# Automation of clinical measurements on radiographs of children's hips

Peter Thompson[1], Medical Annotation Collaborative[2], Daniel C. Perry[2,3], Timothy F. Cootes[1], and Claudia Lindner[1]

[1] The University of Manchester, UK
[2] University of Liverpool, UK
[3] Alder Hey Children's Hospital, UK
peter.thompson-2@manchester.ac.uk

**Abstract.** Developmental dysplasia of the hip (DDH) and cerebral palsy (CP) related hip migration are two of the most common orthopaedic diseases in children, each affecting around 1-2 in 1000 children. For both of these conditions, early detection is a key factor in long term outcomes for patients. However, early signs of the disease are often missed and manual monitoring of routinely collected radiographs is time-consuming and susceptible to inconsistent measurement. We propose an automatic system for calculating acetabular index (AcI) and Reimer's migration percentage (RMP) from paediatric hip radiographs. The system applies Random Forest regression-voting to fully automatically locate the landmark points necessary for the calculation of the clinical metrics. We show that the fully automatically obtained AcI and RMP measurements are in agreement with manual measurements obtained by clinical experts, and have replicated these findings in a clinical dataset. Such a system allows for the reliable and consistent monitoring of DDH and CP patients, aiming to improve patient outcomes through hip surveillance programmes.

**Keywords:** Clinical decision support system · Automated radiographic measurement · Acetabular index · Reimer's migration percentage

## 1 Introduction

This work is concerned with the automatic derivation of clinically applicable metrics on paediatric hip radiographs, specifically acetabular index (AcI) and Reimer's migration percentage (RMP). These metrics are commonly used in the diagnosis of developmental dysplasia of the hip (DDH) and monitoring cerebral palsy (CP) related hip migration. AcI attempts to characterise the dysplasia of the joint by estimating the angle of the acetabular roof, and RMP estimates the proportion of the femoral head that is not covered by the acetabulum.

AcI is a long-standing metric in the determination of hip dysplasia and is considered the most useful for children under the age of 8 years old [3, 15]. The value of AcI for a healthy hip is expected to decrease significantly in the first few years of life, typically falling below 20° by 2 years of age, with 30° sometimes

used as a rough upper bound for a healthy hip [20]. The reliability of the metric has been called into question, with small variations in the orientation of the pelvis [2] and the interpretation of the relevant anatomical features [9] having been identified as significant sources of error. However, AcI remains the preferred radiographic metric for assessing hip dysplasia in children.

RMP is considered a reliable and repeatable metric for assessing the severity of CP related hip migration [18]. Values greater than 33% are usually used to distinguish cases of subluxation, with 100% to distinguish dislocation [20, 19]. RMP is monitored over time to assess the progression of hip migration [7].

There exist many techniques for detecting landmarks in medical images. Several contributions have used Random Forest based techniques [8], sometimes in combination with Markov random fields (MRF) [5, 21]. More recently, convolutional neural networks (CNN) have been applied to the problem. These approaches often attempt to directly regress on the landmark coordinates and are generally improved by the inclusion of a model capable of representing shape constraints, such as SSMs [1], coupled shape models [23], or a graphical component in the CNN itself [11]. Another common CNN approach is to use fully-connected networks to predict heatmaps [16], displacement-maps [24] or both [4].

We use the Random Forest Regression-Voting Constrained Local Models (RFRV-CLM) framework [14, 12] to automatically locate landmark points which are used to calculate AcI and RMP. RFRV-CLM has been used successfully and achieved state-of-the-art performance in several domains of medical imaging, including bone segmentation [13], automatic cephalometry [22], and wrist fracture detection [6]. Studies of the feet and paediatric hips demonstrated that the technology outperforms deep learning networks on smaller datasets [10, 4].

Pham et al. [17] used CNNs to directly regress landmark coordinates in hip radiographs and used those landmarks to calculate RMP. This is the most closely related work to ours, insofar as automatically located points are used to calculate RMP, and is, to our knowledge, the only prior work that performs this task.

**Contributions:** We present a fully automated system to calculate AcI and RMP from hip radiographs. To the best of our knowledge, this is the first system to automatically calculate AcI while also achieving state-of-the-art results for calculating RMP. The system obtains measurements in agreement with clinical experts and has been validated on an independent clinical dataset of 200 images.

