

AUTOMATED MEASUREMENT OF FOOT DEFORMITIES: FLATFOOT, HIGH ARCH, CALCANEAL FRACTURE

MACIEJ SKWIRCZYŃSKI^{✉,1}, TOMASZ GĄCIARZ², MAREK SKOMOROWSKI³ AND WADIM WOJCIECHOWSKI⁴

¹Pedagogical University of Cracow, Podchorążych 2, 30-084 Cracow, Poland, ²Cracow University of Technology, Warszawska 24, 31-155 Cracow, Poland, ³Jagiellonian University, Gołębia 24, 31-007 Cracow, Poland, ⁴Jagiellonian University Medical College, Kopernika 19, 31-501 Cracow, Poland
e-mail: maciej.skwirczynski@up.krakow.pl, tga@pk.edu.pl, marek.skomorowski@uj.edu.pl, wadim.wojciechowski@uj.edu.pl

(Received September 15, 2018; revised March 15, 2019; accepted June 10, 2019)

ABSTRACT

Radiographic measurements of foot deformities are used to determine, among other things, such conditions as flatfoot, high arch, or calcaneal fracture. Those measurements are achieved by estimating four angles. Manual assessment of those angles is time-consuming not to mention inevitable errors of such approximation. To the best of the authors knowledge, currently there is no research focusing on finding those four angles. In this paper an algorithm for automatic assessment of those angles, based on extremely randomized trees, is being proposed. Moreover this diagnostic assisting system was intended to be as generic as possible and could be applied, to some degree, to other similar problems. To demonstrate usefulness of this method, correlations of automated measurements with manual ones against correlations of manual measurements with manual ones are being compared. The significance level for manual-manual measurements comparison is less than 0.001 in case of all four angles. The significance level for automated-manual measurements comparison is also less than 0.001 in all cases. The results show that the search for the aforementioned angles can be automated. Even with the use of a generic algorithm a high degree of precision can be achieved, allowing for a more efficient diagnosis.

Keywords: calcaneal fracture, computer aided diagnostic, extremely randomized trees, flatfoot, high arch, x-rays.

INTRODUCTION

No matter which country we live in or how technically advanced our society is, medical care is vital and often far from perfect. Some problems are still too advanced to solve due to lack of knowledge, tools, or simply not sufficient funding. Some affect only narrow group of people. In many cases we do not find cure for, so called, rare diseases not because lack of means, but due to small amount of data or interest, where the small interest results in little to no founding. That is why computer aided diagnostic is so important even in case of simple or already thoroughly investigated cases such as flatfoot (pes planovalgus), high arch (pes cavus), or calcaneal fracture. It could save both time and money.

Despite the common belief that x-ray is being replaced by more modern diagnostic techniques such as computed tomography (CT) and magnetic resonance imaging (MRI), it still holds a vital role in the art of medical diagnostic (Arif *et al.*, 2018; Wojciechowski *et al.*, 2016). As is stated in Arslan *et al.* (2014), x-ray radiographs are fast, easily accessible, and expose patients to less radiation than CT. In this paper we present an algorithm that enables

automated angles detection on feet x-ray images. We take focus on four angles: calcaneal inclination angle (CIA), talar-first metatarsal angle (TFMA), talar declination angle (TDA), and Boehler angle (BA). Assessment of the first three allows us to determine if we are dealing with pes planovalgus or pes cavus. Assessment of the Boehler angle helps us to evaluate a deformity resulting from a calcaneal fracture. Manual assessment of those angles is tedious and time-consuming. We could easily think of a better use of time for a highly-trained specialist, such as a radiologist. We see the time that could be gained this way worth the effort, even if we are talking about aiding in this part of the diagnostic process and not the complete automation of all the tasks that a radiologist might find unnecessarily monotonous.

The calcaneal pitch or calcaneal inclination angle is, with details depending on the source (Waldt and Woertler, 2014; Yates, 2009), described as formed by two lines. The first line is tangent to the inferior cortex of the calcaneus. It is defined by two points: the anterior extension of the calcaneal tuberosity on the plantar side and the antero-inferior corner of the calcaneus that articulates with the cuboid (Waldt and

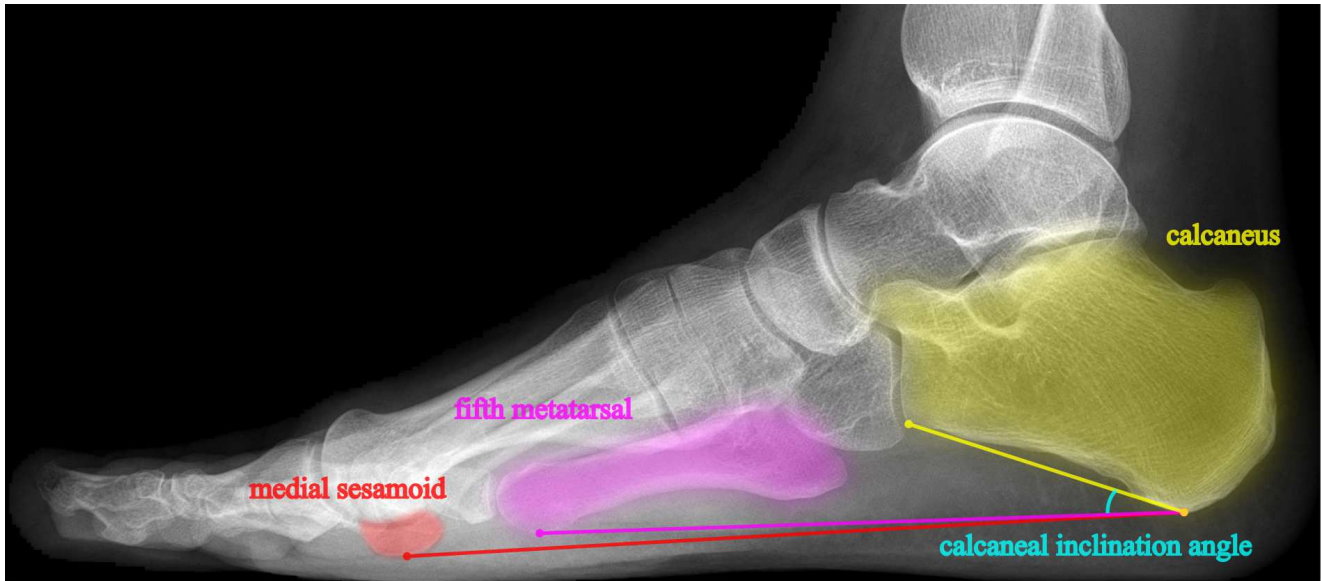


Fig. 1. Lines used to determine the calcaneal inclination angle.

