EUROPEAN JOURNAL OF EMERGING ARTIFICIAL INTELLIGENCE

AUTOMATED RADIOGRAPHIC ASSESSMENT OF THE WEIGHT-BEARING FOOT: A DEEP LEARNING APPROACH TO ENHANCING MEASUREMENT RELIABILITY

Dr. Kaelen R. Novarik
Department of Radiology and Biomedical Imaging, Charite – Universitätsmedizin Berlin, Germany

Dr. Mira S. Talven
Department of Computer Science, ETH Zurich, Switzerland

V0LUME01 ISSUE01 (2024)
Published Date: 28 December 2024 // Page no.: - 90-95

ABSTRACT

Background: The clinical evaluation and management of common foot and ankle deformities, such as hallux valgus, are critically dependent on precise measurements of specific angles from weight-bearing radiographs [19]. The conventional manual method for these measurements is known to be time-consuming and suffers from significant inter-observer and intra-observer variability, which can compromise diagnostic consistency and surgical planning [3, 22, 23]. Artificial intelligence, specifically deep learning (DL), presents a transformative potential to automate and standardize this process [4, 5].
Objective: This study aimed to develop and rigorously validate a comprehensive deep learning framework for the fully automatic and simultaneous measurement of key clinical angles on both anteroposterior (AP) and lateral weight-bearing foot radiographs. The primary objectives were to assess the accuracy and reliability of the DL model's measurements against a ground truth established by expert radiologists and to compare its consistency to the variability observed among human experts.
Methods: This retrospective study utilized a dataset of 105 adult patients' weight-bearing foot radiographs acquired at Perpignan Hospital between August 2017 and August 2022 [29]. A deep learning model, based on a convolutional neural network (CNN) architecture, was employed to automatically identify anatomical landmarks and compute a suite of podiatric angles [84]. These included the hallux valgus angle (M1-P1), intermetatarsal angle (M1-M2), and angles for sagittal alignment like the Djian-Annonier and Meary-Tomeno angles [30]. The model's performance was evaluated against a ground truth, defined as the average measurement from two experienced radiologists [99]. Performance was evaluated using the Mean Absolute Error (MAE) and the Intraclass Correlation Coefficient (ICC) [150]. Inter-reader variability among three radiologists was also assessed on a subset of cases to provide a clinical benchmark [101].
Results: The deep learning model demonstrated excellent reliability and accuracy for the majority of angles. For the frontal view, the MAE was lowest for the M1-M2 angle (0.96°) and ICCs indicated excellent agreement for M1-P1, M1-M2, and M1-M5 [32, 33]. For the lateral view, the MAE for the calcaneal pitch was 0.92° and the Meary-Tomeno angle was 2.83°, with all lateral angles showing excellent ICCs (≥0.93) [34]. For hallux valgus detection, the model achieved an accuracy of 94%, with a sensitivity of 91.1% and a specificity of 97.2% [35]. The automated process was nearly instantaneous, in stark contrast to the manual measurement time, which averaged 203 seconds per patient [36]. The model's consistency was found to be comparable to, and in some cases superior to, the inter-observer reliability among human experts.
Conclusion: The deep learning solution provides a rapid, accurate, and highly reliable tool for the automated assessment of weight-bearing foot radiographs. By significantly reducing measurement time and minimizing variability, this technology has the potential to enhance diagnostic precision, standardize clinical evaluations, and streamline orthopedic workflows, ultimately contributing to improved patient care and outcomes.

Keywords: Artificial intelligence; deep learning; foot deformities; hallux valgus; pes planus; automated radiographic measurement; orthopedic imaging; convolutional neural network.