## 2 Methods

### 2.1 Images

The initial dataset consisted of 450 pelvic radiographs of children aged 2-18 years-old enrolled in the Outcomes Research in Children's Hip Disease (ORCHID) study (IRAS 227197). This was a challenging dataset containing occlusions and cases of severe disease. One third of cases had DDH, one third had no pathological condition, and one third had Perthes disease. Of the 450 images, 50 were randomly selected to be manually measured by 9 clinicians (8 trainee and 1 senior). Of these 50 cases, 30 had DDH, 10 no condition, and 10 Perthes. Of the

remaining 400 images, one was excluded for image quality reasons (parts of the pelvis cropped). One hip per image was used for each image in DataI.

A replication dataset (DataR) was also produced, consisting of 200 DDH cases of varying severity. The children were aged 2-18 years and one image per child was collected from Alder Hey Children's Hospital. These images were each measured by 5 clinicians. For DataR, both hips in each image were used, resulting in a dataset of 400 hips.

### 2.2 Manual measurements

Both AcI and RMP require a horizontal pelvis line, the Hilgenreiner line, to be found in the image. The Hilgenreiner line passes through the triradiate cartilage on both sides of the pelvis (see Figure 1). AcI is defined as the angle between the Hilgenreiner line and the acetabular roof. Perkin's line is perpendicular to the Hilgenreiner line and is drawn from the end of the acetabular roof. RMP is defined as the proportion of the projection of the femoral head onto the Hilgenreiner line that is on the lateral side of the Perkin's line.

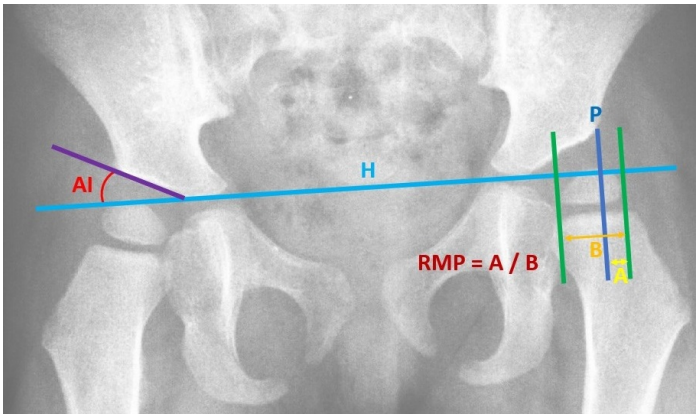



Fig. 1: Diagram showing measurement of AcI and RMP. Line H is the Hilgenreiner line, line P is Perkin's line. AcI is the angle shown on the left. RMP is distance A divided by distance B. Best viewed in colour.

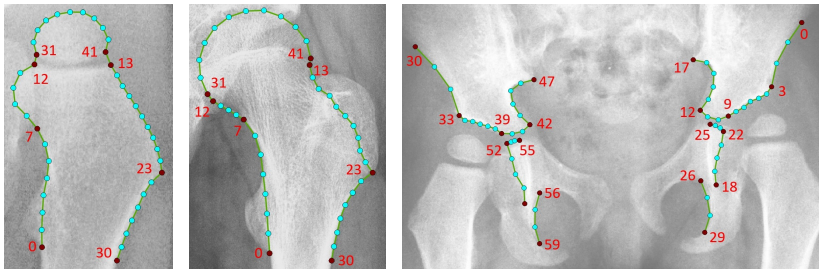
We report the intraclass correlation coefficient type 2 (ICC2; considering both the images and the observers as random effects) and the proportion of variance explained by the differences between observers (PoV) to assess the manual inter-observer agreement for the AcI and RMP measurements.

### 2.3 Automatic search models

Our technique uses the RFRV-CLM framework [14, 12] which allows landmark points to be automatically located in images, taking into account both local image features and global shape. We develop two RFRV-CLM based fully automatic search models, one outlining the proximal femur and the other outlining parts of the pelvis. The models are defined as a set of points, where some points represent localised features in the image, while others are evenly spaced between localised feature points. Both search models are applied in parallel.