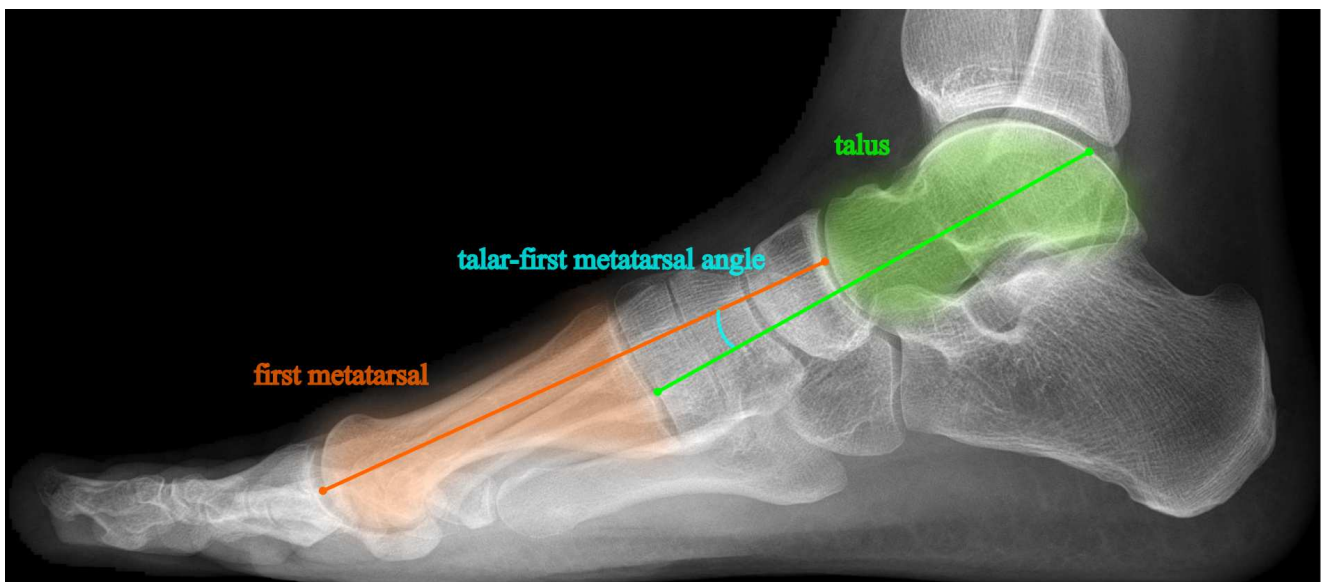


Fig. 2. Lines used to determine the talar-first metatarsal angle.

Woertler, 2014). For the second one we take a line that passes between the lowest point of the calcaneus and the lowest point of the fifth metatarsal. The latter can be used interchangeably with the medial sesamoid bone as shown on Fig. 1.

The Meary's angle or talar-first metatarsal angle is formed by the longitudinal axis of the first metatarsal shaft and the talus as shown on Fig. 2. Measurements of the Meary's angle can also be used to check for the hyperflexibility of the foot, which can be the case when more than an eight degree change is taking place from weight-bearing to nonweight-bearing (Waldt and

Woertler, 2014).

The Boehler angle, also called tuber joint angle, is formed by a line tangent to the superior part of the posterior and a line tangent to the superior part of the anterior of the calcaneus as shown on Fig. 3 (Khoshhal *et al.*, 2005).

The talar declination angle (Fig. 4) is defined by the longitudinal axis of the talus, same as in case of the talar-first metatarsal angle, and a line drawn based on the same principles as the second line in the calcaneal inclination angle case (Waldt and Woertler, 2014).

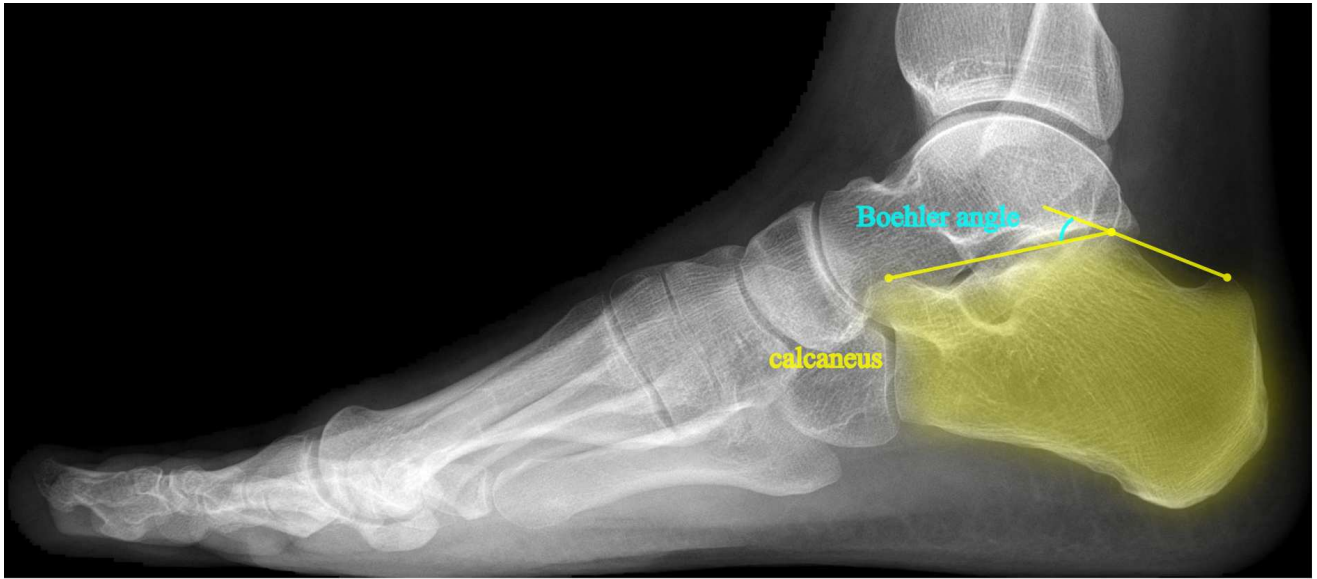


Fig. 3. Lines used to determine the Boehler angle.

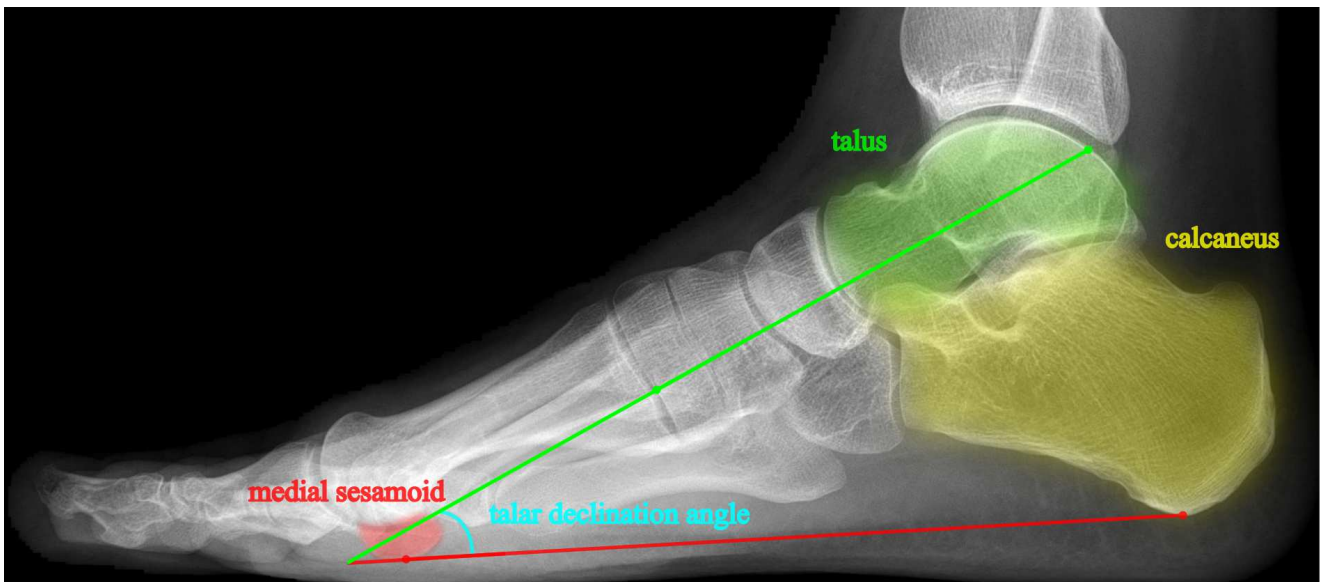


Fig. 4. Lines used to determine the talar declination angle.

