



Artificial intelligence-based solution for x-ray longitudinal flatfoot determination and scaling

Purpose: To develop promising approach for longitudinal flatfoot determination on a base of neural network, which effectively affects the time spending by a radiologist without detection accuracy loss.

Methods: We used 3458 foot radiographs of patients with longitudinal flatfoot and 1726 humans without the foot deformity aged 17-75. Each radiograph used for neural network training was labeled by one radiologist while at testing stage of the study each X-ray image was labeled independently by two radiologists chosen blindly. Diagnostic algorithm was designed on a base of detecting three anatomical points forming the foot arch angle. The artificial intelligence workflow consisted of three-step sequence: a) data preprocessing and preparation for neural network segmentation; b) segments three areas as bounding boxes around required three points; c) location of each of the required points was found inside the relevant area, and appropriate angle measure and flatfoot degree were calculated. The segmentation network was encoder-decoder type convolutional neural network based on U-Net architecture with skip-connections, where ResNet50 is used as encoder, and transposed convolutions were used in decoder for upsampling the result after bottle-neck.

Results: We created effective, robust and fast artificial intelligence-based method, that shows the results in general not worse than radiologists and requires about 6000 times less time.

Conclusions: the artificial intelligence developed is an effective tool for longitudinal flatfoot determination by X-ray image segmentation and the foot arch angle calculation. It may be considered as a rapid assistant as accurate as experienced radiologist.

KEYWORDS: longitudinal flatfoot ■ convolutional neural network ■ foot arch angle ■ radiographs ■ artificial intelligence ■ machine learning ■ semantic segmentation

Introduction

Longitudinal flatfoot (LF) is particular consequence of the foot osteal architecture with flattened or even collapsed longitudinal arch composed by calcaneus, tarsal and metatarsal bones [1-3]. Congenital condition accompanied by valgus leg deformity is relatively rare and considered as a malformation [2,4,5]. Muscular and fascial foot disturbances may as well be associated with rickets, paralysis and trauma [4]. In any case the pathology causes severe leg damage during intensive walking and running due to lack of the foot springiness. Such condition is of crucial importance for professional sport and military service [6].

Great stride of science and technology has resulted in broad variety of diagnostic approaches appeared during past decades for the pathology detection. Each of them has definite advantages and limitations. Traditional and most common manual options for the pathology detection are caliper and ink-stained footprints measurement [7,8]. Both of

the methods are indirect and, hence spare place for the measurements fluctuation. Several novel methods have been proposed to gain precise foot measurements, reduce error rate and assist orthopedists in routine and rudimentary manual estimations. High-sensitive ultrasonic sensor for the height of the longitudinal arch determination was invented and proposed by Hamza et al., as an effective tool for flatfoot diagnostics [9]. The approach, nonetheless based only on the arch height does not take into consideration angles and dimensions, which have particular meaning for quantification of the measurement results. Scientific team, led by Navarro described sensor plate for detection of footprint pressure distribution and subsequent results computation [10]. Although, high cost of the method in comparison with conventional ones sufficiently limits its broad spread.

Machine processing of the obtained data is another promising way to decrease the measurement cost, reduce noise during estimations and increase preciseness. Podoscopy,

Lilian Nitris¹,
Anna Varfolomeeva¹,
Dmitry Blinov^{1*},
Irina Kamishanskaya²,
Alina Lobishcheva²,
Sergey Dydykin³ &
Ekaterina Blinova⁴

¹Research and Development Department, Medical Scientific Department, 10 Tverskoy-Yamskoy Lane, Moscow, 125009 Russia

²Department of Oncology and Radiology, Saint-Petersburg State University, 7-9 University Embankment, Saint-Petersburg, 199034 Russia

³Department of Clinical Anatomy and Operative Surgery, Sechenov University, 8/2 Trubetskaya Street, Moscow, 119991 Russia

⁴Department of Faculty Surgery, National Research Mordovia State University, 68 Bolshevistskaya Street, Saransk, 430005 Russia

*Author for correspondence :
d.blinov@cmai.team

footprinting and three-dimensional the arch reconstruction were used by Lee et al., Maestre-Rendon et al., for development low computational-cost footprint deformities diagnosis sensors and image processing system to make rapid and precise classification [11,12]. At the same time, all aforementioned methods consider shape of a foot as an object for measurements, whereas an experienced practitioner must build the true estimation on bone skeleton of the anatomical area, because a volume of soft tissues consisting of skin, muscles and subcutaneous fat frequently causes discrepancies of measurement.

