Estimation of the Radiographic Parameters for Hallux Valgus from Photography of the Feet Using a Deep Convolutional Neural Network

Kana Inoue

Department of Medical Engineering, Faculty of Engineering, Chiba University

Satoshi Maki (✉ satoshimaki@gmail.com)

Department of Orthopaedic Surgery, Graduate School of Medicine, Chiba University

Satoshi Yamaguchi

College of Liberal Arts and Sciences, Chiba University, Chiba

Seiji Kimura

Department of Orthopaedic Surgery, Graduate School of Medicine, Chiba University

Ryuichiro Akagi

Department of Orthopaedic Surgery, Graduate School of Medicine, Chiba University

Takahisa Sasho

Graduate School of Global and Transdisciplinary Studies, Chiba University, Chiba

Seiji Ohtori

Department of Orthopaedic Surgery, Graduate School of Medicine, Chiba University

Sumihisa Orita

Center for Frontier Medical Engineering, Chiba University

Research Article

Keywords: Foot, hallux valgus, deep learning, artificial intelligence, photograph

Posted Date: January 27th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1258050/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Hallux valgus (HV) is one of the most common forefoot deformities. Early diagnosis and proper evaluation of HV are important for timely management of HV. The purpose of our study was to estimate the radiographic parameters of HV using deep learning to assess the agreement of the predicted measurement with the actual radiographic measurement. There were 131 patients enrolled in this study. A total of 248 radiographs and 337 photographs of the feet were acquired. Radiographic parameters including the HV angle (HVA), M1-M2 angle, and M1-M5 angle were measured. We constructed a convolutional neural network using Xception and made it a regression model. Then, we fine-tuned the model using images of the feet and the radiographic parameters. The coefficient of determination (R^2) and root mean squared error (RMSE) were calculated to evaluate the performance of the model. The radiographic parameters, including the HVA, M1-M2 angle, and M1-M5 angle were predicted with an $R^2=0.684$, RMSE=7.91; $R^2=0.573$, RMSE=3.29; $R^2=0.381$, RMSE=5.80, respectively. The present study demonstrated that our model was able to predict the radiographic parameters of HV from photography. This study shows a potential application of deep learning for HV screening.

Introduction

Hallux valgus (HV), also known as bunion deformity, is one of the most common forefoot deformities and is characterized by a medial deviation of the first metatarsal bone, a lateral deviation of the hallux, and a prominent metatarsal head ¹. The deformity is associated with disabling pain leading to significant morbidity and quality of life issues. The prevalence of HV increases with age, affecting 23% of adults aged 18-65 and 36% of adults over 65. HV is more prevalent in women, and women are diagnosed 2 to 15 times more often than men ^{2,3}.

Early diagnosis and proper evaluation of HV are important because timely management can improve symptoms and quality of life. The most commonly used reference standard measurements for assessing HV are the HV angle (HVA) and the inter-metatarsal angle on foot radiographs. However, radiographs are not the screening tool due to availability, cost, and radiation exposure. Instead of radiographs, more accessible and less invasive assessments have been reported as alternatives, including visual categorical grading scales to assess the extent of deformity in HV ⁴⁻⁶, and a quantitative measurement of the hallux valgus angle using digital photographs ⁷. Although it is feasible to determine the severity of HV to some extent by the appearance of the foot, the limitation is that this requires training and experience for the examiner, and is not automated.

Deep learning is a type of artificial intelligence and has been applied extensively in the medical imaging field in recent years. The application of convolutional neural networks (CNNs) to medical imaging include classification, object detection, and semantic segmentation ⁸⁻¹⁰. Here, we propose a deep learning estimation for the radiographic measurement of HV based on a regression network where the input to the algorithm is digital photographs of the forefoot and the radiographic measurement of HV is computed as output directly. The purpose of our study was to estimate the radiographic parameters of HV using deep

learning, to classify the severity by grade, and to assess the agreement of the predicted measurement with the actual radiographic measurement.