Angles intervals, which allow us to evaluate if we deal with normal range, pes planovalgus, or pes cavus (Waldt and Woertler, 2014), are presented in Table 1.

Table 1. Angles intervals.

angle [deg]	normal range	PP	PC
CIA	$20 < \alpha < 30$	$\alpha \leq 20$	$\alpha \geq 30$
TFMA	$-4 < \alpha < 4$	$\alpha \leq -4$	$\alpha \geq 4$
BA	$\alpha \geq 4$	-	-
TDA	$14 < \alpha < 36$	$\alpha \leq 14$	$\alpha \geq 36$

The paper is structured as follows: Materials and

Methods section first describes, without going into too much detail, used methods and then introduces the data and experiments with previously omitted specific settings. The statistical data can be found in Results section. The last section is dedicated to discussion.

MATERIALS AND METHODS

The extremely randomized trees is a method developed by Geurts *et al.* (2006). It is currently widely used in research in the field of computer aided diagnostic

(Soltaninejad *et al.*, 2017; Pinto *et al.*, 2018; Scalzo *et al.*, 2012). The rationale behind the extremely randomized trees in this case is that the variance reduction is more important than the bias reduction (Briscoe and Feldman, 2011). We need a finite number of correct answers candidates to lower time needed for computations and, what is even more important, we want to avoid high number of false, outlying candidates. As far as the bias in extremely randomized trees is concerned, it is minimized by mechanism of using the full learning sample rather than a bootstrap like in a non-extremely random forests case (Geurts *et al.*, 2006). Moreover the second stage of the proposed algorithm is constructed in such a way that allows it to handle, to some degree, errors occurring due to the bias.

The main idea of the proposed approach is to find characteristic points, which allow us to draw lines and with those we can, using dot product, calculate the angles. Despite the individual characters of those points, we use the same method, since we want the proposed algorithm to be as generic as possible. We do not negate that an individual approach for some parts of the problem at hand could be better. A good candidate for this is finding the long axis of the talus for the talar-first metatarsal angle calculation. By reducing the approach to finding key points we achieve acceptable results with one, easy to understand and implement method. More information can be found in the Results and Discussion sections.

The proposed algorithm is divided into two phases, which are described below in detail. During the first one we extract a certain number of candidates for key points. The second stage leaves us with one, precisely selected point for each in question. Once we have those, we can easily calculate angles, which will allow a radiologist to assess the foot deformities.

FIRST STAGE OF THE PROPOSED ALGORITHM

The first stage of the proposed approach is based on the extremely randomized forest algorithm (Pedregosa *et al.*, 2011). Building of an extremely randomized forest and prediction during the first step is performed on down-sampled images and corresponding data with down-sampling factor two along each axis. Such action considerably increases performance while having no impact on the results (Donner *et al.*, 2013). We train the algorithm with a set of images on which a number of correct key points have been selected by one of the authors (W.W., board-certified radiologist). For the purpose of defining the key points, both the central pixel as well as all points in three pixel radius

are considered a valid choice for the algorithm in question. To further define the key points, we encircle them with a one pixel wide ring consisting of pixels which we consider neither the correct answer nor the background. For the purpose of this research, we define the background as set of five hundred randomly chosen pixels while discarding any points designated as a key point or the ring encircling it.

For each key and background point we take one hundred pixels, randomly selected by taking samples from a multivariate normal distribution with standard deviation – sigma set to one hundred, and subtract the grey values of those. By doing so, we create feature sets containing one hundred elements each. Only those are used during the first stage for the purpose of identification of each pixel class. Such a solution allows us to avoid issues related to the image brightness fluctuations (Lepetit and Fua, 2006). The number of classes is the sum of all key points plus one for the background.

We use forty fully expanded extremely randomized trees with at least two samples required to split an internal node and the Gini impurity (Gini) is being utilized to measure the quality of a split. Each of forty trees votes on the class to which the currently inspected point should belong to. This process provides us with a number of candidate sets for key points used in the second step of this algorithm.

SECOND STAGE OF THE PROPOSED ALGORITHM

Since we are using only key points candidates obtained during the first stage, we can work in this part of the algorithm without worrying about the efficiency. Having that in mind, we restore the previous resolution of the images. We base this step on the extremely randomized trees as well. Analogically as before, we use forty fully grown trees with at least two samples required to split and the Gini as quality criterion, but this time what interests us in each point candidate passing through the forests is their probability of being assigned to one of the classes. To estimate those probabilities we take the spatial positions of each key point plus its vicinity and a ring created in the same manner as the ring in the first step. During the construction of the forests the first ones are used as positive and the last ones as negative samples. This means that we take each point extracted during the first stage and we treat it as a valid choice if it is closer to the correct key point and its neighbours within three pixel radius. We record the probability, which was evaluated by the forest. If a currently tested point is closer to the one pixel wide ring, which is encircling correct points area, we discard it.

From all of the predicted points we consider only those, which the probability of being the correct one is higher than the chosen level. Finally, we calculate the mean of their positions, which results with one final point for each class, background excluded, from the first stage. As mentioned at the beginning of this chapter, all that is left to do is to calculate the angles. That is achieved by a calculation based on the dot product.

DATA

Our data consists of forty-eight right foot radiographs. The feet in those images are presented in a standing lateral position with a resolution of 1024×450 pixels. Those radiographs were analysed manually, with significant time periods between each case, three times by one of the authors (W.W.). On each image twelve points were annotated and they coordinates were saved.

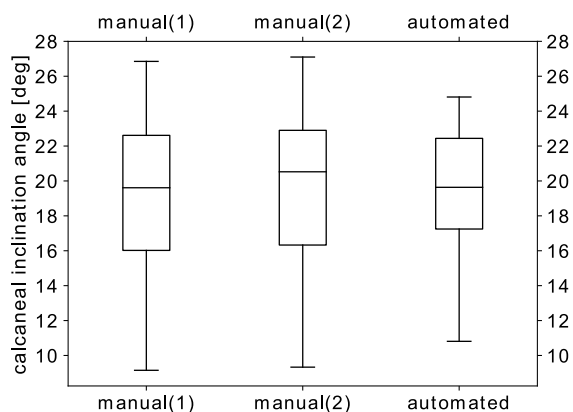


Fig. 5. Box-and-whisker plot for the calcaneal inclination angle.