Analysis of foot radiographs allows to calculate true estimations, angles and indexes, which reflect anatomical architecture of the foot [13]. X-ray is one of the most common and relatively cheap way to capture skeletal image of the area. Jian et al., in 2014 described Cloud image processing and analysis based flatfoot classification method [14]. The authors compared four different computational algorithms for foot X-ray pictures analysis and classification used cloud solution for the images collection and processing. Using technology of deep machine learning (ML) and neural networks (NN) as another kind of artificial intelligence (AI) should have made significant progress in foot deformity determination and taken advantage in scaling of LF. So, the development of novel AI-based solution for LF determination and scaling was the main objective of our study.

Materials and Methods

■ Ethics

The study protocol was reviewed and approved by Saint-Petersburg State University and Sechenov University Ethic Committees at joint meeting with Care Mentor Laboratory representatives on April 2, 2016 (Report No. 127/04/16).

■ Source and labeling of X-ray images

Source of the radiographs: Foot X-ray images have been taken from 3458 patients of both sexes with LF and 1726 humans without the foot deformity aged 17-75 at Saint-Petersburg State University Clinical Hospital and Sechenov University Traumatology Clinic from 2016 until 2019.

Gender characteristics of the study participants: Gender characteristics of the participants are shown at TABLE 1. Informed consent has been obtained from each subject. All the radiographs were depersonalized at the site of the University Clinic before processing through training or testing.

Radiographs labelling: All collected images were randomly assigned at 3:1 ratio to training and testing sets respectively. Radiographs used for the CNN training were labeled by 5 equally-educated and well-experienced radiologists with more than 10 years at a position (each image by single practitioner). The images were blindly designated to the radiologist's personal account in Care Mentor labeling software particularly developed for the study and secured by login and password. Labeling process comprised consequently logging in, browsing through the images pending list, searching and highlighting of three anatomical points described below. The software allowed the radiologists correct their marks' position until the data having sent to the image preprocessing.

At testing stage of the study each X-ray image was labeled independently by two radiologists chosen blindly. It made possible further consideration of the CNN and the radiologist's results divergence as well as variability of the practitioner's opinion.

■ X-ray method of longitudinal flatfoot detection

Currently there are several approaches to X-ray-based detection of LF. The most conventional way is by the foot arch angle [3,15-17]. The blunt angle is formed by intersection of two lines: the first of them is drawn through the lowest two points at the lower edge of the fifth metatarsal bone, while the other one connects two most prominent point on the lower edge of calcaneus (FIGURE 1) [17]. The angle value 165° and more is considered as LF [3].

Table 1. Gender characteristics of the study participants.

Presence of the foot deformity	Sex	Number	Average age, M ± MSE
Longitudinal flatfoot	Male	2051	37.2 ± 3.4
	Female	1407	34.4 ± 4.1
Without the deformity	Male	1114	43.7 ± 2.8
	Female	612	46.1 ± 3.6
Total		5184	40.5 ± 3.2

We find it more correct to use another method of LF determination measuring Costa-Bertani's angle [17], according to which a radiologist first searches three points on the radiograph FIGURE 2 as follows: point A is the lowest margin of cuneonavicular joint; point B is lower margin of calcaneus; and point C corresponds with the lowest margin of the first metatarsal. BAC angle the authors consider as the foot arch angle, whereas the perpendicular drawn from point A to the segment BC (ground) corresponds with the arch height (h). Described way allows building the angle based

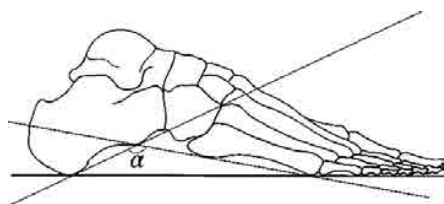


Figure 1. Detection of LF by the foot arch angle estimation [3].

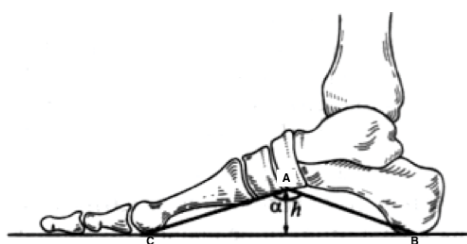


Figure 2. Detection of LF by both the arch angle (α) and height (h) estimation.

on three definite anatomical points unlike the [3] method, that hypothetically presumes the angle summit locate offside the foot lower osteal margin.

