

RESEARCH

Open Access



Contributions of age, gender, body mass index, and normalized arch height to hallux valgus: a decision tree approach

Do-Young Jung¹, Sung-hoon Jung² and Gyeong-tae Gwak^{3*}

Abstract

Background Hallux valgus (HV) is a common toe deformity with various contributory factors. The interactions between intrinsic risk factors of HV, such as arch height, sex, age, and body mass index (BMI) should be considered. The present study aimed to establish a predictive model for HV using intrinsic factors, such as sex, age, BMI, and arch height based on decision tree (DT) model.

Methods This is retrospective study. The study data were based on the fifth Size Korea survey, of the Korea Technology Standard Institute. Among 5,185 patients, 645 were excluded due to unsuitable age or missing data, and 4,540 (males = 2,236 and females = 2,304) were selected for inclusion in the study. Seven variables (i.e., sex, age, BMI, and four normalized arch height variables) were used to develop the prediction model for the presence of HV using a DT model.

Results The DT model correctly classified 68.79% (95% confidence interval [CI] = 67.25–70.29%) of the training data set (3,633 cases). The predicted presence of HV based on the DT was verified against the testing data set (907 cases) and showed an accuracy of 69.57% (95% CI = 66.46–72.55%).

Conclusions The DT model predicted the presence of HV on the basis of sex, age, and normalized arch height. According to our model, women aged over 50 years and those with lower normalized arch height were at high risk of HV.

Keywords Hallux valgus, Normalized arch height, Decision tree

Background

Hallux valgus (HV) is a common toe deformity characterized by excessive angulation, rotation, and lateral deviation of the great toe at the first metatarsophalangeal

(MTP) joint [1]. Symptoms of HV include poor-fitting shoes, plantar foot pain, and pain in the medial first MTP joint [2]. The resultant deformity causes pain and discomfort, which decreases the physical function of the foot and has been identified as a risk factor for falls in older people [3, 4]. Also, Traumatic events if not adequately treated could result in degeneration of the joint [5, 6]. Therefore, HV causes several effects that are not limited to the feet [3, 7, 8]. Understanding the risk factors of HV is important for its prevention and management [9, 10].

Various intrinsic and extrinsic factors contribute to HV, such as age [10, 11], sex [2, 4, 10, 11], body mass index (BMI) [10], foot morphology [12, 13], and shoe

*Correspondence:

Gyeong-tae Gwak
tarutear@yonsei.ac.kr

¹ Department of Physical Therapy, College of Tourism & Health, Joongbu University, Geumsan, Republic of Korea

² Department of Physical Therapy, Division of Health Science, Baekseok University, Cheonan, Republic of Korea

³ Department of Physical Therapy, Graduate School, Yonsei University, Wonju, Republic of Korea



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

characteristics [12, 14]. HV is more prevalent in women and older adults [11], and is associated with increased BMI and pes planus [2, 10, 15, 16]. Zhao et al. suggested that sex, age, and obesity affect the arch structure [17]. Therefore, interactions between intrinsic risk factors of HV, such as arch height, and sex, age, and BMI should be considered.

The decision tree (DT) is one of the most popular classification techniques due to its capability of presenting results in a simple format and modelling nonlinear relationships [18–20]. Previous studies of the contribution of intrinsic factors, such as age, sex, BMI, and pes planus, to HV progression assumed linear relationships between outcome and predictor variables [2, 21, 22]. The present study aimed to establish a predictive model for HV using intrinsic factors, such as sex, age, BMI, and arch height, based on a DT model.

Methods

Study participants and data

The data of this study were based on the fifth Size Korea survey, which were publicly accessible data, of the Korea Technology Standard Institute (<https://sizekorea.kr/human-info/meas-report?measDegree=5>). Size Korea was a part of the fifth National Anthropometric Survey, and the fifth survey was conducted among 14,200 Korean civilians aged 0–90 years between 2003 and 2004. Foot measurements were performed at ages 10–69 years among 5,185 participants. Among the 5,185 participants, 645 were excluded due to ineligible age or missing data, and a total of 4,540 participants (males = 2,236 and females = 2,304) were included for analysis. The Institutional Review Board (IRB) of Joongbu University approved this retrospective study and waived the requirement for written informed consent (No. JIRB-2022051101–01).