The femur model included 42 points along the boundary of the proximal femur, ignoring the trochanters as their shape is not relevant in this study. Since in young children the femoral head is separated from the femoral neck, fixed points were placed on either side of the epiphysis gap and connected by a continuous curve. In older children, the gap disappears and the curve follows the contour of the femoral head. This way, the same model can be used across the range of ages despite the change in topology of the visible bone. Figure 2 shows the femur model as applied to (a) a 2-year-old patient, for whom the femoral head is separated and (b) an 18-year-old patient, for whom it is connected.

The pelvis model included 60 points in total and was defined to outline at least the regions of the pelvis relevant to measuring AcI and RMP, i.e. the acetabular roof and triradiate cartilage. Other distinctive features of the pelvis were also included to provide more context to the shape model (as part of the RFRV-CLM framework). The impact of including these extraneous features is demonstrated in Section 3.2. The pelvis model is shown in Figure 2c.




(a) (b) (c)

Fig. 2: The femur model for (a) a 2-year-old and (b) an 18-year-old subject, and (c) the pelvis model. Localised feature points are shown in red and labelled with their index. The indices of the evenly spaced points are consecutive to these.

For each of the 449 images in DataI, we manually placed point positions as shown in Figure 2 to create the ground truth. Based on the latter, we ran three-fold cross-validation experiments (using randomised folds) to assess the performance of the search models.

Two separate models were used to represent the femur and pelvis because the position of an extremely migrated proximal femur is highly variable, and thus the statistical modes of shape variation of a single model may not be sufficient to capture the range of position in these outlying cases. A comparison of the performance of two separate models with a single model is given in Section 3.2.

The performance of the search models was assessed using the point-to-point and point-to-curve errors between the manual ground truth annotations and the automatically located point positions.

### 2.4 Automatic Measurements

The automatic point positions used to calculate AcI and RMP for DataI were obtained from the three-fold cross-validation experiments. To also automatically calculate AcI and RMP for the replication dataset DataR, we retrained the proximal femur and pelvis search models using all images and manual annotations from DataI. We then ran these models over all images in DataR, and calculated AcI and RMP based on the automatically located point positions.

As per Figure 2c, points 9 and 39 were used to define the Hilgenreiner line, points 5 and 8 to define the acetabular roof, and point 3 to define the Perkin's line. The bounds of the femoral head were taken to be the points of maximum separation in points 31-41 in Figures 2a and 2b when projected onto the Hilgenreiner line. AcI and RMP were then calculated as described in Section 2.2.

## 3 Results

### 3.1 Manual measurements

The agreement of the manual measurements is given in Table 1. The ICC2 scores for AcI suggest good agreement on DataI and moderate agreement on DataR. For RMP, the ICC2 scores suggest good agreement for both DataI and DataR. Similarly, the PoV explained by the difference between observers suggests good agreement between observers in all experiments. Overall, the manual inter-observer agreement was slightly better on DataI compared to DataR.

Table 1: Inter-observer agreement metrics for the manual measurements. (AcI = acetabular index; RMP = Reimer's migration percentage; ICC2 = intraclass correlation coefficient type 2; PoV = proportion of variance explained by the differences between observers).

| | AcI | | RMP | |
|---|---|---|---|---|
| Dataset | ICC2 | PoV (%) | ICC2 | PoV (%) |
| DataI | 0.77 | 3.6 | 0.88 | 4.1 |
| DataR | 0.66 | 9.4 | 0.83 | 5.7 |

### 3.2 Landmark detection

The results of the three-fold cross-validation experiments to assess the landmark detection performance of the search models are shown in Table 2.

Table 2: Performance of the automatic landmark detection, including all femur and pelvis points (based on applying both search models in parallel).

| Measure | Mean | Median | 95th %ile | 99th %ile |
|---|---|---|---|---|
| Point-to-curve (mm) | 1.02 | 0.81 | 1.83 | 2.98 |
| Point-to-point (mm) | 2.72 | 2.29 | 4.93 | 7.24 |

The performance of the landmark detection when two separate models were used for the femur and pelvis, compared to using a single model incorporating all femur and pelvis points, is shown in Figure 3a. It can be seen that the two separate models significantly outperform the combined model. Figure 3b compares the results of the complete pelvis model including all points, as shown in Figure 2c, with a reduced pelvis model including only the parts of the pelvis relevant to measuring AcI and RMP (points 3-12 and 33-42 as in Figure 2c). For a fair comparison, only the points in the reduced model were included when analysing the results for the complete model in Figure 3b. The results show that the complete pelvis model has a slight but consistent advantage over the reduced pelvis model.