We should emphasize that each aforementioned X-ray-based way of measurements puts an exhaustive burden on a radiologist making the physician spend up to 10-12 minutes just on measurements. Accuracy of estimations directly depends on radiologist's experience and X-ray image quality. Experienced physician can overcome even bad quality of the radiograph, but it costs the specialist extra-time and attention.

■ Workflow

Overview of proposed workflow: Our proposed workflow is depicted on FIGURE 3. The workflow consists of three major steps. The first step (e.g. section A) deals with data preprocessing and preparation for the NN segmentation. On the second step (e.g. section B) fully convolutional neural network (CNN) segments three areas as bounding boxes around required three points. On the third step (e.g. section C), location of each of the required points is found inside the relevant area, and appropriate angle measure and flatfoot degree are calculated.

Data pre-processing: All our longitudinal footprint case was randomly divided into training, validating and testing parts in proportion 0.6:0.15:0.25. Training part was

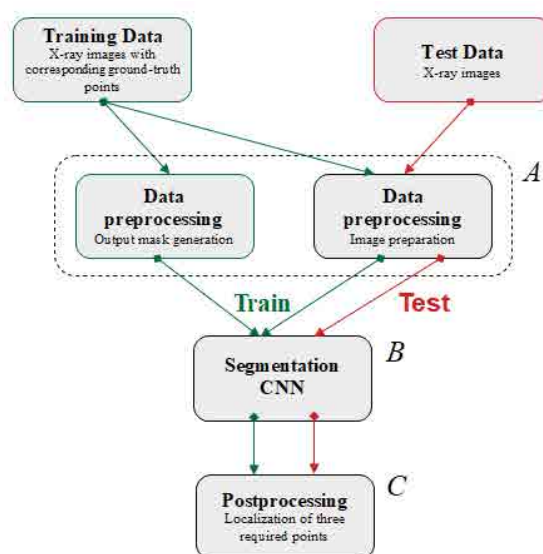


Figure 3. Proposed workflow: section A – data preprocessing and preparation; section B – CNN segments three areas as segment boxes; section C – foot arch angle calculation.

used to train CNN, validating part was used for validate CNN quality during training, and testing part was used to test quality of trained CNN, our method overall quality and difference between radiologists' markings. For training and validating, we used one marking from only one radiologist for each case. For testing, to calculate difference between radiologists' markings, we used two marking from two different radiologists for each case. Each our training longitudinal footprint case contained gray-scale X-ray image and marking - a position (in pixels) of three points for the calculation of the angle required.

Input images had significantly various resolution (approximately from 800 pixels to 4000 pixels on one side), various scope (some images covered only foot, some captured part of the *tibia* bone) and different contrast level. Also, due to the different X-ray apparatus from which the items were obtained, images had different

both detailing quality and noises. Examples of the input footprint images are shown in FIGURE 4. These images were used for the CNN input.

We used position of three points to generate Boolean mask the same size as relevant input image was. For each point, if point position is (x, y) , then mask has value of 1 in the bounding box with corners $(x-k, y-k)$, $(x-k, y+k)$, $(x+k, y-k)$, $(x+k, y+k)$, where k – parameter of bounding box size, that can be changed for different image scale within limit, that bounding boxes do not overlap. So we created the mask with three 1-value bounding boxes as the CNN output.

To form dataset for the CNN training, each image and corresponding mask were rescaled to the size of 512×512 pixels. Since input images had very different resolutions and aspect ratios (in pixels), cropping or padding preceded