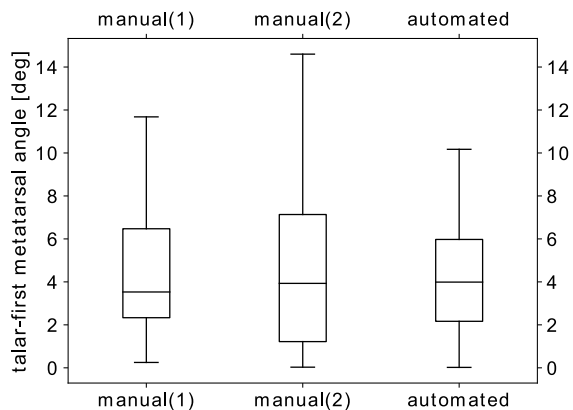


Fig. 6. Box-and-whisker plot for the talar-first metatarsal angle.

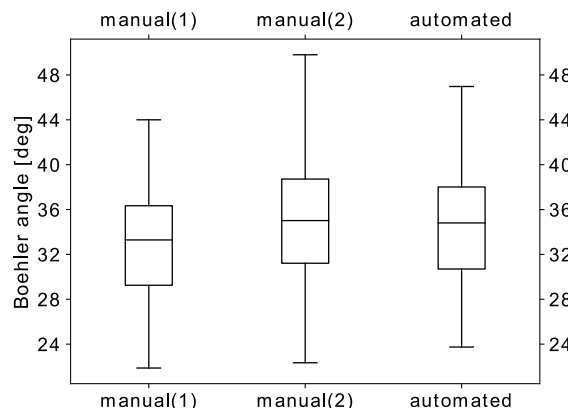


Fig. 7. Box-and-whisker plot for the Boehler angle.

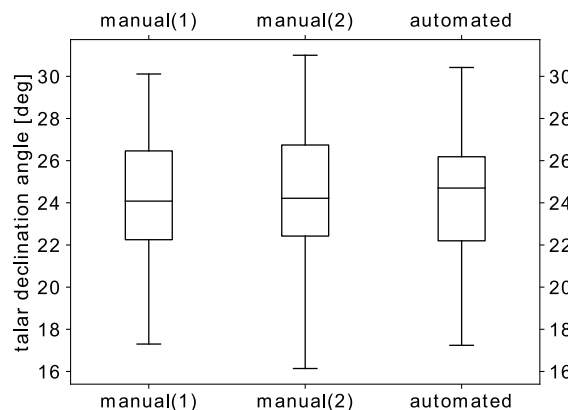


Fig. 8. Box-and-whisker plot for the talar declination angle.

EXPERIMENTS

To fully utilize our radiographs we use a leave-one-out cross-validation. We divide each radiograph in four sub-parts for the purpose of searching for key points candidates in each of them in a different thread.

The random forest classifier was trained using forty extremely randomized trees, which were grown to their full depth. Decreasing the number of those trees decreases the prediction time, but at the cost of precision. The latter is also true when we increase the number of the trees.

We use the Gini for the purpose of the qualitative measurement of a split. We could potentially use the Entropy method, but we have decided in favour of the Gini over Entropy due to the fact that the first one is slightly more efficient while avoiding the use of square root and both of those methods are producing the same results.

Table 2. Comparison with other methods.

First stage		ERF classifier	RF classifier	ERF classifier	RF classifier	CNN
Second stage		ERF classifier	RF classifier	ERF regressor	RF regressor	ERF classifier
Angle						
	CIA	0.80	0.82	0.88	0.65	0.70
	TFMA	0.49	0.49	0.20	0.20	0.43
	BA	0.47	0.33	0.35	0.39	0.38
	TDA	0.48	0.55	0.28	0.52	0.52

After the completion of the First stage of the proposed algorithm, we obtain circa five thousand key point candidates per image. The received data contain pixel coordinates, class membership information, and the probability of this membership. We can narrow predictions by narrowing probability range, but in our case attempts at such manipulations did not provide us with any improvements, but in fact turned out to be to the detriment of the process.

For the purpose of this research, the introduction of one pixel wide rings provides us with the much desired improvement to the Bohlear angle correlation, which has proven to be most problematic in the whole process. Increasing the ring width has nullified any previously achieved improvements. In our case, choosing one pixel width has proven to be beneficial, however this must not be the case for other possible applications of this particular procedure.

Table 3. Average angles differences.

angle[deg]	Man. -auto	Man.(1) -Man.(2)	Man.(1) -Man.(3)	Man.(2) -Man.(3)
CIA	1.86	0.83	0.64	0.86
TFMA	2.07	2.06	1.63	2.23
BA	4.50	4.27	2.40	3.58
TDA	2.58	1.90	2.18	2.18

At the beginning of second stage of the proposed algorithm we restore 1024×450 pixels resolution of the images. Forty extremely randomized trees, fully grown in each forest, are trained with spatial information of pixels appointed in the first stage as valid choice and the one pixel wide ring, which is used to disqualify candidates.

Then we take into account only those predictions, for which the probability of being the correct one is higher than the chosen level, in our case 0.8. We pick such a value due to the fact that it gives us best results when applied for all points. By doing so, we ensure generic character of the proposed algorithm,

but nothing stands in the way of setting a different value for all points or even different values for different points when working on another problem.

Table 4. P-values obtained from Wilcoxon's test.

angle	Man. -auto	Man.(1) -Man.(2)	Man.(1) -Man.(3)	Man.(2) -Man.(3)
CIA	0.52	0.11	0.73	0.08
TFMA	0.72	0.64	0.06	0.20
BA	0.08	0.28	0.43	0.21
TDA	0.93	0.91	0.01	0.01

In order to end the process with a single result for each class, we calculate mean coordinates from points with probability equal or greater than the level mentioned previously.

FINISHING

Since our problem comes down to the evaluation of the angles, we calculate them using a dot product and inverse cosine. We designed a GUI, which allows for predictions and is able to display effects of those prediction. It is also capable of saving and loading them from files.

The algorithm we described earlier as a two step process due to the third stage's status as non-essential, for the purpose of this algorithm all that is required are points and the final stage could be omitted or modified to meet the requirements of any possible tasks it could be used for in the future.

RESULTS

To demonstrate usefulness of the proposed method, we compare correlations, expressed with Spearman's rank-order correlation coefficient (Jones *et al.*, 2011), of automated measurements with manual ones against correlations of manual measurements with manual ones. We assume the 0.001 significance level.