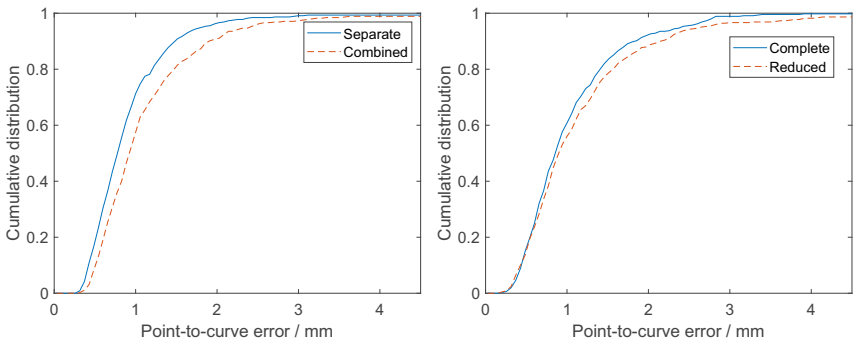



Fig. 3: Cumulative distribution functions (CDFs) showing the point-to-curve results comparing (a) two separate to a single combined model, and (b) the complete pelvis model to a reduced version including only a minimal set of points.

### 3.3 Automatic measurements

Table 3 gives the results of comparing the fully automatically obtained measurements to those made manually by the clinical experts. These show good agreement between the automatic results and the manual measurements for both AcI and RMP across the original and replication datasets. The linear mixed-effects model results demonstrate that the automatic measurements give an unbiased estimate of the manual measurements (given the p-values are not significant), and are thus in agreement with the clinicians.

Table 3: Comparison of automatic and manual measurements showing the type 2 intraclass correlation coefficient (ICC2), percentage-in-range (PiR), and bias estimates with associated p-values based on a linear mixed-effects model.

| | AcI | | | | RMP | | | |
|---|---|---|---|---|---|---|---|---|
| Dataset | PiR | ICC2 | Bias | p | PiR | ICC2 | Bias | p |
| DataI | 84 | 0.86 | -0.70 | 0.46 | 86 | 0.92 | -0.66 | 0.71 |
| DataR | 77 | 0.86 | 0.03 | 0.97 | 82 | 0.90 | -1.57 | 0.26 |

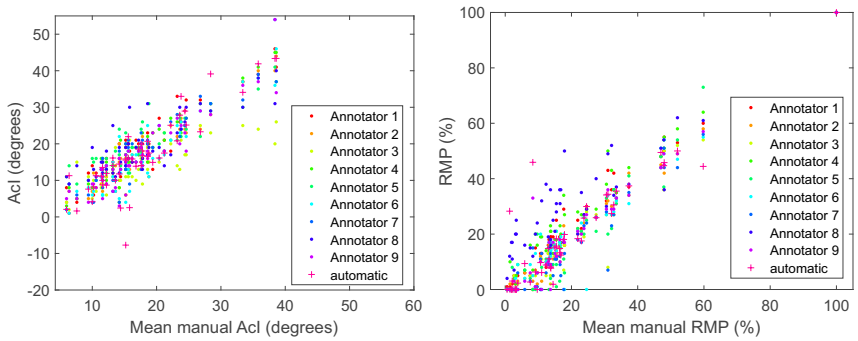
Figure 4 shows scatter plots of the automatic measurements for DataI and DataR alongside all of the manual measurements, demonstrating the agreement between the manual and automatic measurements for both AcI and RMP.

Pham et al. [17] report RMP mean absolute errors of 4.5±4.3 and 4.9±3.9 percentage points between their automatic RMP measurement system and their two raters. The equivalent values for our system are comparable, being 4.5±6.5 on DataI and 3.3±3.7 on DataR, indicating similar performance. However, these values are not directly comparable as the mean absolute error metric also depends on the distribution of RMP values for each of the datasets.

