitano N, et al. (2009) Radiographic Measurements in Patients with Hallux Valgus Before and After Proximal Crescentic Osteotomy: J Bone Jt Surg Am 91(6): 1369-1376.
14. Schneider W, Csepan R, Knahr K (2003) Reproducibility of the radiographic metatarsophalangeal angle in hallux surgery. J Bone Joint Surg Am 85(3): 494-499.
15. de Carvalho KAM, Walt JS, Ehret A, Tazegul TE, Dibbern K, et al. (2022) Comparison between Weightbearing-CT semiautomatic and manual measurements in Hallux Valgus. Foot Ankle Surg 28(4): 518-525.
16. Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 39(2): 175-191.
17. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, et al. (2003) Segment Anything.
18. Suvorov R, Logacheva E, Mashikhin A, Remizova A, Ashukha A, et al. (2021) Resolution-robust Large Mask Inpainting with Fourier Convolutions.
19. enesmsahin. simple-lama-inpainting n.d.
20. R Core Team (2024) R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria.

21. Wickham H (2016) NY: Springer Science+Business Media, LLC; Ggplot2. New York, USA.

22. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1(8476): 307-310.

23. Gamer M, Lemon J, Singh IFP (2019) irr Various Coefficients of Interrater Reliability and Agreement 2019.

24. Won Elm DE, Altman PD, Egger PM, J Pocock PS, Gotzsche PC, et al. (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 370(9596): 1453-1457.