Foot scan measurements

During the fifth Size Korea survey, foot scans were performed using the INFOOT 3D Digitizer model IF-21Series (I-Ware Laboratory Co., Ltd, Osaka, Japan) (Fig. 1). ISO 15535 (general requirements for establishing anthropometric databases, 2003) was used to ensure the validity of measurement techniques and data reliability. The 13 non-reflective green velvet markers were placed on specific anatomical landmarks of the foot. Participants were asked to step with the right foot in a resting stance position on to the glass footplate inside the laser scanner, while the left foot was placed next to the scanner on a step of the same height to distribute their body weight equally over both feet. Scans were performed using a scan pitch of 1.0 mm and an optical laser scanning procedure to measure anthropometric data of the



Fig. 1 The device used for the measurement of foot scan

foot. Participants were instructed not to move their feet during scanning.

Data processing

BMI was calculated as body weight divided by height squared (kg/m^2). The five variables related to foot scan measurements were defined as follows (Fig. 2): First, HV angle (HVA) was calculated as the angle between two lines (one line connecting the medial side of the first MTP joint to the medial side of the heel and the other line connecting the medial side of the first MTP joint to the medial side of the hallux). Participants were assigned to HV ($\text{HVA} > 10^\circ$) and non-HV ($\text{HVA} \leq 10^\circ$) groups [23, 24]. Second, the total foot length (TFL) was calculated as the distance between the most posterior aspect of the heel and the tip of the longest toe measured along the foot axis. Third, instep length (IL) was calculated as the perpendicular distance between the first MTP joint to the most posterior aspect of the heel. Fourth, instep

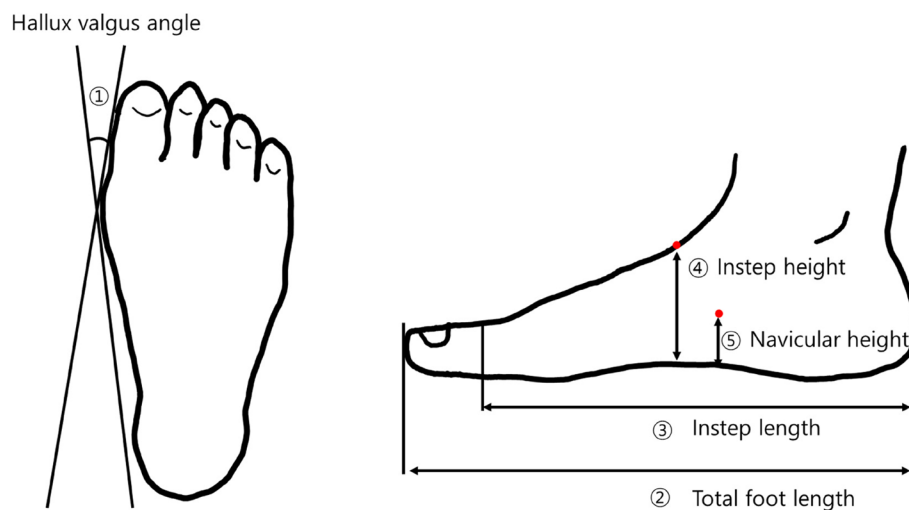


Fig. 2 The measurement of foot scan measurement variables

height (IH) was calculated as the distance between the highest point of the instep and the supporting surface (measurement taken at 50% foot length). Fifth, navicular height (NH) was calculated as the distance between the navicular tuberosity and the supporting surface. We calculated four normalized arch height indices, i.e., NH/TFL, NH/IL, IH/TFL, and IH/IL, from four foot scan measurements.