## 4 Discussion and conclusions

We have presented an automatic system to calculate AcI and RMP, two radiographic measurements that are considered among the most important in the diagnosis and monitoring of hip diseases in children. This system is the first to automatically calculate AcI and performs comparably to the only previously reported system to automatically calculate RMP. For both AcI and RMP, we have replicated our findings in an independent dataset and demonstrated that our automatically obtained measurements are in agreement with clinical experts. The system could also be used to derive other common radiographic measurements, such as acetabular depth ratio, Wilberg's angle, and neck shaft angle.

The system was developed and evaluated using a challenging dataset containing (i) occlusions; (ii) cases of severe DDH and Perthes disease; and (iii) a range




(a) AcI - DataI (b) RMP - DataI

(c) AcI - DataR (d) RMP - DataR

Fig. 4: Plots showing the automatic measurements in the context of all manual measurements for each dataset, sorted by the mean manual measurement.

of ages with a large variation in anatomy. The initial dataset also contained a number of post-operative cases in which the appearance of the acetabular roof was ambiguous. These cases account for the largest differences in AcI between the automatic and manual measurements, including the outlier in Figure 4a. A limitation of this work is that, although RMP is primarily used to monitor the progression of CP, the dataset did not include any CP patients. Being able to automatically measure RMP has important implications for national surveillance systems of hip subluxation in patients with CP. The application of our system to images of children with CP is an important direction for future work. A more general problem is the fairly large spread in the manual measurements. This makes the evaluation of an automatic system challenging but also highlights the need for a more consistent method of making these radiographic measurements.

**Acknowledgements** Peter Thompson was supported by the Wellcome Trust University of Manchester TPA [209741/Z/17/Z] and an Artificial Intelligence in Health and Care Award (AI_AWARD02268). Claudia Lindner was funded by the Medical Research Council, UK (MR/S00405X/1) as well as a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (223267/Z/21/Z). Daniel C. Perry is funded by the National Institute for Health Research (NIHR) through a NIHR Research Professorship. Manual measurements were provided by the Medical Annotation Collaborative (Grace Airey, Mohammed Ali, Bishoy Bessada, Benjamin Gompels, Tom Hughes, Mustafa Javaid, Zarmina Kakakhel, Mohammed Khattak, James Redfern, David Samy, Myles Simmons, Lucy Stead). This report is independent research funded by the National Institute for Health Research (Artificial Intelligence, An Automated System for Measuring Hip Dysplasia in Children with Cerebral Palsy, AI_AWARD02268). The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

## References

1. Arik, S.Ö., Ibragimov, B., Xing, L.: Fully automated quantitative cephalometry using convolutional neural networks. Journal of Medical Imaging **4**(1), 014501 (2017)
2. Van der Bom, M., Groote, M., Vincken, K., Beek, F., Bartels, L.: Pelvic rotation and tilt can cause misinterpretation of the acetabular index measured on radiographs. Clinical Orthopaedics and Related Research® **469**(6), 1743–1749 (2011)
3. Broughton, N., Brougham, D., Cole, W., Menelaus, M.: Reliability of radiological measurements in the assessment of the child's hip. The Journal of bone and joint surgery. British volume **71**(1), 6–8 (1989)
4. Davison, A.K., Lindner, C., Perry, D.C., Luo, W., Cootes, T.F., et al.: Landmark localisation in radiographs using weighted heatmap displacement voting. In: International Workshop on Computational Methods and Clinical Applications in Musculoskeletal Imaging. pp. 73–85. Springer (2018)
5. Donner, R., Menze, B.H., Bischof, H., Langs, G.: Global localization of 3d anatomical structures by pre-filtered hough forests and discrete optimization. Medical image analysis **17**(8), 1304–1314 (2013)
6. Ebsim, R., Naqvi, J., Cootes, T.: Fully automatic detection of distal radius fractures from posteroanterior and lateral radiographs. In: Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures, pp. 91–98. Springer (2017)
7. Hägglund, G., Lauge-Pedersen, H., Persson, M.: Radiographic threshold values for hip screening in cerebral palsy. Journal of children's orthopaedics **1**(1), 43–47 (2007)
8. Han, D., Gao, Y., Wu, G., Yap, P.T., Shen, D.: Robust anatomical landmark detection with application to mr brain image registration. Computerized Medical Imaging and Graphics **46**, 277–290 (2015)
9. Kay, R.M., Watts, H.G., Dorey, F.J.: Variability in the assessment of acetabular index. Journal of Pediatric Orthopaedics **17**(2), 170–173 (1997)