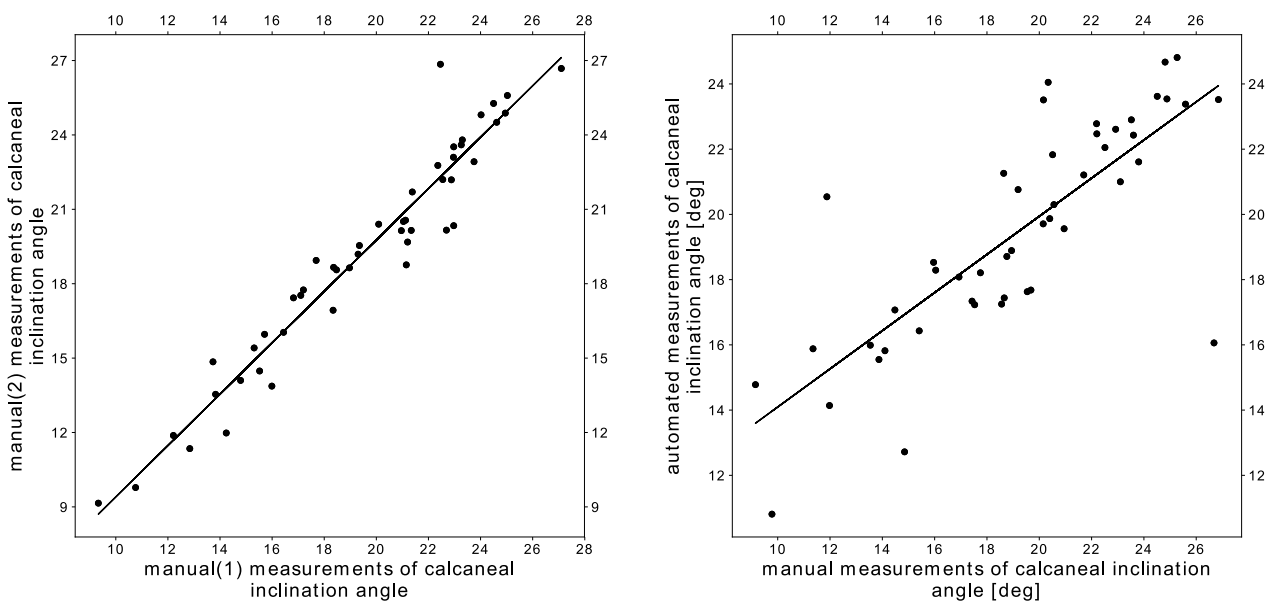


Fig. 9. Correlation charts for the calcaneal inclination angle.

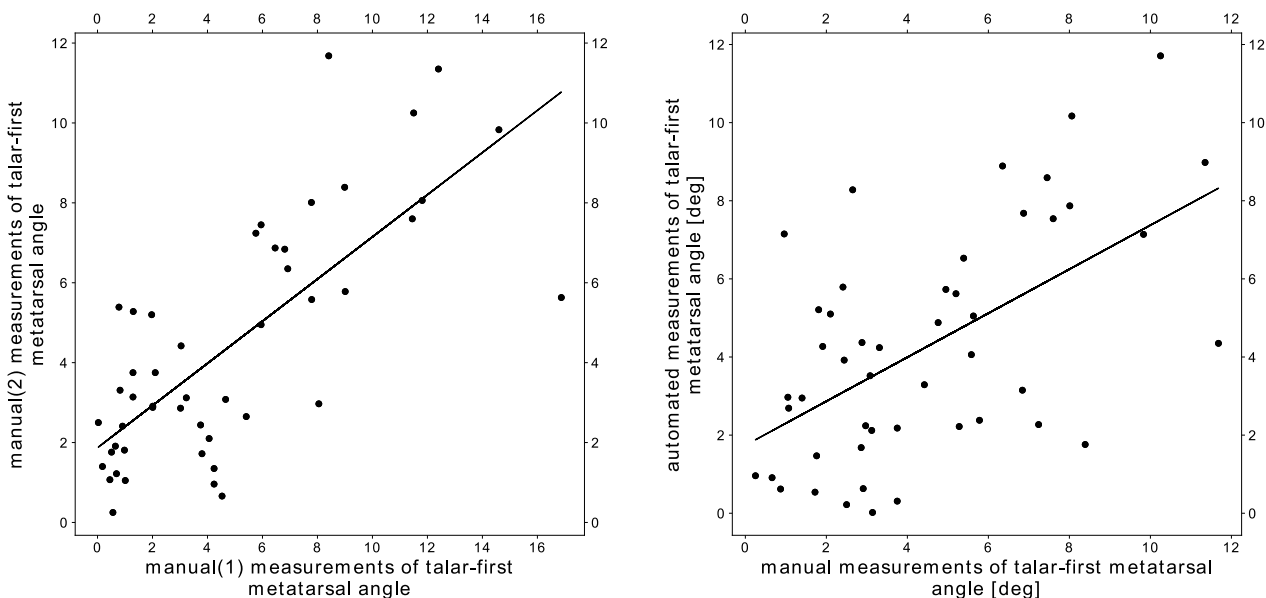


Fig. 10. Correlation charts for the talar-first metatarsal angle.

The proposed algorithm gives the best results in case of the calcaneal inclination angle with 0.80 correlation coefficient. The Bohlear angle appears to be most problematic. In its case correlation obtained by the automated predictions equals 0.47, but in comparison the correlation between manual predictions does not appear to be significantly better and achieves 0.51. Results of the remaining two, the

talar-first metatarsal angle and the talar declination angle, are in-between with the correlation coefficients 0.49 and 0.48 respectively.

We place the Bland-Altman charts, with 95 percent limits, in the same manner as the correlation charts (Figs. 13 to 16). The intervals of agreements of the automated measurements and manual ones are very

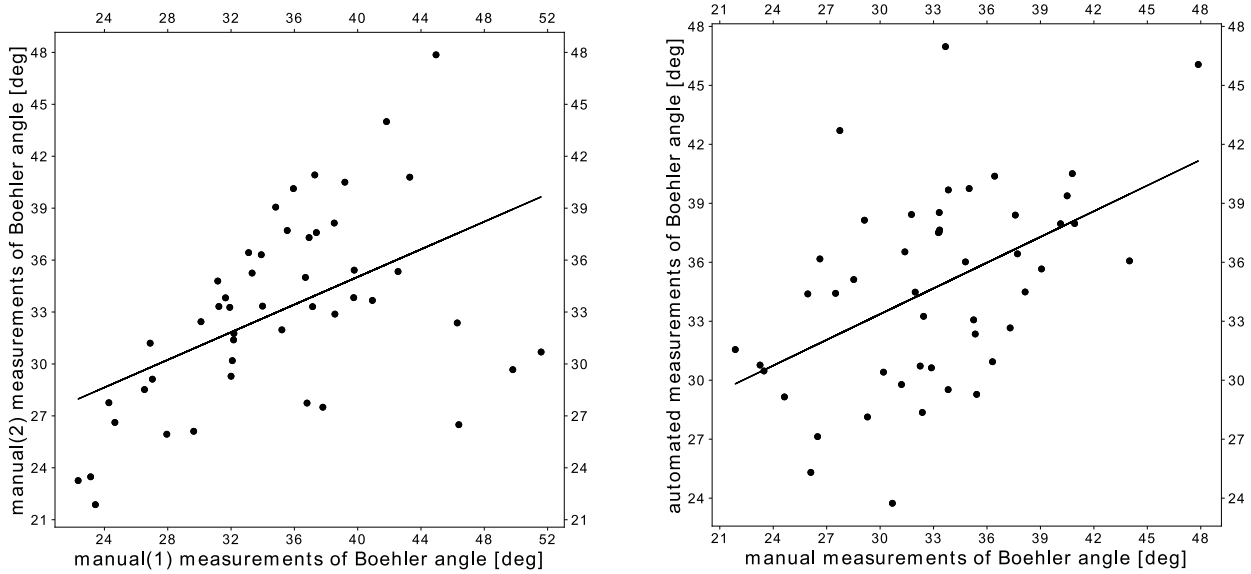


Fig. 11. Correlation charts for the Boehler angle.