Results

Baseline patient characteristics

Table 1 shows the baseline characteristics and radiographic data of the participants.

Model performance

The radiographic parameters of hallux valgus, which included the HVA, M1-M2 angle, and M1-M5 angle were predicted with a coefficient of determination (R^2) 0.684, root mean squared error (RMSE) 7.91; R^2 0.573, RMSE 3.29; R^2 0.381, RMSE 5.80, respectively. Agreement between the predicted and ground truth radiographic parameters was substantial for the HVA, but was only fair for the M1-M2 and M1-M5 angles. Scatter plots of the predicted radiographic parameters versus the ground truth are shown in Figs. 1a-c. Confusion matrices of the CNN model to predict the severity grade of HV are indicated in Table 2. The Cohen's weighted κ coefficient for agreement between predicted and actual severity grade of HV was 0.809, indicating substantial agreement.

Discussion

The present study demonstrates that a CNN can predict the radiographic parameters of HV from photographs. Moreover, the agreement between the predicted HV grade by the CNN and the actual grade was substantial. This study indicates a potential application of a CNN for the screening of HV.

The three radiographic parameters of HV, namely the HVA, M1-M2 angle, and M1-M5 angle were estimated from photography using the CNN. The predicted HVA showed substantial agreement with the ground truth, and the predicted M1-M2, and M1-M5 angles showed fair agreement with the ground truth. There have been several studies that have attempted to predict the HV angle from a photograph of the foot. Nix et al. measured the hallux valgus angle quantitatively using digital photographs and showed that the measurement was reliable and highly correlated with the radiographic measurement (Pearson's $r = 0.96$)¹¹. However, in that study, the photographs were taken by an experienced rater, and under strict parameters, to ensure that the position and the angle of the camera matched perfectly to those from the radiography. Yamaguchi et al. demonstrated the photographic HVA, which is the angle between the two lines drawn on the medial side of the hallux and the medial side of the foot on a photograph was reproducible and useful as a non-radiographic measurement to quantify HV deformity⁷. Moreover, the photographic HVA showed a strong correlation with the HVA measured on the radiograph ($R^2 = 0.891$, $P < .001$)⁷. Although the agreement between the HVA predicted by the CNN and the HVA measured on the radiograph is not as good as the agreement found by the previous two studies, the strength of the present study is that the prediction is automatic and does not require expertise or experience for either the

acquisition of the photograph or the rater. Theoretically, this could enable remote screening for HV, without necessitating an initial office visit.

The HV grade based on the HVA predicted by the CNN was substantially associated with the HV grade determined by radiographic measurement (Cohen's weighted κ coefficient = 0.809). There have been a few reports of grading hallux valgus based on the appearance of the foot using photographs⁴⁻⁶. Menz et al. reported that the Manchester scale, consisting of 4 standardized photographs of the feet, was highly correlated with the HVA (Spearman's $\rho=0.73$, $P<0.01$), and moderately associated with the M1-M2 angle (Spearman's $\rho=0.49$, $P<0.01$) obtained from radiographs.⁵ The Manchester scale showed high test-retest reliability, and the participants' self-rated Manchester scale was correlated strongly with scores obtained by expert raters.⁶ The agreement between the observations of the participants and the expert raters was substantial (weighted kappa = 0.71 to 0.80).⁶ Another self-report instrument for HV with a five-grade foot appearance, which weighted kappa scores (left and right feet combined), indicated 0.45 for the agreement between the participants and expert raters, 0.53 at 1-2 months and 0.51 at 3-6 months, for participant repeatability, and 0.82 for expert rater repeatability.⁴ The agreement between the predicted HVA and the ground truth HVA in the present study was comparable to that of this previous study. Since the severity of the HV grade is generally evaluated in 10 – 15 degree increments, the HV grade predicted by the CNN is reliable enough to detect the difference in HV grade given that the RMSE of the angle was 7.91. This study indicates the potential application of a CNN for the screening of hallux valgus. Using the CNN, the HV angles and HV grades can be estimated with substantial accuracy, without requiring experience of the examiner. Thus, this study could be the first step towards the application of a CNN using non-radiographic images for screening of hallux valgus in the near future.