10. Lauder, J., Harris, J., Layton, B., Heire, P., Sorani, A., DeSancha, M., Davison, A., Sammut-Powell, C., C., L.: A fully automatic system to assess foot collapse on lateral weight-bearing foot radiographs: A pilot study. Computer Methods and Programs in Biomedicine **213**(106507), 106507 (2022), doi:10.1016/j.cmpb.2021.106507
11. Li, W., Lu, Y., Zheng, K., Liao, H., Lin, C., Luo, J., Cheng, C.T., Xiao, J., Lu, L., Kuo, C.F., et al.: Structured landmark detection via topology-adapting deep graph learning. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16. pp. 266–283. Springer (2020)
12. Lindner, C., Bromiley, P.A., Ionita, M.C., Cootes, T.F.: Robust and accurate shape model matching using random forest regression-voting. IEEE transactions on pattern analysis and machine intelligence **37**(9), 1862–1874 (2014)
13. Lindner, C., Thiagarajah, S., Wilkinson, J.M., Wallis, G.A., Cootes, T.F., arcOGEN Consortium, et al.: Accurate fully automatic femur segmentation in pelvic radiographs using regression voting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 353–360. Springer (2012)
14. Lindner, C., Thiagarajah, S., Wilkinson, J.M., Wallis, G.A., Cootes, T.F., arcOGEN Consortium, et al.: Fully automatic segmentation of the proximal femur using random forest regression voting. IEEE transactions on medical imaging **32**(8), 1462–1472 (2013)
15. Parrott, J., Boyd, R.N., Dobson, F., Lancaster, A., Love, S., Oates, J., Wolfe, R., Nattrass, G.R., Graham, H.K.: Hip displacement in spastic cerebral palsy: repeatability of radiologic measurement. Journal of Pediatric Orthopaedics **22**(5), 660–667 (2002)
16. Payer, C., Štern, D., Bischof, H., Urschler, M.: Regressing heatmaps for multiple landmark localization using cnns. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 230–238. Springer (2016)
17. Pham, T.T., Le, M.B., Le, L.H., Andersen, J., Lou, E.: Assessment of hip displacement in children with cerebral palsy using machine learning approach. Medical & Biological Engineering & Computing **59**(9), 1877–1887 (2021)
18. Pons, C., Rémy-Néris, O., Médée, B., Brochard, S.: Validity and reliability of radiological methods to assess proximal hip geometry in children with cerebral palsy: a systematic review. Developmental Medicine & Child Neurology **55**(12), 1089–1102 (2013)
19. Reimers, J.: The stability of the hip in children: a radiological study of the results of muscle surgery in cerebral palsy. Acta Orthopaedica Scandinavica **51**(sup184), 1–100 (1980)
20. Scrutton, D., Baird, G.: Surveillance measures of the hips of children with bilateral cerebral palsy. Archives of disease in childhood **76**(4), 381–384 (1997)
21. Urschler, M., Ebner, T., Štern, D.: Integrating geometric configuration and appearance information into a unified framework for anatomical landmark localization. Medical image analysis **43**, 23–36 (2018)
22. Wang, C.W., Huang, C.T., Lee, J.H., Li, C.H., Chang, S.W., Siao, M.J., Lai, T.M., Ibragimov, B., Vrtovec, T., Ronneberger, O., Fischer, P., Cootes, T., Lindner, C.: A benchmark for comparison of dental radiography analysis algorithms. Medical image analysis **31**, 63–76 (2016), doi:10.1016/j.media.2016.02.004
23. Wirtz, A., Mirashi, S.G., Wesarg, S.: Automatic teeth segmentation in panoramic x-ray images using a coupled shape model in combination with a neural network. In: International conference on medical image computing and computer-assisted intervention. pp. 712–719. Springer (2018)

24. Zhang, J., Liu, M., Shen, D.: Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks. IEEE Transactions on Image Processing **26**(10), 4753–4764 (2017)