Data analyses

The prediction model for HV was constructed using a classification and regression tree (CART) derived using the `rpart` function of RStudio (version 1.1.463; RStudio, Inc., Boston, MA, USA) based on the R program [25]. Seven variables (sex, age, BMI, NH/IL, IH/IL, NH/TFL, and IH/TFL) were used to develop the prediction model for HV using a CART model. The prediction model used gini impurity to select splits during classification. The subjects from each group were randomly divided into training (80% of subjects) and test (remaining 20%) data sets using ‘`createDataPartition`’ function of the `caret` package (<http://topepo.github.io/caret/index.html>). A total of 3,633 cases were included in the training data set (1,282 cases with HV and 2,351 cases without HV), and 907 cases were included in the test data set (320 cases with HV and 587 cases without HV). The no information rates (classified by chance) were 64.71% in the training data set and 64.72% in the test data set due to imbalanced class.

Results

Figure 3 shows the DT for HV, which had five terminal nodes. The DT showed that sex was the first predictor of HV, with 75.72% (1350/1783) of the male subjects

included in the control group. In the subgroup of subjects who were not males (i.e., female subjects), age was the second predictor. Subjects whose age was not in 10–40 (i.e., aged 50–60) were more likely to have HV (73.67%; 221/330). In the subgroup of subjects who were aged 10–40, NH/IL was the third predictor. The cutoff value of NH/IL for absence of HV was ≥ 0.21 . In the subgroup of subjects with $\text{NH/IL} < 0.21$, IH/IL was the fourth predictor. The cutoff value of IH/IL for absence of HV was 0.32. Details of the five divisions with cutoff values of the respective predictors and number of cases classified in each subgroup are presented in Fig. 3.

The DT model correctly classified 68.79% (95% confidence interval [CI]=67.25–70.29) of the training data set (3,633 cases). The predicted presence of HV based on the DT was verified against the test data set (907 cases) and an accuracy of 69.57% (95% CI=66.46–72.55%) was obtained (Table 1).

Discussion

Previous studies have suggested that three-dimensional (3D) foot scanning could be used for HV diagnosis [23, 26, 27]. Zhou et al. (2013) reported high correlation between HVAs measured on 3D foot scan and radiography ($r=0.70$), and that the mean HVA measured by 3D foot scan was 4.9° smaller than that measured by radiography [23]. HV is diagnosed at $\text{HVA} \geq 15^\circ$ on radiography; therefore, in the present study, participants with $\text{HVA} \geq 10^\circ$ on the 3D foot scan were assigned to the HV group.

We included seven variables (sex, age, BMI, NH/IL, IH/IL, NH/TFL, and IH/TFL) in the prediction model for HV. The DT with classification and regression tree selected four variables (sex, age, NH/IL, and IH/IL) from