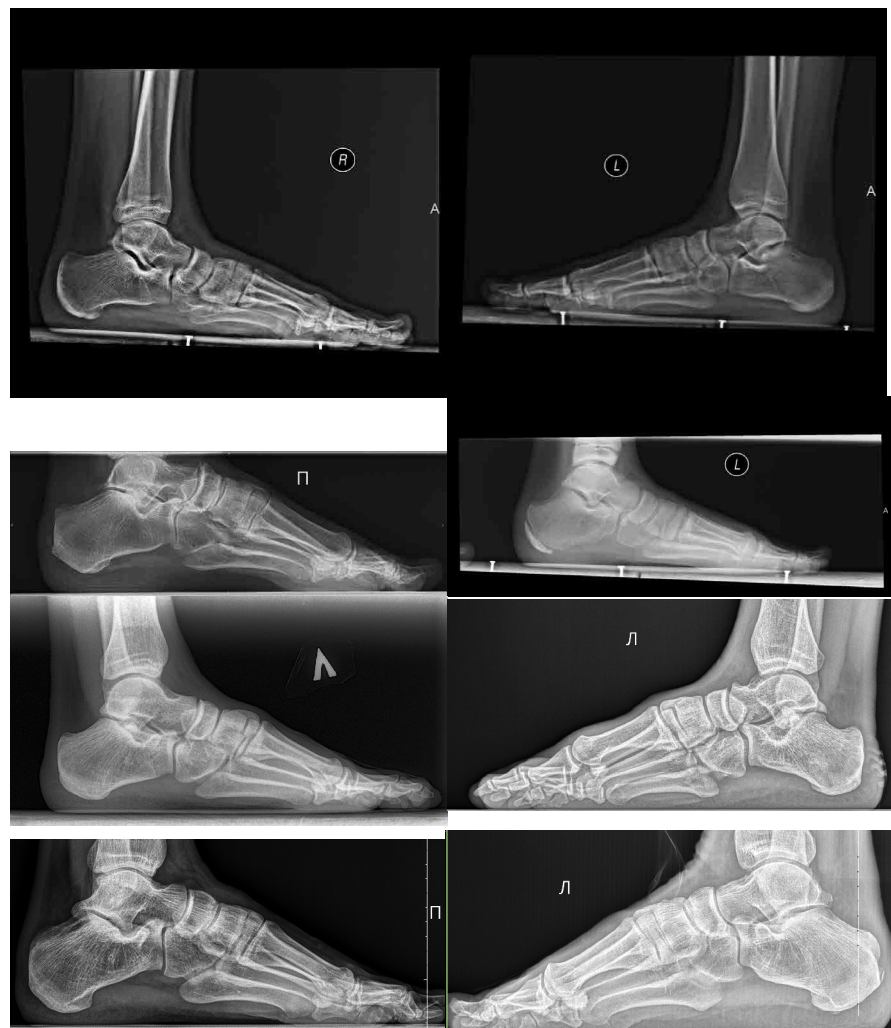


Figure 4. Examples of input X-ray images (some of them are cropped for better representation).

rescaling: the informative part (non-black part with footprint) is rectangle, but it is often surrounded by black frame. So we removed part of black frame to get square image, if it possible, or added parts of black frame otherwise, and then the modified image was rescaled. The same operations are performed over corresponding masks.

Several data augmentation steps, such as translation, rotation, sharpening, weak affine transformations, contrast normalization and addition of Gaussian noise were employed during training to increase the diversity of the training data. It's worth to note, that augmentation was applied during training at each training epoch, and all of augmentation steps are random (for example, rotate image at random angle from -5 to +5 degree, sharp image with random parameters from given interval, etc). So, all training images during training procedure become different from each other, but

remain similar and contain the same meaningful information.

CNN architecture: Our segmentation network is encoder-decoder [1] type CNN, which is suitable solution for our semantic segmentation task. Our initial network architecture is schematically presented on FIGURE 5A.

In the first initial layers of the neural network, spatial information is present in the activations of the current layer: these layers of CNN activate simple features like different parts of lines, angles, simple textures, etc. In later layers, because of using convolutions, which aggregate information from previous layers, spatial information gets transferred to semantic information at the cost of specific knowledge on the localization of these structures. Here, for example, the original U-Net architecture reduces an input image of size 388×388 to a size of