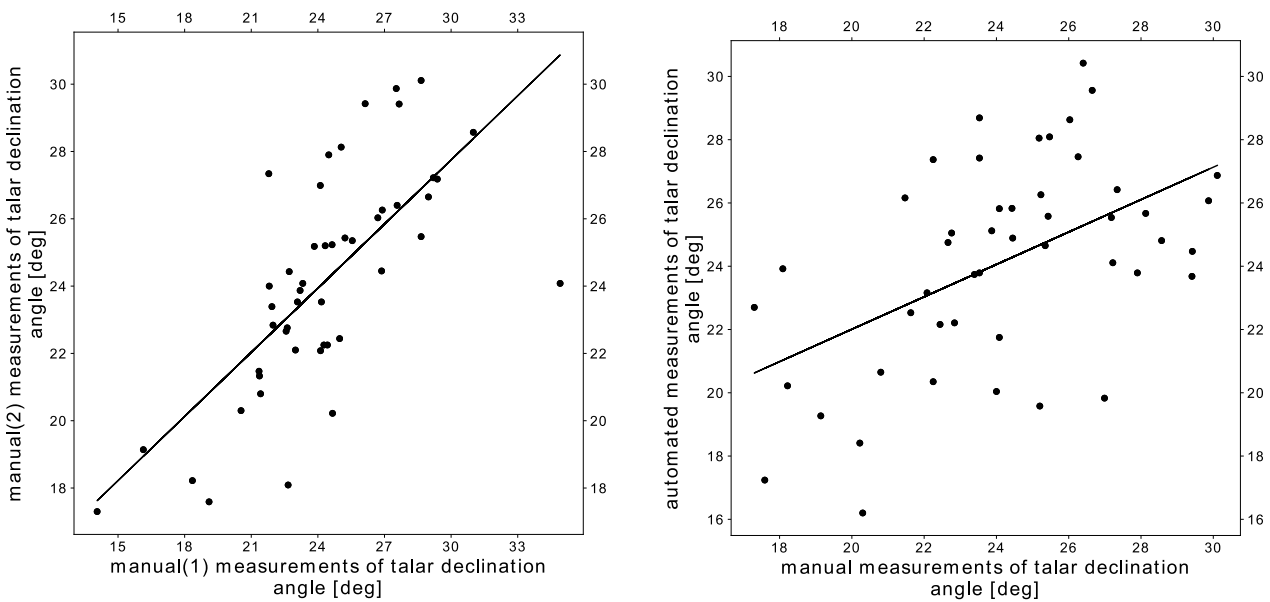


Fig. 12. Correlation charts for the talar declination angle.

similar, as can be seen on those plots, to the intervals of agreements of the manual measurements and manual ones. The points are uniformly scattered between the limits implicating no particular trend. In addition the differences in all cases are small.

In Table 2 we compare Spearman’s rank-order correlation coefficients of the proposed method in

which we use the extremely randomized forest (ERF) classifier during both stages, with those of the methods in which different combinations of the ERF classifier and regressor as well as randomized forest (RF) classifier and regressor are being used. In the last column of Table 2, we present the correlations obtained while using the convolutional neural network (CNN) during the first stage.

Table 5. Landmark localization accuracy and its effect on each angle calculation error (standard measurement uncertainty).

angle	vector 1				vector 2				mean error of angle [deg]
	mean error of								
	point 1		point 2		point 3		point 4		
	x	y	x	y	x	y	x	y	
CIA	2.86	1.88	2.74	1.26	5.60	1.81	2.27	1.03	0.30
TFMA	2.23	1.34	2.27	1.49	1.24	1.72	1.93	2.31	0.46
BA	1.46	2.05	1.56	1.62	2.32	1.79	1.51	2.12	0.51
TDA	5.60	1.81	1.93	2.31	1.24	1.72	1.93	2.31	0.42

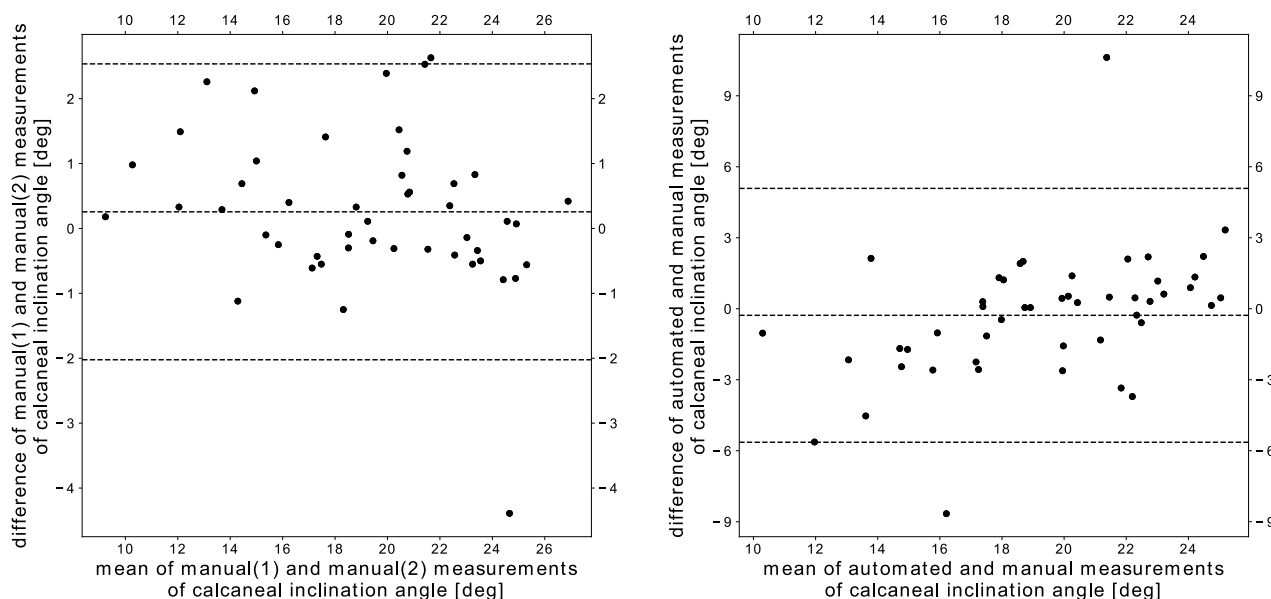


Fig. 13. Bland-Altman charts for the calcaneal inclination angle.

We use scatter, Bland-Altman and box-and-whisker plots to visualize the nature of the relationship between pairs of automatically and manually predicted angles (Altman and Bland, 1983; Hunter, 2007). To ease the comparison the correlation charts are placed so, that the automated measurements with manual ones are aligned with correlations of the manual measurements with manual ones (Figs. 9 to 12).

The box-and-whisker plots are shown in Figs. 5 to 8. In Table 3 we present average angles differences.