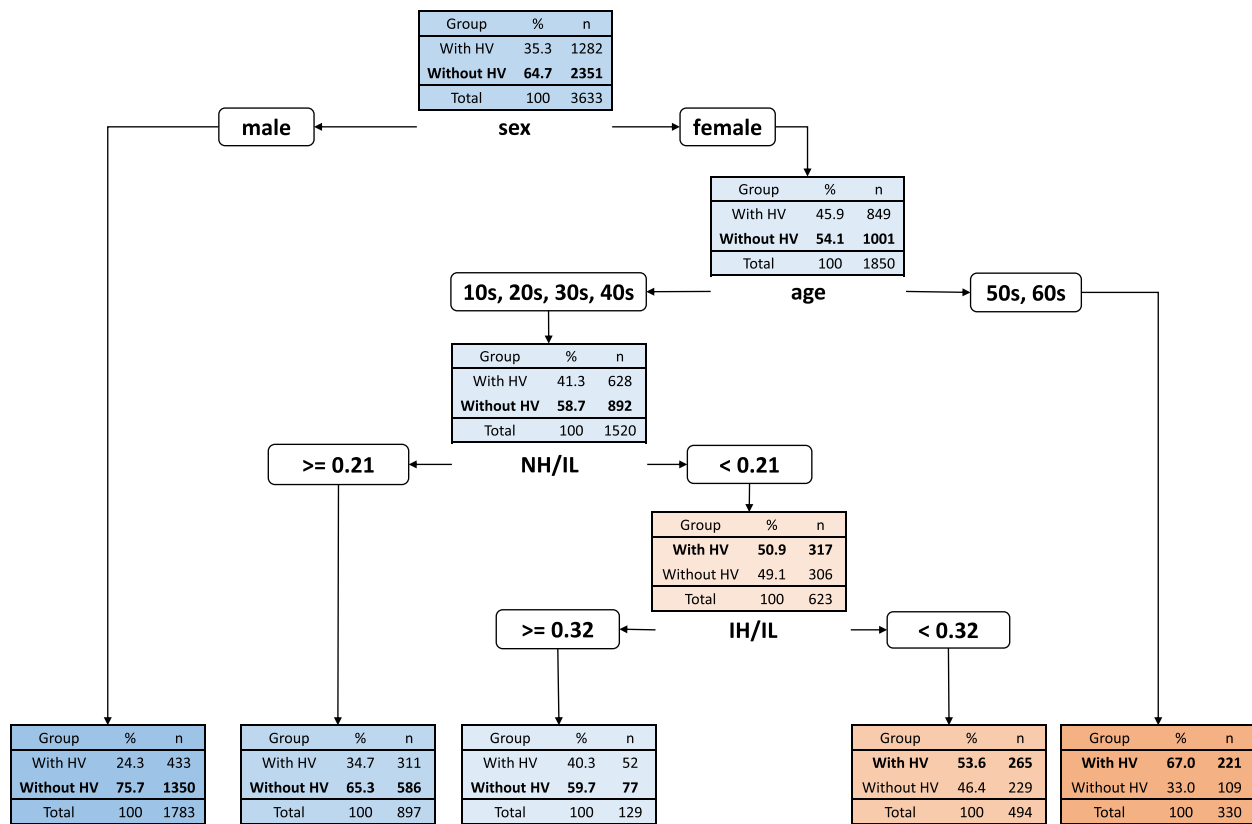


Fig. 3 The decision tree model predicting the presence of hallux valgus

Table 1 Performance of the classification tree as calculated during the model training and testing

Calculated evaluation measures		Predicted values		
		Hallux Valgus	Control	
Training data set	Hallux Valgus	796	338	Sensitivity: 37.91%
	Control	486	2013	Specificity: 85.62%
	Overall accuracy (95% CI)			68.79% (67.25, 70.29) %
Test data set	Hallux Valgus	113	207	Sensitivity: 35.31%
	Control	69	518	Specificity: 88.25%
	Overall accuracy (95% CI)			69.57% (66.46, 72.55) %

seven variables to predict the presence of HV. Our prediction model for HV selected sex and age as the first and second predictors of HV, respectively, which is consistent with previous studies. Previous studies showed that HV is more prevalent in women and older adults [4, 22]. A meta-analysis of 76 surveys (496,957 participants) reported that the pooled estimate of HV prevalence in females (30%) was 2.3-fold higher than that in males (13%) [11]. In addition, a previous study showed an increase in HV prevalence with age: 7.8% in juveniles (16 studies, $n=73,030$), 23% in adults aged 18–65 years (15 studies, $n=23,790$), and 35.7% in older adults (37 studies,

$n=16,001$) [11]. A previous study showed that participants with HV were more likely to be older and female, and have musculoskeletal co-morbidities [28].

Many studies have suggested that pes planus affects the etiology of HV [12, 29]. Despite the commonly held belief that pes planus plays an important role in HV development, the association between navicular height, a measure of pes planus, and HV is controversial [30, 31]. Bryant et al. reported no significant difference in navicular height between the HV group and controls [30]. In contrast, Komeda et al. reported that each point on the medial longitudinal arch in the HV group was

significantly lower than the corresponding points in the control group [31]. However, NH and IH cannot be used alone to quantify the pes planus. The roughly triangular shape of the arch indicates a relationship between arch length and height [32]. Therefore, in this study, arch height was normalized by foot length (TFL and IL), to improve the validity and reliability of measurements. In a previous study, NH and IH were divided by TFL and IL, respectively [33]. In the present study, arch heights normalized by TFL (NH/TFL and IH/TFL), and IL (NH/IL and IH/IL) were used to develop the prediction model. However, our prediction model for HV only selected arch heights normalized by IL (NH/IL and IH/IL), which may be because measurement of TFL could be skewed by foot deformities, such as HV and claw toes [33]. Therefore, our results suggest that normalized arch height, especially arch height divided by IL, should be considered when evaluating subjects with HV.