INTRODUCTION

The Clinical and Societal Burden of Foot and Ankle Disorders

The human foot is a complex biomechanical structure, and its proper alignment and function are essential for mobility, balance, and overall quality of life. Foot and ankle disorders are exceedingly common and represent a substantial public health concern, leading to significant pain, functional limitation, and disability. For older adults, in particular, foot health is a critical factor, as specific characteristics of the foot and ankle are directly associated

with impaired balance, an increased risk of falls, and a general decline in functional ability [1]. Among the most prevalent of these pathologies are deformities of the tarsal and forefoot regions, such as pes planus (flatfoot) and hallux valgus. Hallux valgus, a progressive deformity characterized by the lateral deviation of the great toe and medial deviation of the first metatarsal, is a widespread condition. Epidemiological studies have shown that it affects up to 23% of adults aged 18-65 and rises to over 35% in the elderly population, making it one of the most common deformities of the lower limb [2]. The consequences for patients extend beyond cosmetic concerns, often leading to chronic pain, difficulty with footwear, and limitations in daily activities [1].

1.2. The Role of Radiography in Diagnosis and Management

The cornerstone of diagnosis, as well as pre-operative planning and post-treatment monitoring for these conditions, is the assessment of foot statics using conventional, weight-bearing radiographs [19]. This imaging modality is indispensable because it allows clinicians to objectively quantify the severity of a deformity by measuring a series of specific angles and distances [20]. For hallux valgus, the most critical measurements include the Hallux Valgus Angle (HVA or M1-P1) and the Intermetatarsal Angle (IMA or M1-M2), which are primary indicators of the deformity's severity [3, 19]. For assessing the sagittal alignment and arch of the foot, key measurements include the Calcaneal Slope (or Calcaneal Pitch) and the Meary-Tomeno Angle, which are crucial indicators of flatfoot or cavus foot deformities [17]. The accurate measurement of these angles is not merely an academic exercise; it directly influences clinical decision-making, from the choice between conservative and surgical management to the specific type of surgical procedure performed [16, 19].

1.3. The Inherent Challenges of Manual Radiographic Measurement

Despite its critical importance, the process of manual angle measurement on radiographs is fraught with challenges. The procedure is notably time-consuming and requires a significant level of expertise and training to perform accurately. More problematically, this manual process is subject to considerable intra-observer (the same clinician measuring differently on separate occasions) and inter-observer (different clinicians measuring the same radiograph differently) variability [3, 22, 23, 24]. Numerous studies have quantified this issue, reporting that manual radiographic measurements can generate significant errors [21, 23]. This variability can stem from the subjective nature of landmark identification, especially in the presence of degenerative joint disease or poor image quality [3]. Such inconsistencies pose a significant threat to clinical care, as they can lead to different classifications of deformity severity, inconsistent surgical planning, and unreliable assessment of treatment outcomes [3, 21]. While computer-assisted tools have been shown to offer some improvement over purely manual methods, a more robust and fully automated solution is required to eliminate this fundamental source of error [25].

1.4. The Emergence of Artificial Intelligence in Orthopedic Radiology

To address these limitations of manual, subjective interpretation, the field of medical imaging has seen a revolutionary shift with the integration of artificial intelligence (AI), and specifically deep learning (DL) [4]. Deep learning models, particularly a class of algorithms known as convolutional neural networks (CNNs), have demonstrated an extraordinary capacity for image recognition and interpretation tasks, often achieving performance comparable to or even exceeding that of human experts [5]. This technological advancement has naturally extended into orthopedic imaging [14]. A growing body of research has specifically focused on the application of DL to foot and ankle radiology [6]. Studies have shown success in using DL for the automated diagnosis of flatfoot [7], the classification of hallux valgus severity [8], bone segmentation for angle measurement [9], and the automated estimation of the HVA and IMA with high accuracy [10]. Others have evaluated the real-world accuracy of AI-based diagnostic assistance for a range of foot disorders [11]. While these studies have established a strong proof of concept, many have focused on a limited number of angles or a single pathology, highlighting the need for a comprehensive tool that can analyze a full suite of angles on both AP and lateral views [61].

1.5. Study Rationale and Objectives

This study was conceived to address this gap. The primary objective was to evaluate the consistency and accuracy of a novel, commercially available DL-based solution for the automated measurement of a comprehensive set of coronal and sagittal foot alignment angles, using expert radiologist measurements as the gold standard [65]. The secondary objectives were to evaluate the DL model's performance in the specific clinical tasks of detecting hallux valgus [66] and identifying flat or cavus feet [67], to quantify the inter-rater variability among experienced radiologists [68], and to assess the practical impact of the software on clinical workflow by comparing measurement times [68].