We also perform the Wilcoxon’s tests, which do not allow for the rejection of a zero hypothesis, which assumes that the manually calculated angles are equal to the angles determined by the algorithm. This is due to the fact that the p-value is, in case of the manual-automated comparison, greater than the assumed confidence level of 0.05. This is also true for the manual(1)-manual(2) comparison. In case of the

manual(1)-manual(3) and the manual(2)-manual(3) comparisons p-value for the talar declination angle measurements allows us to reject the zero hypothesis. All p-values are presented in Table 4.

Landmark localization accuracy and how it is affecting the error in the angle calculation can be found in Table 5.

DISCUSSION

We found and proved correlations between automated and manual predictions, but there is room for improvement. It is important to use consistent method when determining key points for the learning set. As we have mentioned above, calculation of the calcaneal inclination angle can be done in different ways. Mixing them together in same learning set could lead to serious

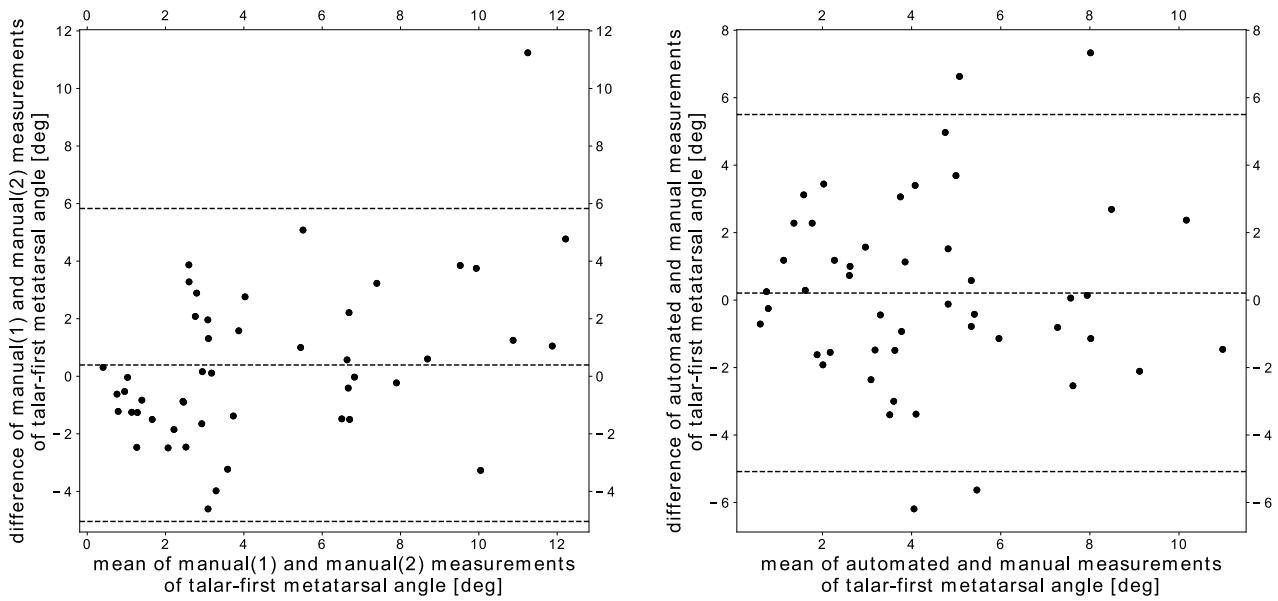


Fig. 14. Bland-Altman charts for the talar-first metatarsal angle.

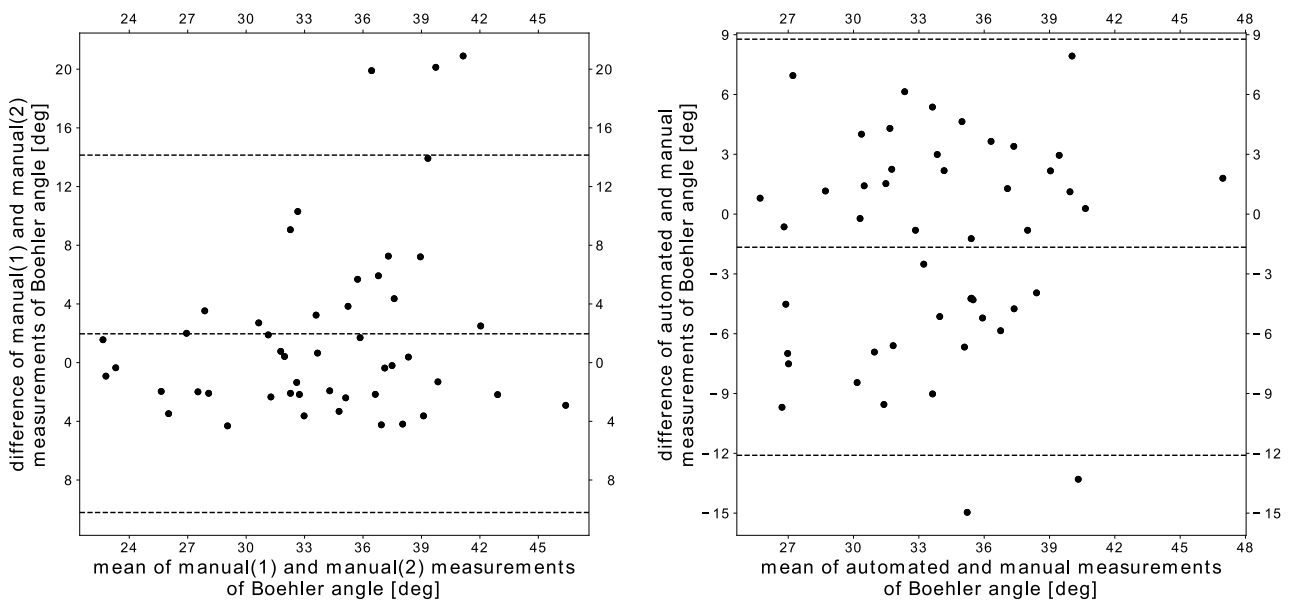


Fig. 15. Bland-Altman charts for the Boehler angle.

discrepancies. In the future research that could be taken into consideration.

In Experiments section we indicate fragments of the algorithm, which can be easily altered to achieve different results, depending on a problem at hand. It is also possible to change attributes defining the points in the first stage or to replace the second stage with a

method allowing for more off-centre placement of the object. Such modification could as well be done for the purpose of solving another problem or to improve aspects of solving this one. It is worth mentioning that the individual approach to each point could give better results, but that was not the intention of this paper. In the future research an individual approach could be implemented and compared with current results.