Several limitations should be noted in the present study. First, the dataset used in the study included 337 images of the foot, which is relatively small for a deep learning study. A larger number of images for training could contribute a better agreement between the predicted and actual radiographic parameters. Second, we did not strictly match the position of the feet when taking radiographs and photographs. Third, the self-photography may vary across the participants because they just followed the instruction sheet and had no further assistance. However, this inconsistency in foot position or self-photography represents the clinical setting in which the CNN would be used.

In conclusion, the CNN model was able to predict the radiographic parameters of HV from photography with fair to substantial agreement. Also, the agreement between the actual HV grades and the grades predicted by CNN was substantial. Our study demonstrates the potential application of a CNN for the screening of HV.

Methods

Patients

The present study was conducted using secondary data from the previous prospective study. The study was approved by the local institutional review board of the Graduate School of Medicine, Chiba University. The additional requirement for informed consent was waived by the local institutional review board of the Graduate School of Medicine, Chiba University because of the retrospective analysis. All procedures involving human participants were in accordance with the 1964 Declaration of Helsinki and its later amendments. Patients who visited our foot and ankle clinic from February to June in 2016 and underwent weight-bearing dorsal X-rays of the foot were recruited. Patients with acute inflammatory diseases such as cellulitis and gout, or with a history of ankle surgery, fractures, or dislocation within the past year, were excluded. There were a total of 131 patients enrolled in this study.

Photography of the Feet

Patients took photographs and also received radiographs during their outpatient visits. Patients photographed their feet using a digital camera or smartphone according to an instruction sheet, which was given to each subject to standardize the foot position for the photograph. The camera or smartphone type was not specified. Participants who did not have a camera or smartphone were provided with a digital camera (IXY 150, Canon, Ota Ward, Tokyo). Participants did not receive additional instructions or assistance from the research staff. This was to simulate a situation where a patient uses a smartphone app to take a picture of his or her foot to be diagnosed with HV. The images of the feet were divided into right and left and cropped to a minimum region, which included the ankle to the toes. (Fig.2a, b) The background was removed semi-automatically using PowerPoint (Microsoft Corporation, Redmond, WA). (Fig.2c). In order to correctly identify the big toe and little toe, the photo of the left foot was flipped horizontally so that all foot orientations were recognized as the right foot. A total of 346 images of feet were acquired. One hundred and seventeen patients took images of both feet, and 43 of the 117 took twice. In addition, 14 patients took pictures of one foot, and 3 out of 14 took images twice.

Radiographic dataset

For the weight-bearing, dorsoplantar-view radiographs of the feet, patients were instructed to stand in a relaxed position, distribute the weight evenly on both feet, and keep the feet parallel. The central beam is angled to approximately 15-20 degrees towards the heel at a distance of 100 cm, directly parallel to the long axis of the foot, and centered on the second tarsometatarsal joint ¹². The radiographic parameters of HV including the HVA, M1-M2 angle, and M1-M5 angle were measured. The HVA refers to the angle formed by the axis of the proximal phalanx of the hallux and the axis of the first metatarsal (Fig 3a). The M1-M2 angle is the angle formed by the longitudinal axis of the first and second metatarsals (Fig 3b). The M1-M5 angle is the angle formed by the longitudinal axis of the first and fifth metatarsals (Fig 3c).

Hallux valgus, defined as HVA of $\geq 20^\circ$, was classified as mild [20° , 30°), moderate [30° , 40°) or severe ($>40^\circ$) ¹³.

HV was measured using the angular measurement function of the Picture Archiving and Communication System and rounded to the nearest whole number for analysis. All images were measured by a board-

certified orthopaedic surgeon (SM 14-years of experience).

A total of 248 radiographs were taken, with 117 patients taking radiographs of both feet and 14 people taking radiographs of one foot.