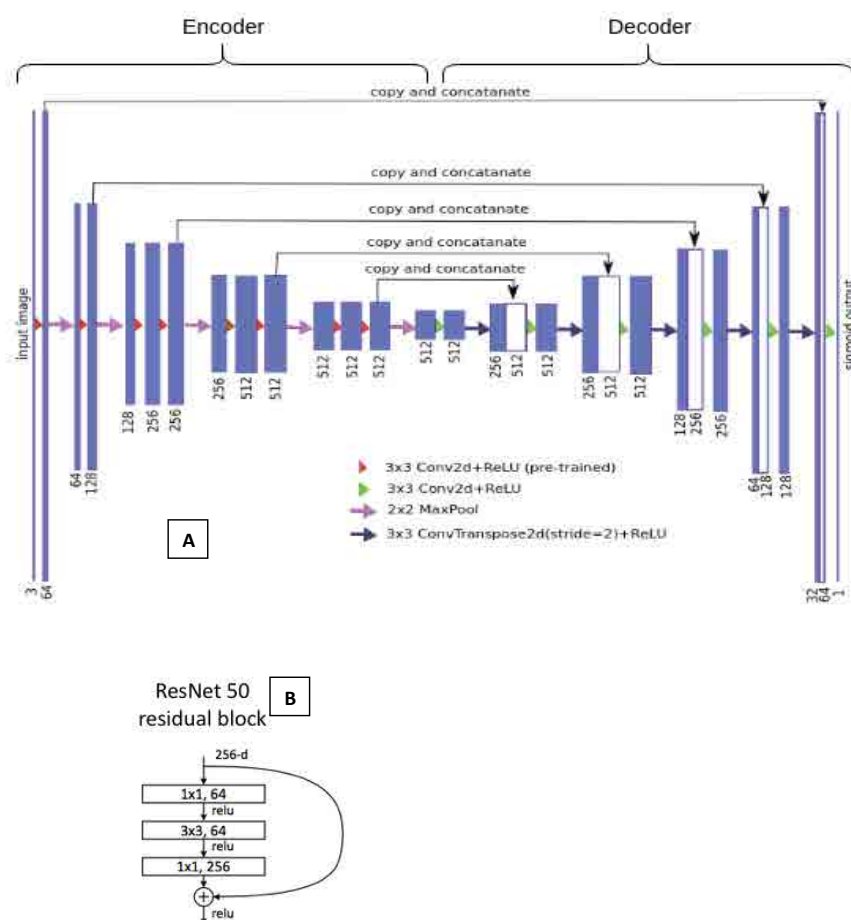


Figure 5. Basic U-Net architecture (A) and ResNet50 residual block (B). Our network is based on U-Net architecture [1] with skip-connections, where ResNet50 [11] is used as encoder, and transposed convolutions [7] are used in decoder for upsampling the result after bottle-neck. The encoder output grid size is 16-16, the last fully convolutional layer output matches the input dimension.

28 × 28 in the U-Net bottleneck. Ronneberger et al., [17] introduced skip-connections to allow utilization of spatial and semantic information at later layers, since the spatial information from earlier stage can be fused in the neural network at later layers. Thus the neural network at later layers can utilize both semantic and spatial information: connect features from earlier layers with features from later layers, as it is shown by arrows on FIGURE 5A.

Common approach in deep learning is transfer learning using pre-trained NN models. Neural networks pre-trained on another task, e.g. a natural image classification data set, can be used as initialization of the network weights when training on a new task. The first layers of neural networks learn simple features and basic structures like blobs and edges, so this knowledge can be transferred from one task to others. This concept is very useful for medical imaging, where there is no possibility to obtain large datasets like natural image datasets. In our work, we use pre-trained on ImageNet [18] ResNet50 [19] model as encoder, and decoder was trained from scratch. ResNet50 is a deep residual network, showed good quality on different tasks and which is easy to train because of using residual connections between inner blocks shown on FIGURE 5B.

CNN training: Input images and their corresponding segmentation maps were used to train the network with Adam [20] optimization. The binary cross-entropy function was calculated pixel-wise with different weights for each class pixels as in equation (1):

$$L = - \sum_{ij} w_{ij} (y * \log(p) + (1 - y) * \log(1 - p)) \quad (1)$$

where $\hat{p}$ – predicted probability, y – Indicator of ground-truth class (0 or 1) and w – per-pixel weight matrix. As most of the pixels in each image belong to the zero-label, we balanced the learning process by using fixed weights that were inversely proportional to the population ratios. We trained CNN with Adam optimizer for 500 epochs with standard parameters: betas 0.9 and 0.99, the initial learning rate was chosen to be 0.0001, with reducing learning rate on plateau.

CNN post-processing: For each area D_k of the three predicted areas from the CNN, we found its mass center as the location of relevant point (x_k, y_k) :

$$x_k = \frac{\sum_{(x_j, y_j) \in D_k^{xj}} x_j}{|D_k|}, y_k = \frac{\sum_{(x_j, y_j) \in D_k^{yj}} y_j}{|D_k|}, k = 1, 2, 3 \quad (2)$$

Found this way three required points, we can simply compute angle measure, as it is shown above on FIGURE 2.

■ Quality measures

CNN quality evaluation: We used the Dice score as a main metric for segmentation CNN quality evaluation. We referred to the foreground areas in the ground truth as object A, and object B for the predicted areas.