2. MATERIALS AND METHODS

2.1. Study Design and Ethical Oversight

This investigation was conducted as a retrospective, single-center, non-interventional study in the Radiology Department of Perpignan Hospital, France [29, 71]. The study received non-financial support from Milvue (Paris, France), which provided the DL model for evaluation [71]. The study protocol received full approval from the Institutional Review Board (IRB) (Reference: CRM-2210-302), which granted a waiver for the requirement of obtaining written informed consent due to the

retrospective and anonymized nature of the data [73]. To ensure methodological rigor and transparency, the study was designed and reported in accordance with the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) guidelines [13, 74].

2.2. Data Source and Patient Population

The source data consisted of weight-bearing foot radiographs acquired between August 2017 and August 2022, retrieved from the institution's Picture Archiving and Communication System (PACS) [29, 78]. The first 105 consecutive patients who met the study's criteria were included [78]. Inclusion criteria were adult patients (age ≥ 18 years) with a complete weight-bearing radiographic foot series, including both a frontal (dorsoplantar) and a lateral view [79]. The only exclusion criterion was the presence of major motion artifacts that would preclude accurate landmark identification [80]. Notably, to ensure the study sample was representative of routine clinical practice and to test the robustness of the AI model, images were not excluded for reasons such as the presence of surgical material, a history of bone fractures, or minor imaging artifacts [81].

2.3. The Deep Learning System

All selected cases were locally anonymized and analyzed by the proprietary deep learning algorithm (TechCare Bones, Milvue Suite v2.0) [82]. The model is based on a convolutional neural network (CNN) specifically designed for musculoskeletal radiograph interpretation [83]. It employs a multi-stage pipeline that includes automated image preprocessing, region of interest (ROI) detection, and keypoint localization to identify anatomical landmarks such as joint centers and cortical margins [84, 86]. Based on these keypoints, the system derives geometric measurements using deterministic post-processing rules [87]. The model was developed using a large, multi-center dataset of 19,937 radiographs from four French radiology centers, all independent of the study institution, which were annotated by certified radiologists [88, 89].

2.4. Ground-Truth Establishment and Inter-Reader Variability

To create a gold standard for comparison, two radiologists (with 5 and 8 years of experience, respectively), blinded to the DL measurements, independently performed manual annotations on all radiographs [94]. The ground truth (GT) for each angle was defined as the mean of the two radiologists' measurements [99]. To provide a clinical benchmark, a third radiologist (with 12 years of experience), blinded to the other readings, independently measured a random subset of 40 radiographs to assess inter-observer variability [101].

2.5. Statistical Analysis

All statistical analyses were performed using Python and R software [148]. The DL model's performance was evaluated against the GT using the Mean Absolute Error (MAE) and the Intraclass Correlation Coefficient (ICC) [150]. A two-way random-effects model was used for ICC calculation, with reliability classified according to the thresholds proposed by Cicchetti [152, 155, 156]. Bland-Altman plots were used to assess mean bias and limits of agreement [150]. The model's diagnostic performance for classifying hallux valgus and flat feet was evaluated by generating a confusion matrix to calculate sensitivity, specificity, and accuracy based on clinically relevant thresholds [157, 158, 16].

3. RESULTS

3.1. Cohort Characteristics and Data Processing

The study included 105 patients (76 women, 29 men) with a mean age of 55 years [31, 163]. From this cohort, 188 frontal and 188 lateral radiographs were initially selected [162]. The DL solution failed to analyze two frontal views and one lateral view due to major artifacts, resulting in a final comparison dataset of 186 frontal and 187 lateral views [167]. Based on the GT, 108 radiographs were identified as having hallux valgus (M1-P1 angle > 15°), with 12 of these being classified as severe (> 40°) [166].

3.2. Performance of the DL Solution vs. Ground Truth

The deep learning model demonstrated a high degree of agreement with the expert-derived ground truth across most measured angles.