Our prediction model selected NH/IL and IH/IL as the third and fourth predictors of HV with cutoff values of 0.21 and 0.32, respectively. The cutoff values in our model were similar to the results of previous studies [34, 35]. Aboelnasr et al. reported that the optimal cutoff value for diagnosing pes planus was $NH/IL \leq 0.195$ [34]. In addition, Hillstrom et al. reported lower IH/IL values in asymptomatic healthy adults with pes planus (0.33 ± 0.03) compared to other foot types, such as pes rectus (0.36 ± 0.03) and pes cavus (0.38 ± 0.03) [35]. Because low normalized arch height is associated with pes planus, our results showed that pes planus would affect the etiology of HV. Considering the node position, our model showed that pes planus could affect the presence of HV in women (aged 10–40 years) but not in men. These results are consistent with the findings of Dufour et al., who reported that pes planus was associated with increased odds of having HV and forefoot pain in women, but pes planus was not associated with HV in men [36]. Similarly, a previous study reported no difference in the odds of having HV based on the presence of pes planus in men, but women with pes planus showed higher odds of having HV than women without pes planus [13]. Therefore, further studies are needed to investigate the risk factors for HV in men.

Previous studies have attempted to predict the risk of lower extremity musculoskeletal disorders [37–39]. Kernozek et al. showed acceptable performances of HV predictive models based on logistic regression of clinical (accuracy of 91.5%) and biomechanical (accuracy of 93.3%) variables [39]. However, the study had a relatively small sample size (40 participants with HV and 51 healthy controls) and did not perform cross-validation. In addition, both models in the study showed high specificity and relatively low sensitivity. Similarly,

our model showed high specificity (85.62% and 88.25% in the training and test data sets, respectively) and low sensitivity (37.91% and 35.31% in the training and test data sets, respectively). There are possible reasons for these results. First, the numbers of participants in the HV and non-HV groups were not equal. Therefore, class imbalance could affect the performance of the prediction model. Second, extrinsic and intrinsic factors could contribute to HV development [29]. However, we only used the intrinsic factors, such as sex, age, and pes planus, to predict HV. Therefore, it is necessary to consider the extrinsic factors, such as footwear, occupation, and excess weight bearing, to improve the performance of predictive models for HV. In this study, a predictive model was developed using large number of foot scan data. Our prediction model is easy to interpret, which is one of the strengths of decision tree model [20]. The cutoff value could also be used clinically for evaluation and intervention.

Conclusions

Our DT model for predicting the presence of HV selected sex, age, and normalized arch height as predictors of HV. According to our model, women aged over 50 years or with low normalized arch height are at high risk for HV. Our model could be used to identify high-risk patients for HV and to recommend conservative management based on the suggested cutoff values.

Acknowledgements

Not applicable

Authors' contributions

Gyeong-tae Gwak and Do-young Jung wrote the main manuscript text and Sung-hoon Jung prepared figures and tables. All authors reviewed the manuscript. The author(s) read and approved the final manuscript.

Funding

This paper was supported by Joongbu University Research & Development Fund, in 2022 and Baekseok University.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the sizekorea repository, [<https://sizekorea.kr/human-info/meas-report?measDegree=5>].