The Dice score was evaluated as (3):

$$DICE(A, B) = \frac{2 |A \cap B|}{|A| + |B|} \quad (3)$$

where the Dice score was in the interval (0,1). A perfect segmentation yields a Dice score of 1.

Overall quality evaluation: We calculate mean absolute angle error as by (4):

$$Err = |a - a_{true}| \quad (4)$$

where $\hat{\alpha}$ – predicted angle, α_{true} – angle, calculated from ground truth points.

■ Statistics

We presented the results obtained as Mean (M) ± Mean square error (MSE) value. Variants distribution was assessed by ANOVA. *T*-test was used to compare differences between the groups, and they were considered as significant at $p < 0.05$. SPSS software (IBM Inc., USA) was used for statistical data proceeding.

Results

■ Segmentation quality

Our test set contained 1296 cases. Mean Dice quality on our test set was 0.946. The largest set of errors was related to segmentation mask borders shape and did not affect overall method quality. Examples of the CNN output and how it localized required areas were shown in FIGURE 6A.

■ Quality comparison

For evaluation our method quality, for every test case we compared the markings from two different radiologists, who worked independently, with the inference of the CNN solution. Examples of comparison are shown on FIGURE 6B, where light and dark green lines and points refer to radiologists markings, and red color refers to the CNN marking. In

angle measurement, average difference between radiologists' markings is *1.18 degree*, and average difference between out method received angle and radiologist's marking is *1.27 degree* ($p>0.05$). Mean deviation for every flatfoot degree is shown on TABLE 2. There were not significant differences in the angle measurements between the radiologists and the artificial intelligence-based solution in cases of the foot pathology. At the same time, intergroup analysis of the

radiographs without foot pathology showed that the angle value Mean calculated by the CNN deviated from human-made measurements approximately twice bigger then between two independently working specialists.

■ Working time assessment

The obtained results are shown at TABLE 2. The time spent by the radiologists on searching for the three anatomical points on radiograph

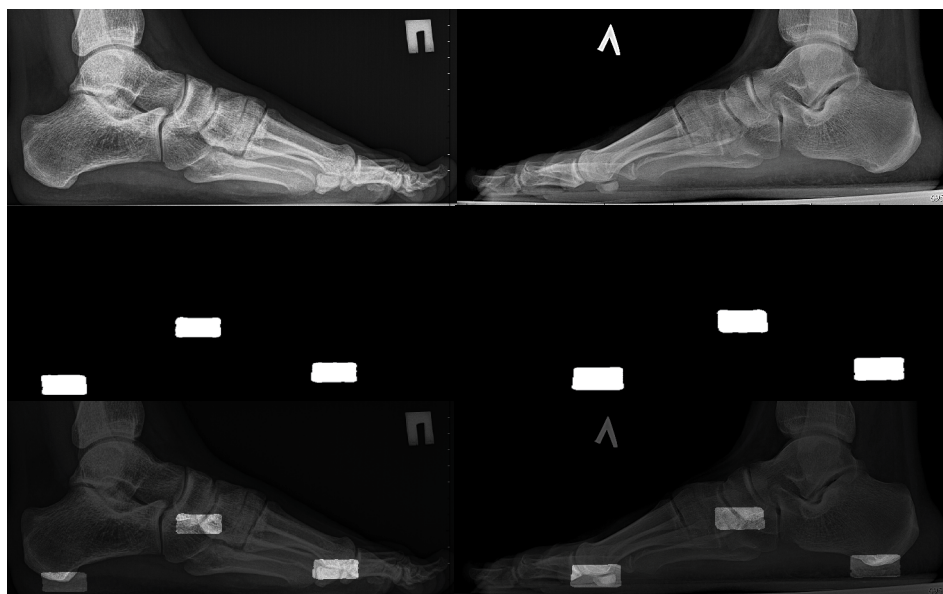


Figure 6. (A). Top to bottom: CNN input, CNN output mask, mask applied to the image, highlighting predicted areas.