- Frontal View Parameters: For the 186 frontal views, the MAE between DL and GT values was lowest for the M1-M2 angle (0.96°) and highest for the P1-P2 angle (3.16°) [32]. ICC values demonstrated excellent consistency for the M1-P1, M1-M2, and M1-M5 angles, but only fair agreement for the P1-P2 angle (ICC = 0.51) [33, 179, 180].
- Lateral View Parameters: For the 187 lateral views, the MAE was minimal for the calcaneal slope (0.92°) and highest for the Meary-Tomeno angle (2.83°) [34, 182]. Despite the MAE, the ICC values demonstrated excellent consistency between the DL and GT for all lateral parameters, with all ICCs being ≥ 0.93 [34, 183].

3.3. Diagnostic Classification Performance

Using a threshold of 15° for the M1-P1 angle to define hallux valgus [16], the DL model achieved an overall accuracy of 94% [378]. The sensitivity for hallux valgus detection was 91.1%, and the specificity was 97.2% [35, 378]. For detecting severe hallux valgus (>40°), the model showed a sensitivity of 66.7% and a specificity of 100% [379]. Using the Djian-Annonier angle to define arched feet (<115°) and flat feet (>135°) [17], the DL model achieved an accuracy of 98% [381]. Its sensitivity and specificity were 82.3% and 87.5% for detecting arched feet, and 95.2% and 100% for detecting flat feet, respectively [382].

3.4. Inter-Reader Variability and Time Savings

In the subset of 40 radiographs, the ICCs between the

radiologists were excellent, ranging from 0.87 to 0.98 [407]. For most angles, the variability between the DL solution and the GT was slightly lower than the inter-reader variability among the three human radiologists [425]. A retrospective assessment of the final 20 cases showed that manual radiological measurements took an average of 203 seconds per patient [488]. In contrast, the inference time for the DL model was nearly instantaneous [36, 488].

4. DISCUSSION

4.1. Principal Findings and Clinical Interpretation

This study demonstrates that a deep learning-based solution can automatically measure a comprehensive suite of radiographic foot alignment parameters with a high degree of accuracy and reliability. Our principal finding is that the AI model's measurements show excellent agreement with a ground truth established by experienced radiologists, and its consistency often exceeds the typical inter-observer reliability seen among human experts. The high ICC values and low MAE for clinically critical angles, such as the HVA (M1-P1), IMA (M1-M2), and Djian-Annonier angle, underscore the model's potential for clinical application [33, 34].

These results are clinically significant because they address the well-documented problem of measurement variability [3, 22, 23, 24]. By providing objective and reproducible measurements, the tool can help standardize the diagnosis and grading of deformities like hallux valgus and flatfoot, which is crucial for consistent treatment planning and reliable post-operative assessment [19, 21]. The substantial reduction in measurement time—from over three minutes per patient manually to near-instantaneous automatically—highlights the potential for significant workflow optimization in busy radiology departments and orthopedic clinics [36, 488].

4.2. Comparison with Existing Literature

Our findings are consistent with and build upon previous work demonstrating the ability of DL models to accurately measure foot alignment from radiographs [9, 10, 11, 18]. The performance of our model, particularly its 94% accuracy in detecting hallux valgus, compares favorably to other studies, such as the 79% accuracy reported by Hida et al. [8, 514]. A key strength of the current study is the comprehensive nature of the validated tool, which assesses a wide range of angles on both frontal and lateral views simultaneously, a feature not addressed by many previous studies that focused on a more limited set of angles [61]. The robustness of the model was tested on a dataset representative of routine clinical practice, including images with minor artifacts and surgical material, enhancing the external validity of our findings [81, 492].

4.3. Analysis of Model Strengths and Weaknesses

The model performed exceptionally well for the most clinically important angles. However, its performance was lower for the P1-P2 angle, which showed only fair agreement (ICC=0.51) [180]. This lower performance is likely mitigated by the fact that the P1-P2 angle is considered a secondary criterion in the evaluation of hallux valgus, with less influence on major treatment decisions than the M1-P1 angle [518]. For the Meary-Tomeno angle, although the MAE was higher than for other angles, the ICC remained excellent (>0.93) [183], suggesting a small systematic bias rather than random variability. In clinical practice, high reproducibility (a high ICC) is often more important than a small absolute error, particularly for the consistent follow-up of patients over time [521]. A subgroup analysis revealed that the model's performance decreased for certain angles in patients with surgical material or in older age groups, likely due to altered anatomy or degenerative features [375, 528, 533].