Declarations

Ethics approval and consent to participate

This study used data already obtained by sizekorea project under the Bioethics and Safety Act of the Korean government. All procedures were performed in accordance with the principles of the World Medical Association's Declaration of Helsinki. The Institutional Review Board (IRB) of Joongbu University approved this retrospective study and waived the requirement for written informed consent (No. JIRB-2022051101–01).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 4 December 2022 Accepted: 31 March 2023
Published online: 11 April 2023

References

- Deschamps K, Birch I, Desloovere K, Matricali GA. The impact of hallux valgus on foot kinematics: A cross-sectional, comparative study. *Gait Posture*. 2010;32:102–6.
- Hecht PJ, Lin TJ. Hallux valgus. *Med Clin North Am*. 2014;98:227–32.
- Menz HB, Lord SR. Foot problems, functional impairment, and falls in older people. *J Am Podiatr Med Assoc*. 1999;89:458–67.
- Cho NH, Kim S, Kwon DJ, Kim HA. The prevalence of hallux valgus and its association with foot pain and function in a rural Korean community. *J Bone Joint Surg Br*. 2009;91:494–8.
- Caizzi G, Coviello M, Franchini A, Riefoli F, Palmiotto F, Moretti B. Post-traumatic instability of the first metatarsophalangeal joint: a novel surgical technique of capsular reconstruction in a young kickboxer. *Med Glas (Zenica)*. 2023;20.
- Fabeck LG, Zekhnini C, Farrokh D, Descamps P-Y, Delincé PE. Traumatic hallux valgus following rupture of the medial collateral ligament of the first metatarsophalangeal joint: A case report. *J Foot Ankle Surg*. 2002;41:125–8.
- Saro C, Jensen I, Lindgren U, Felländer-Tsai L. Quality-of-life outcome after hallux valgus surgery. *Qual Life Res*. 2007;16:731–8.
- Kim S, You K, Jung D. Between-side Comparisons of Iliotibial Band Flexibility and the Tibial Torsion Angle in Subjects with an Asymmetric Hallux Valgus Angle. *J KEMA*. 2018;2:11–5.
- Deenik AR, De Visser E, Louwerens JWK, Malefijt MDW, Draijer FF, De Bie RA. Hallux valgus angle as main predictor for correction of hallux valgus. *BMC Musculoskelet Disord*. 2008;9:1–6.
- Nishimura A, Kato K, Fukuda A, Nakazora S, Yamada T, Uchida A, et al. Prevalence of hallux valgus and risk factors among Japanese community dwellers. *J Orthop Sci*. 2014;19:257–62.
- 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.
- Nguyen USDT, Hillstrom HJ, Li W, Dufour AB, Kiel DP, Procter-Gray E, et al. Factors associated with hallux valgus in a population-based study of older women and men: the MOBILIZE Boston Study. *Osteoarthritis Cartilage*. 2010;18:41–6.
- Golightly YM, Hannan MT, Dufour AB, Renner JB, Jordan JM. Factors associated with hallux valgus in a community-based cross-sectional study of adults with and without osteoarthritis. *Arthritis Care Res (Hoboken)*. 2015;67:791–8.
- Munteanu SE, Menz HB, Wark JD, Christie JJ, Scurrah KJ, Bui M, et al. Hallux Valgus, By Nature or Nurture? A Twin Study. *Arthritis Care Res (Hoboken)*. 2017;69:1421–8.
- Atbaşı Z, Erdem Y, Kose O, Demiralp B, Ilkbahar S, Tekin HO. Relationship Between Hallux Valgus and Pes Planus: Real or Fiction? *J Foot Ankle Surg*. 2020;59:513–7.
- Kang M-H, Moon D-C, Park D-J. Comparison of Foot Posture, Joint Position Sense, Dynamic Balance, and Plantar Pressure between Young Adults with and without Obesity. *J Musculoskelet Sci Technol*. 2022;6:58–63.
- Zhao X, Gu Y, Yu J, Ma Y, Zhou Z. The Influence of Gender, Age, and Body Mass Index on Arch Height and Arch Stiffness. *J Foot Ankle Surg*. 2020;59:298–302.