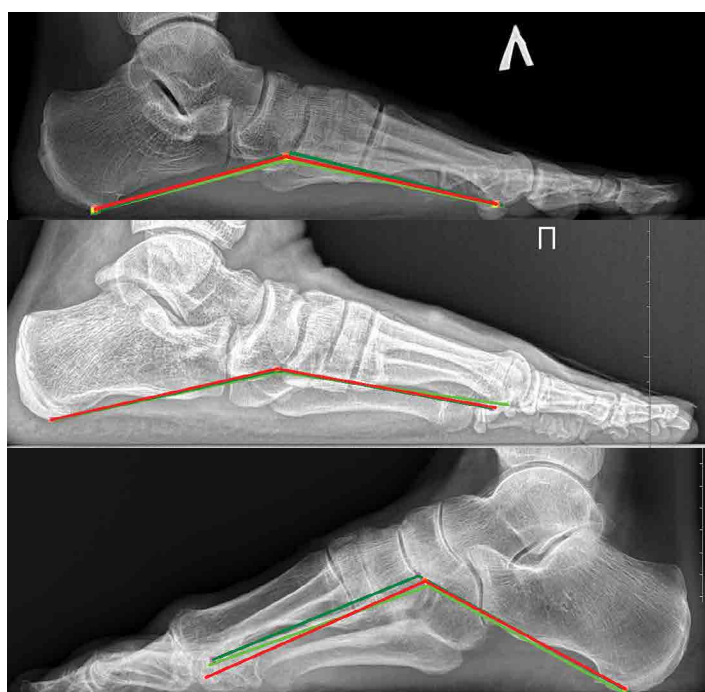


Figure 6. (B). Comparison of radiologists and proposed methods markings.

Table 2. Mean deviations according to flatfoot degree and time spending for image processing.

Flatfoot degree	Radiologists estimation M ± MSE		CNN vs radiologists deviation, M ± MSE	
	Angle deviation	Time spent on estimation	Angle deviation	Time spent on estimation
0 (125-130)	0.60 ± 0.21	667.72 ± 72.81	1.04 ± 0.17 [†]	0.10 ± 0.02 [‡]
1 (131-140)	1.10 ± 0.22		1.19 ± 0.18	
2 (141-155)	1.27 ± 0.19		1.30 ± 0.21	
3 (>155)	1.13 ± 0.17		1.36 ± 0.18	

Note: [†]p<0.05 (t-test) when compared with radiologists-estimated value; [‡]p<0.05 (t-test) when compared with the time spent on measurement by the radiologists.

and subsequent the foot arch angle estimation was averaged at 667.7 ± 72.8 sec. The time spent by our artificial intelligence solution was obtained using GPU model Titan V and averaged at 0.10 ± 0.02 sec ($p=0.001$ when compared with the radiologists).

Discussion

Longitudinal foot deformity has particular meaning in childhood and for adults because of severe outcomes deteriorating health condition and life quality. The pathology diagnostics based on the foot arch both shape and volume evaluation, footprints measurements associates with high level of incorrect decisions largely because all of the approaches are indirect. Direct, and importantly more correct way to handle the issue is to take into consideration bone-referred points that allow to calculate both the arch features, angle and height. From this point of view X-ray foot scanning seems more appropriate approach for doing the estimations. At the same time, conventional sequence of plain measurements and calculations is utterly exhaustive and time-consumable, so optimization of the algorithm may be very helpful and useful. Implementation of deep learning methods for detection of anatomical features crucial for the pathology detection and subsequent calculation of the arch angle may be useful tool for routine radiological practice.

We choose convolutional neural network as a base of our method because of its recognized efficiency and stability. Also, its rapidity allows to receive almost real-time estimation's result. CNNs enable us to work with both complex and basic features of images, which makes successfully work with different input data possible.

We use semantic segmentation CNN as robust and successful method for task, there the case is to find some mask on the image. We

choose U-net architecture, because it's proved to be successful in training on small data, and using of ResNet50 encoder allow us to use pre-trained weights, which also helps training on the small amount of data.

As a result, segmentation quality of our method was high and comparably equal with the quality of trained practitioners. Moreover, CNN-based model spent no more than 0.1 sec on getting the result versus 667.72 sec of a practitioner's average time for the analysis.

However, it is worth noting that using the method on different X-ray machine, the results may differ in quality because of disparate image characteristics (mostly, different image noises). Such output discrepancies can be improved by additional training CNN on the cases, obtained from new X-ray machine. Also, if there are specific cases, where radiologist may mark that the quality of method is weak, the method can be additionally trained on them for quality improvement.

One of the most important shortcomings we met was radiologist labeling of the images used for the CNN training, since we had found out that the specialist's accuracy in anatomical points determination had been varied within 1.18 degree. Therefore, next step in developing effective and autonomous system, which should span computational identification of anatomical objects and machine visualization.