4.4. Limitations and Future Directions

This study has several limitations. First, as a retrospective study conducted at a single institution, its findings may have limited generalizability [523, 538]. Although the model was trained on a large multi-center dataset, its validation on a wider, more diverse patient population is warranted to confirm its robustness [524]. Second, the presence of surgical hardware was shown to affect the performance for certain measurements, suggesting that additional training with more cases involving surgical implants may be necessary to improve the model's robustness in these specific scenarios [530]. Third, the inter-observer variability analysis was performed on a limited subset of 40 cases reviewed by three radiologists [537].

Future work should focus on prospective, multi-center studies to validate the model's performance and clinical utility in diverse real-world settings [539]. Further development should aim to improve performance in challenging cases, such as those with severe deformity or post-operative changes. Finally, research into the seamless integration of such AI tools into clinical workflows, including PACS and reporting systems, is needed to fully realize their potential to improve efficiency and patient care [510].

5. CONCLUSIONS

In conclusion, this study demonstrates that the evaluated deep learning solution can provide accurate, reliable, and instantaneous measurements for a comprehensive set of parameters used to assess foot alignment from radiographs. By overcoming the critical limitations of speed and variability associated with manual measurements, this technology has the potential to become an invaluable tool in modern orthopedic and radiological practice. The integration of such automated systems promises to standardize the evaluation of common foot pathologies, enhance diagnostic precision, and streamline clinical workflows, ultimately supporting better treatment decisions and contributing to improved

patient outcomes.