CNN model construction

The CNN architecture was constructed using Python 3.6.7 and Keras 2.2.4 with Tensorflow 2.0.0 at the backend. The models were separately constructed for the HVA, M1-M2 angle, and M1-M5 angle. In this study, we adapted the Xception architectural model, which had been previously trained using images with ImageNet^{14,15}. The input images were resized to 299×299 pixels. We replaced the final layer of the model with a global average pooling layer and a fully connected layer to make the classification model into a regression model¹⁶. Then, we fine-tuned the pre-trained model using the photographs of the feet and the measured radiographic parameters. The first 26 layers were frozen and the weights were not modified during the training process, then the rest of the layers were retrained with our data. The network was trained over 1000 epochs with a learning rate of 0.1, which was reduced if no improvement was seen. Adam was used for the optimizer, and the RMSE was used for the loss function. The data augmentation was done by ImageDataGenerator including a rotation angle range of 90° , a width shift range of 0.1, a height shift range of 0.1, a shear range of 0.1, and a horizontal flip of 50%. The CNN was trained and validated using a computer with a GeForce RTX 2060 graphics processing unit (NVIDIA, Santa Clara, CA).

Performance evaluation of the model

To train and evaluate the CNN model, we performed a five-fold cross-validation. Photographs of the foot were randomly divided into five equal-sized independent subgroups. Images taken from the same patients were assigned to the same subgroup. In each iteration, data from the four subgroups were selected as a training set and the remaining independent subgroups served as validation data. In the validation phase, the performance to predict the radiographic parameters was assessed using the remaining independent subgroup. This cross-validation process was repeated five times. The R^2 and the RMSE were calculated to evaluate the performance of the CNN using the `sklearn.metrics.r2_score` and the square root of `sklearn.metrics.mean_squared_error` function from the Scikit-learn library (version 0.23.2), respectively. The severity of hallux valgus was graded as normal, mild, moderate, and severe¹³ from the degree of the predicted HVA angle, and the agreement with the severity grading based on the radiographic HVA measurement was evaluated using a Cohen's kappa coefficient. The Cohen's kappa coefficient was calculated using `sklearn.metrics.cohen_kappa_score` function from the Scikit-learn library.

Declarations

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to their containing information that could compromise the privacy of research participants but are available from the corresponding author on reasonable request.

Acknowledgements

This work was supported by a research grant funded by the JOA-Subsidized Science Project Research 2020-1 and JSPS KAKENHI Grant Number JP20K18052.

The authors gratefully acknowledge Dr Shuhei Toba for providing the main part of the regression code, Dr Yasukuni Mori for their support in editing the code, and Minako Morimoto for assisting with the acquisition of the datasets.

Authors' contributions

KI and SM wrote the paper. SY collected the patient data. SM designed computational experiments. KI implemented the experiments, prepared the figures and tables, and performed statistical analyses. SM measured the radiographs SY, SK, RA, TS, SO, and SO supervised the work. All authors reviewed the submitted version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. Wülker, N. & Mittag, F. The treatment of hallux valgus. *Dtsch. Arztebl. Int.* **109**, (2012).
2. Nix, S., Smith, M. & Vicenzino, B. Prevalence of hallux valgus in the general population: A systematic review and meta-analysis. *Journal of Foot and Ankle Research* (2010). doi:10.1186/1757-1146-3-21
3. Piqué-Vidal, C., Solé, M. T. & Antich, J. Hallux Valgus Inheritance: Pedigree Research in 350 Patients With Bunion Deformity. *J. Foot Ankle Surg.* (2007). doi:10.1053/j.jfas.2006.10.011
4. Roddy, E., Zhang, W. & Doherty, M. Validation of a self-report instrument for assessment of hallux valgus. *Osteoarthr. Cartil.* (2007). doi:10.1016/j.joca.2007.02.016
5. Menz, H. B. & Munteanu, S. E. Radiographic validation of the Manchester scale for the classification of hallux valgus deformity. *Rheumatology* (2005). doi:10.1093/rheumatology/keh687
6. Menz, H. B., Fotoohabadi, M. R., Wee, E. & Spink, M. J. Validity of self-assessment of hallux valgus using the Manchester scale. *BMC Musculoskelet. Disord.* **11**, (2010).
7. Yamaguchi, S. *et al.* Nonradiographic measurement of hallux valgus angle using self-photography. *J. Orthop. Sports Phys. Ther.* **49**, 80–86 (2019).
8. Galbusera, F., Casaroli, G. & Bassani, T. Artificial intelligence and machine learning in spine research. *Jor Spine* **2**, e1044 (2019).