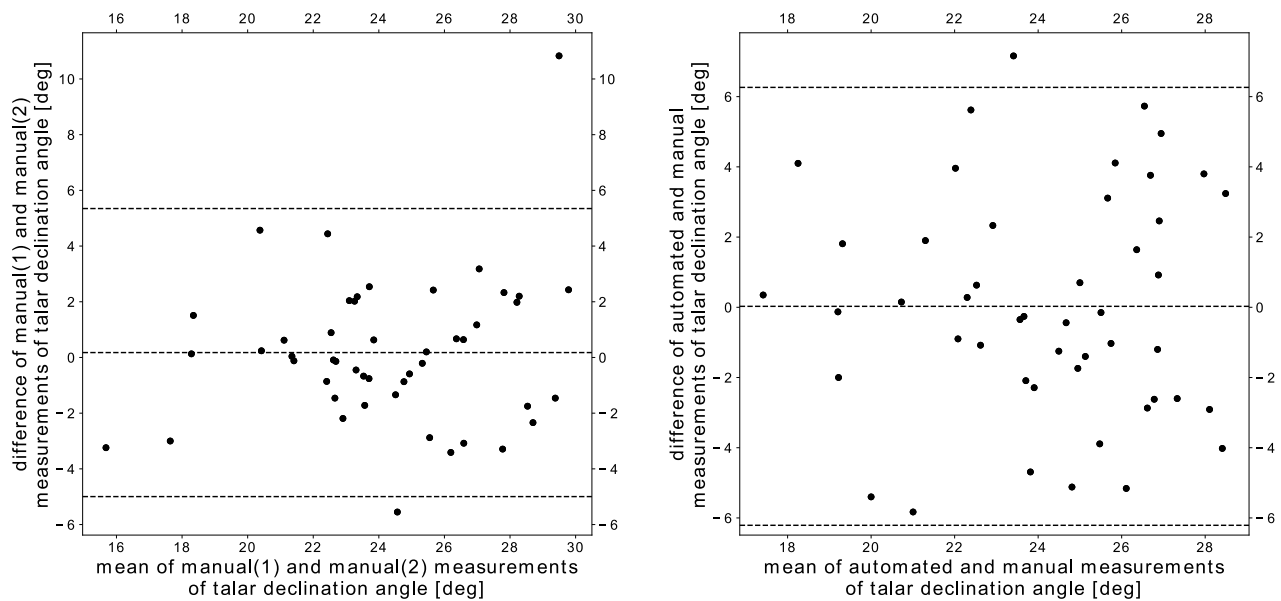


Fig. 16. Bland-Altman charts for the talar declination angle.

Although the interest in computer aided diagnostics is growing rapidly in the recent years, there is comparatively little research in the area of automated standing lateral position foot angles assessment. A recent study by Kao *et al.* (2018) presents algorithm allowing for automated flatfoot detection. The process focuses on measurement of the calcaneal inclination angle only and as such allows for an individual approach. Even taking into account a large number of x-rays used in this research, the proclaimed success rate is 73.33% of the cases. While the remainder required manual correction. An even earlier paper by Yang *et al.* (2015) also describes a method for the calcaneal inclination angle measurement. The first step of this mutual information based approach necessitates manual isolation of both the calcaneus and the fifth metatarsal bones. In our case we are able to achieve comparable results while maintaining generic character of the algorithm.

In regards of performance optimization each radiograph could be easily divided into sub-parts. We can then search for key points in each one of them in a different thread. Number of those depends on the utilized CPU. Taking that into account, the choice of GPU seems to be most promising in regard to computation time.

ACKNOWLEDGEMENTS

We are grateful to Zbysław Tabor for insightful comments.

REFERENCES

- Altman DG, Bland JM (1983). Measurement in medicine: The analysis of method comparison studies. *J of the R Stat Soc ser D The Statistician* 32:307–17.
- Arif SMMRA, Knapp K, Slabaugh G (2018). Fully automatic cervical vertebrae segmentation framework for x-ray images. *Comput Methods and Programs in Biomed* 157:95 – 111. [doi: 10.1016/j.cmpb.2018.01.006].
- Arslan G, Yirgin IK, Tasguzen A (2014). A measuremental approach to calcaneal fractures. *Eur J of Trauma and Emerg Surg* 40:593–599. [doi: 10.1007/s00068-013-0359-2].
- Briscoe E, Feldman J (2011). Conceptual complexity and the bias/variance tradeoff. *Cogn* 118:2 – 16. [doi: 10.1016/j.cognition.2010.10.004].
- Donner R, Menze BH, Bischof H, Langs G (2013). Global localization of 3d anatomical structures by pre-filtered hough forests and discrete optimization. *Med Image Anal* 17:1304 – 1314. [doi: 10.1016/j.media.2013.02.004].
- Geurts P, Ernst D, Wehenkel L (2006). Extremely randomized trees. *Machine Learning* 63:3–42. [doi: 10.1007/s10994-006-6226-1].

- Hunter JD (2007). Matplotlib: A 2d graphics environment. *Comput in Sci Eng* 9:90–5. [doi: 10.1109/MCSE.2007.55].
- Jones E, Oliphant T, Peterson P, *et al.* (2011). SciPy: Open source scientific tools for Python. <http://www.scipy.org/>. [Accessed 24 May 2018].
- Kao EF, Lu CY, Wang CY, Yeh WC, Hsia PK (2018). Fully automated determination of arch angle on weight-bearing foot radiograph. *Comput Methods and Programs in Biomed* 154:79 – 88. [doi: 10.1016/j.cmpb.2017.11.009].
- Khoshhal K, El Fouhil A, Al-Nakshabandi N, M Zamzam M, A Al-Boukai A, Zamzami M (2005). Böhler's and gissane's angles of the calcaneus in the saudi population. *Saudi Med J* 25:1967–70.
- Lepetit V, Fua P (2006). Keypoint recognition using randomized trees. *IEEE Trans on Pattern Anal and Mach Intell* 28:1465–79. [doi: 10.1109/TPAMI.2006.188].
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011). Scikit-learn: Machine learning in Python. *J of Mach Learn Res* 12:2825–30.
- Pinto A, Pereira S, Rasteiro D, Silva CA (2018). Hierarchical brain tumour segmentation using extremely randomized trees. *Pattern Recognit* 82:105–17. [doi: 10.1016/j.patcog.2018.05.006].
- Scalzo F, Hamilton R, Asgari S, Kim S, Hu X (2012). Intracranial hypertension prediction using extremely randomized decision trees. *Med Eng and Phys* 34:1058–1065. [doi: 10.1016/j.medengphy.2011.11.010].
- Soltaninejad M, Yang G, Lambrou T, Allinson N, Jones TL, Barrick TR, Howe FA, Ye X (2017). Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in flair mri. *Int J of Comput Assist Radiol and Surg* 12:183–203. [doi: 10.1007/s11548-016-1483-3].
- Waldt S, Woertler K (2014). *Measurements and Classifications in Musculoskeletal Radiology*. New York, NY 10001, USA: Thieme Medical Publishers.
- Wojciechowski W, Molka A, Tabor Z (2016). Automated measurement of parameters related to the deformities of lower limbs based on x-rays images. *Comput in Biol and Med* 70:1 – 11. [doi: 10.1016/j.combiomed.2015.12.027].
- Yang CH, Chou KT, Chung MB, Chuang KS, Huang TC (2015). Automatic detection of calcaneal-fifth metatarsal angle using radiograph: A computer-aided diagnosis of flat foot for military new recruits in taiwan. *PLOS ONE* 10:1–10. [doi: 10.1371/journal.pone.0131387].
- Yates B (2009). *Merriman's Assessment of the Lower Limb*. New York: Churchill Livingstone.