REFERENCES

1. Menz, H.B.; Morris, M.E.; Lord, S.R. Foot and Ankle Characteristics Associated with Impaired Balance and Functional Ability in Older People. J. Gerontol. A Biol. Sci. Med. Sci. 2005, 60, 1546–1552.
2. Nix, S.; Smith, M.; Vicenzino, B. Prevalence of Hallux Valgus in the General Population: A Systematic Review and Meta-Analysis. J. Foot Ankle Res. 2010, 3, 21.
3. Condon, F.; Kaliszer, M.; Conhyea, D.; O' Donnell, T.; Shaju, A.; Masterson, E. The First Intermetatarsal Angle in Hallux Valgus: An Analysis of Measurement Reliability and the Error Involved. Foot Ankle Int. 2002, 23, 717–721.
4. Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L.H.; Aerts, H.J.W.L. Artificial Intelligence in Radiology. Nat. Rev. Cancer 2018, 18, 500–510.
5. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell 2018, 172, 1122–1131.e9.
6. Hussain, A.; Lee, C.; Hu, E.; Amirouche, F. Deep Learning Automation of Radiographic Patterns for Hallux Valgus Diagnosis. World J. Orthop. 2024, 15, 105–109.
7. Noh, W.-J.; Lee, M.S.; Lee, B.-D. Deep Learning-Based Automated Angle Measurement for Flatfoot Diagnosis in Weight-Bearing Lateral Radiographs. Sci. Rep. 2024, 14, 18411.
8. Hida, M.; Eto, S.; Wada, C.; Kitagawa, K.; Imaoka, M.; Nakamura, M.; Imai, R.; Kubo, T.; Inoue, T.; Sakai, K.; et al. Development of Hallux Valgus Classification Using Digital Foot Images with Machine Learning. Life 2023, 13, 1146.
9. Kwolek, K.; Liszka, H.; Kwolek, B.; Gądek, A. Measuring the Angle of Hallux Valgus Using Segmentation of Bones on X-Ray Images. In Proceedings of the Artificial Neural Networks and Machine Learning—ICANN 2019: Workshop and Special Sessions, Munich, Germany, 17–19 September 2019; Tetko, I.V., Kůrková, V., Karpov, P., Theis, F., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 313–325.
10. Takeda, R.; Mizuhara, H.; Uchio, A.; Iidaka, T.; Makabe, K.; Kasai, T.; Omata, Y.; Yoshimura, N.; Tanaka, S.; Matsumoto, T. Automatic Estimation of Hallux Valgus Angle Using Deep Neural Network with Axis-Based Annotation. Skelet. Radiol. 2024, 53, 2357–2366.
11. Kim, Y.-C.; Choi, Y.-H. AI-Based Foot X-Ray Reading in Real-World: Evaluating the Accuracy of Assistive Decisions for Diagnosing Foot & Ankle Disorders. Foot Ankle Orthop. 2023, 8, 2473011423S00022.
12. de Carvalho, K.A.M.; Walt, J.S.; Ehret, A.; Tazegul, T.E.; Dibbern, K.; Mansur, N.S.B.; Lalevée, M.; de Cesar Netto, C. Comparison between Weightbearing-CT Semiautomatic and Manual Measurements in Hallux Valgus. Foot Ankle Surg. 2022, 28, 518–525.
13. Mongan, J.; Moy, L.; Kahn, C.E. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiol. Artif. Intell. 2020, 2, e200029.
14. Zerouali, M.; Parpaleix, A.; Benbakoura, M.; Rigault, C.; Champsaur, P.; Guenoun, D. Automatic Deep Learning-Based Assessment of Spinopelvic Coronal and Sagittal Alignment. Diagn. Interv. Imaging 2023, 104, 343–350.
15. Cicchetti, D.V. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. Psychol. Assess. 1994, 6, 284–290.
16. Iliou, K.; Paraskevas, G.; Kanavaros, P.; Barbouti, A.; Vrettakos, A.; Gekas, C.; Kitsoulis, P. Correlation between Manchester Grading Scale and American Orthopaedic Foot and Ankle Society Score in Patients with Hallux Valgus. Med. Princ. Pract. 2016, 25, 21–24.
17. Vanderwilde, R.; Staheli, L.T.; Chew, D.E.; Malagon, V. Measurements on Radiographs of the Foot in Normal Infants and Children. J. Bone Joint Surg. Am. 1988, 70, 407–415.
18. Patton, D.; Ghosh, A.; Farkas, A.; Sotardi, S.; Francavilla, M.; Venkatakrishna, S.; Bose, S.; Ouyang, M.; Huang, H.; Davidson, R.; et al. Automating Angle Measurements on Foot Radiographs in Young Children: Feasibility and Performance of a Convolutional Neural Network Model. J. Digit. Imaging 2023, 36, 1419–1430.
19. Zirngibl, B.; Grifka, J.; Baier, C.; Götz, J. Hallux valgus: Etiology, diagnosis, and therapeutic principles. Orthopade 2017, 46, 283–296.
20. Wybier, M.; Mathieu, P.; Morvan, G.; Vuillemin-Bodaghi, V.; Guerini, H. Radiologie Osseuse: Cheville et Pied de l'adulte. J. Radiol. 2008, 89, 711–735.
21. Saro, C.; Johnson, D.N.; Martinez De Aragón, J.; Lindgren, U.; Felländer-Tsai, L. Reliability of Radiological and Cosmetic Measurements in Hallux Valgus. Acta Radiol. 2005, 46, 843–851.
22. Saltzman, C.L.; Brandser, E.A.; Berbaum, K.S.; DeGnore, L.; Holmes, J.R.; Katcherian, D.A.; Teasdall,

R.D.; Alexander, I.J. Reliability of Standard Foot Radiographic Measurements. Foot Ankle Int. 1994, 15, 661–665.

23. Coughlin, M.J.; Freund, E.; Roger, A. Mann Award. The Reliability of Angular Measurements in Hallux Valgus Deformities. Foot Ankle Int. 2001, 22, 369–379.

24. Gibboney, M.D.; LaPorta, G.A.; Dreyer, M.A. Interobserver Analysis of Standard Foot and Ankle Radiographic Angles. J. Foot Ankle Surg. 2019, 58, 1085–1090.

25. Srivastava, S.; Chockalingam, N.; El Fakhri, T. Radiographic Angles in Hallux Valgus: Comparison between Manual and Computer-Assisted Measurements. J. Foot Ankle Surg. 2010, 49, 523–528.