9. Yang, S. *et al.* Diagnostic accuracy of deep learning in orthopaedic fractures: a systematic review and meta-analysis. *Clin. Radiol.* **75**, (2020).
10. Kalmet, P. H. S. *et al.* Deep learning in fracture detection: a narrative review. *Acta Orthop.* **91**, 215–220 (2020).
11. Nix, S., Russell, T., Vicenzino, B. & Smith, M. Validity and reliability of hallux valgus angle measured on digital photographs. *J. Orthop. Sports Phys. Ther.* **42**, 642–8 (2012).
12. Tanaka, Y. *et al.* Radiographic analysis of hallux valgus in women on weightbearing and nonweightbearing. *Clin. Orthop. Relat. Res.* (1997). doi:10.1097/00003086-199703000-00026
13. Tachibana, S. Etiology, clinical condition, diagnosis (in Japanese). *Hallux Valg. Pract. Guidel.* Tokyo Nankodo 5–19 (2014).
14. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the Inception Architecture for Computer Vision. (2015).
15. Russakovsky, O. *et al.* ImageNet Large Scale Visual Recognition Challenge. (2014).
16. Toba, S. *et al.* Prediction of Pulmonary to Systemic Flow Ratio in Patients with Congenital Heart Disease Using Deep Learning-Based Analysis of Chest Radiographs. *JAMA Cardiol.* 1–9 (2020). doi:10.1001/jamacardio.2019.5620

Tables

Table 1. Baseline characteristics of the patients

Variable	Patients (n=131)
Age, y	61.2 ± 13.3
Sex, n	
Male	34
Female	97
Laterality, n	
Right	123
Left	125
Radiographic parameters, °	
Hallux valgus angle	26.2 ± 16.2
M1-M2 angle	13.9 ± 5.1
M1-M5 angle	32.8 ± 7.0

Values are mean ± SD

Table 2. The confusion matrix showing the relationship between the ground truth and the predicted severity grade of HV. The matrices indicate the share of images correctly deduced in the diagonal line (shaded), and grades to which the misclassified images were assigned are shown in the horizontal plane. The Cohen’s weighted κ coefficient for agreement between predicted and actual severity grade of HV was 0.809, showing a substantial agreement.

		Predicted severity grade of HVA			
Ground truth severity grade of HVA		Normal	Mild	Moderate	Severe
	Normal	132	12	2	0
	Mild	24	25	6	0
	Moderate	7	17	33	16
	Severe	5	2	13	43

HV; Hallux valgus

Figures

Figure 1

A scatter plot of the ground truth vs. predicted HVA (a), M1-M2 angle (b), and M1-M5 angle (c).

HVA, hallux valgus angle

Figure 2

Image preprocessing for the convolutional neural network model training and validation. An original photograph of the feet (**a**). The images of the feet were divided into right and left and cropped to a minimum region, which included the ankle to the toes (**b**). The background was removed semi-automatically using PowerPoint (**c**).

Figure 3

The radiographic parameters of HV including the HVA, the M1-M2 angle, and the M1-M5 angle. The HVA refers to the angle, which is formed by the axis of the proximal phalanx of the hallux and the axis of the first metatarsal (a). The M1-M2 angle is the angle which is formed by the longitudinal axis of the first and second metatarsals (b). The M1-M5 angle is the angle which is formed by the longitudinal axis of the first and fifth metatarsals (c).

HVA, hallux valgus angle