Thus, the artificial intelligence developed during the study is an effective tool for longitudinal flatfoot determination by X-ray image segmentation and the foot arch angle calculation. It may be considered as a rapid assistant as accurate as experienced radiologist.

Conclusions

Established CNN-based method is an effective, robust and rapid solution for LF

detection based on the foot X-ray image preprocessing, anatomical points capturing and subsequent computation of results presenting as the foot arch angle value with the pathology degree according to appropriate scale. The AI solution shows segmentation quality results not worse than trained radiologists and requires about 6000 times less time for the parameters estimation and the inference issuing.

Author Contributions

Anna Varfolomeeva, Lilian Nitris and Dmitry Blinov contributed to the study conception and design. Material preparation, data collection and analysis were performed by Anna Varfolomeeva,

Irina Kamishanskaya, Ekaterina Blinova and Alina Lobishcheva. The first draft of the manuscript was written by Anna Varfolomeeva, Sergey Dydykin and Dmitry Blinov and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We are extremely grateful to the radiologists who labeled foot X-ray images, and whose driven devotion made the study go. We thank all our critics and colleagues for their assistance and valuable advices.

REFERENCES

1. Deland JT. Adult-acquired flatfoot deformity. *J. Am. Acad. Orthop. Surg.* 16: 399-406, (2008).
2. Dare DM, Dodwell ER. Pediatric flatfoot: cause, epidemiology, assessment, and treatment. *Curr. Opin. Pediatr.* 26: 93-100, (2014).
3. Smyth NA, Aiyer AA, Kaplan JR *et al.* Adult-acquired flatfoot deformity. *Eur. J. Orthop. Surg. Traum.* 27: 433-439, (2017).
4. Carr JB II, Yang S, Lather LA. Pediatric pes planus: a state-of-the-art review. *Pediatrics.* 137: 1-12, (2016).
5. Ford SE, Scannell BP. Pediatric flatfoot: pearls and pitfalls. *Foot. Ankle. Clin.* 22: 643-656, (2017).
6. Pavone V, Vescio A, DiSilvestri CA *et al.* Outcomes of the calcaneo-stop procedures for the treatment of juvenile flatfoot in young athletes. *J. Child. Orthop.* 12: 582-589, (2018).
7. Chou PY, Chen YY, Liao CH. The influence of different imaging postures on flatfoot evaluation. *Chin. J. Rad. Tech.* 33: 63-68, (2009).
8. Kouchi M, Mochimaru M, Tsuzuki K *et al.* Interobserver errors in anthropometry. *J. Hum. Ergol.* 28: 15-24, (1999).
9. Hamza AO, Ahmed HK, Khider MO. A new noninvasive flatfoot detector. *J. Clin. Eng.* 40: 57-63, (2015).
10. Navarro LA, García DO, Villavicencio EA *et al.* Opto-electronic system for detection of flat foot by using estimation techniques: Study and approach of design. *Annual. Intl. Conf. IEEE. Eng. Med. Biol. Soc.* 10: 5768-5771, (2010).
11. Lee YC, Lin G, Wang MJJ. Comparing 3D foot scanning with conventional measurement methods. *J. Foot. Ankle. Res.* 7: 44-48, (2014).
12. Maestre-Rendon JR, Rivera-Roman TA, Sierra-Hernandez JM *et al.* Low computational-cost footprint deformities diagnosis sensor through angles, dimensions analysis and image processing techniques. *Sensors.* 17: 2700-2712, (2017).
13. Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning. *Arxiv. Lib.* 10: 1-31, 2018.
14. Jian MS, Shen JH, Chen YC. Cloud image processing and analysis based flatfoot classification method. *Int. J. Comput.* 8: 90-98, (2014).
15. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE. T. Pattern. Anal.* 39: 2481-2495, (2017).
16. He K, Zhang X, Ren S. Deep Residual Learning for Image Recognition. *IEEE Comput. Vis. Pattern. Recognit.* 10: 1-4, (2016).
17. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Arxiv. Lib.* 15: 1-8, (2015).
18. <http://www.image-net.org/index>
19. Kaiming He, Xiangyu Zhang, Shaoqing Ren *et al.* Deep Residual Learning for Image Recognition. *Arxiv. Lib.* 15: 1-12, (2015).
20. <https://arxiv.org/abs/1412.6980>.