- Abu-Nimeh S, Nappa D, Wang X, Nair S. A comparison of machine learning techniques for phishing detection. In: Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit. 2007. p. 60–9.
- Kantardzic M. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons; 2011.
- Mendelek F, Caby I, Pelayo P, Kheir RB. The application of a classification-tree model for predicting low back pain prevalence among hospital staff. *Arch Environ Occup Health*. 2013;68:135–44.
- Kilmartin TE, Wallace WA. The Significance of Pes Planus in Juvenile Hallux Valgus. *Foot Ankle Int*. 1992;13:53–6.
- Roddy E, Zhang W, Doherty M. Prevalence and associations of hallux valgus in a primary care population. *Arthritis Care Res (Hoboken)*. 2008;59:857–62.
- Zhou J, Hlavacek P, Xu B, Chen W. Approach for measuring the angle of hallux valgus. *Indian J Orthop*. 2013;47:278–82.
- Klein C, Kinz W, Zemsch A, Groll-Knapp E, Kundi M. The hallux valgus angle of the margo medialis pedis as an alternative to the measurement of the metatarsophalangeal hallux valgus angle. *BMC Musculoskelet Disord*. 2014;15:1–8.
- Allaire J. RStudio: integrated development environment for R. Boston: MA; 2012. p. 770.
- Zhou J, Hlavacek P, Xu B, Chen W. New approach for measuring the angle of Hallux Valgus. *Advances in Environment, Biotechnology and Biomedicine, WSEAS, Zlin*. 2012;:300–4.
- Kim S. Analysis of 3D Foot Shape Features in Elderly with Hallux Valgus Using Multi-Dimensional Scaling Method. 2012; April:37–43.
- Abhishek A, Roddy E, Zhang W, Doherty M. Are hallux valgus and big toe pain associated with impaired quality of life? A cross-sectional study. *Osteoarthritis Cartilage*. 2010;18:923–6.
- Perera AM, Mason L, Stephens MM. The pathogenesis of hallux valgus. *J Bone Joint Surg Am*. 2011;93:1650–61.
- Bryant A, Tinley P, Singer K. A comparison of radiographic measurements in normal, hallux valgus, and hallux limitus feet. *J Foot Ankle Surg*. 2000;39:39–43.
- Komeda T, Tanaka Y, Takakura Y, Fujii T, Samoto N, Tamai S. Evaluation of the longitudinal arch of the foot with hallux valgus using a newly developed two-dimensional coordinate system. *J Orthop Sci*. 2001;6:110–8.
- Hill M, Naemi R, Branthwaite H, Chockalingam N. The relationship between arch height and foot length: Implications for size grading. *Appl Ergon*. 2017;59:243–50.
- Williams DS, McClay IS. Measurements used to characterize the foot and the medial longitudinal arch: Reliability and validity. *Phys Ther*. 2000;80:864–71.
- Abolnasr EA, Hegazy FA, Zaghoul AA, El-Talawy HA, Abdelazim FH. Validation of normalized truncated navicular height as a clinical assessment measure of static foot posture to determine flatfoot in children and adolescents: A cross sectional study. *Foot*. 2018;37 January:85–90.
- Hillstrom HJ, Song J, Kraszewski AP, Hafer JF, Mootanah R, Dufour AB, et al. Foot type biomechanics part 1: Structure and function of the asymptomatic foot. *Gait Posture*. 2013;37:445–51.
- Dufour AB, Casey VA, Golightly YM, Hannan MT. Characteristics associated with hallux valgus in a population-based foot study of older adults. *Arthritis Care Res (Hoboken)*. 2014;66:1880–6.
- Henriquez M, Sumner J, Faherty M, Sell T, Bent B. Machine Learning to Predict Lower Extremity Musculoskeletal Injury Risk in Student Athletes. *Front Sports Act Living*. 2020;2 November.
- Ferreira VMLM, Oliveira RR, Nazareno TS, Freitas LV, Mendonça LD. Interaction of foot and hip factors identifies Achilles tendinopathy occurrence in recreational runners. *Phys Ther Sport*. 2020;45:111–9.
- Kernozek TW, Elfessi A, Steriker S. Clinical and Biomechanical Risk Hallux Valgus. *J Am Podiatr Med Assoc*. 2003;93:97–103.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.