

Landmark-free Assessment of Lower-limb Alignment with Implicit Neural Shape Functions from Knee Radiographs

Zhisen Hu^{*1}, Antti Kemppainen^{2,3}, David Johnson^{4,5,6}, Egor Panfilov², Huy Hoang Nguyen², Timothy Cootes¹, Claudia Lindner¹, and Aleksei Tiulpin⁷

¹ Division of Informatics, Imaging and Data Sciences, The University of Manchester, United Kingdom

² Research Unit of Health Sciences and Technology, University of Oulu, Finland

³ Medical Research Center Oulu, University of Oulu and Oulu University Hospital, Finland

⁴ Department of Trauma and Orthopaedics, Stockport NHS Foundation Trust, Stepping Hill Hospital, United Kingdom

⁵ School of Health and Society, University of Salford, United Kingdom

⁶ School of Biological Sciences, The University of Manchester, United Kingdom

⁷ Weill Cornell Medicine, Cornell University, United States

zhisen.hu@postgrad.manchester.ac.uk

Abstract. Radiographic assessment of lower-limb alignment (LLA) is important for predicting joint health and surgical outcomes in total knee arthroplasty. Traditional measurement methods are manual and time-consuming, while recent machine learning approaches typically rely on locating a fixed set of anatomical landmarks. This dependence limits flexibility and may require re-annotation when clinical definitions change. To address this, we propose an automated workflow using Implicit Neural Shape Functions (INSF). Rather than relying on explicit landmark coordinates, we encode the anatomy into a compact latent space and regress clinical alignment measurements directly from these latent codes. This architecture allows for rapid extendability to new tasks without altering the backbone representation. We trained our method on an internal dataset of 566 knee radiographs, each annotated with the outline of the femur and tibia. We evaluated it on both an internal test dataset of 50 patients and a separate external set of 402 preoperative cases from the MRKR dataset. Performance was comparable to state-of-the-art landmark-based methods and manual agreement, while offering a flexible shape representation that can be extended to additional measurement tasks. The MRKR measurement data is available at <https://huggingface.co/datasets/imedslab/mrkr-knee-alignment>.

Keywords: Lower-limb Alignment · Shape Model · Deep Learning.

*Corresponding Author

1 Introduction

Knee osteoarthritis (OA) is a common and significant health issue that heavily burdens healthcare systems [1]. Total knee replacement (TKR) may be offered as treatment for end-stage knee OA. Nevertheless, TKR is invasive, involving prosthesis implantation at the knee joint, and around 10% of patients are dissatisfied following TKR [2,3]. Pre-operative and post-operative lower-limb alignment (LLA) affects the outcomes following TKR, with radiographs revealing anomalies such as deformities of the femur and tibia, as well as incorrect positioning of the implants [4,5]. Accurate assessment of LLA in radiographs is important for successful treatment outcomes and long-term joint health. Traditional LLA measurement methods are manual and time-consuming. Machine learning-based automated techniques have now been widely used in the medical imaging area [6,7], including orthopaedics [8,9]. Such automated methods for measuring LLA in knee radiographs are potentially clinically valuable for reducing costs and improving the efficiency of the knee OA treatment pathway.

Recent machine learning approaches [10,11,12] for measuring LLA primarily rely on point-based models to predict a predefined set of anatomical landmarks [8,13], from which the angles representing LLA are subsequently computed. These landmarks must be densely and consistently annotated across images, which is both time-consuming and costly to perform. An example of generating angles such as anatomical tibio-femoral angle (aTFA), anatomical medial proximal angle (aMPTA), and joint line convergence angle (JLCA) from landmark positions is shown in Fig. 1a. While this strategy enables direct computation of established clinical metrics, it inherently constrains the model to a discrete and predefined geometric representation of anatomy.

Deep implicit shape representations [14] have been introduced to model shapes using a compact latent space and a signed distance function (SDF) auto-decoder. DISSM developed by Raju *et al.* [15] adopts this auto-decoder idea and proposes a complete workflow for modelling 3D liver and larynx shapes from CT images. In the orthopaedic field, for example, Pai *et al.* [16] employed neural shape models to quantify bone shape parameters from knee MRI.

In this study, we leverage Implicit Neural Shape Functions (INSF) to model bone morphology from X-ray images using a SIREN-based auto-decoder framework [17]. Instead of relying on explicit anatomical landmarks, the proposed method encodes bone geometry into a compact latent space and directly regresses LLA measurements from these learned representations using a multilayer perceptron (MLP). This continuous shape representation captures global anatomical structure and enables flexible adaptation to new clinical tasks without requiring predefined landmark definitions. To model multiple anatomical structures, the implicit representation is structured with one output channel per bone, allowing independent yet coordinated encoding of femoral and tibial geometry. The method was validated on both internal and external datasets (Fig. 1b). To our knowledge, this is the first application of INSF to knee radiographs and the first use of implicit neural representations for automated LLA assessment.

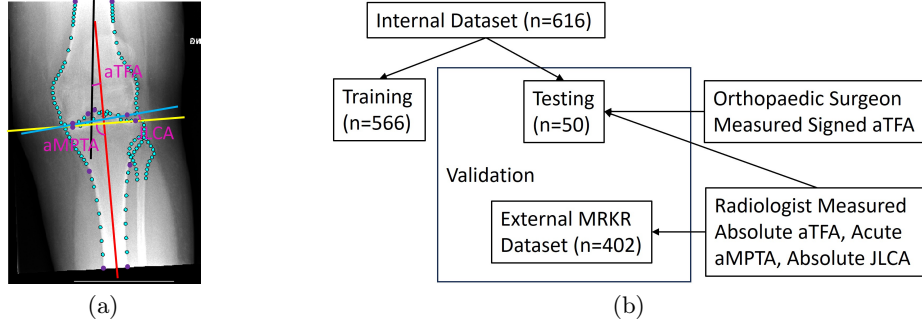


Fig. 1. (a) Angles derived from landmarks. In landmark-based approaches, landmark positions (purple points) are used to fit several lines which form the angles. Black: anatomical femoral axis; Red: anatomical tibial axis; Blue: femoral joint line; Yellow: tibial joint line. aTFA is formed by the anatomical axes of femur (black) and tibia (red). aMPTA is formed by the anatomical tibial axis (red) and the tibial joint line (yellow), usually measured on the medial side. JLCA is formed by the femoral (blue) and tibial (yellow) joint lines. (b) Our study design. We trained our model with the internal training set and validated it using both internal and external testing sets.

2 Methodology

2.1 Overview

The overall inference workflow of our automated LLA assessment framework is illustrated in Fig. 2. ① We first align all the shapes using two reference points at the corners of the tibial plateau. ② Then, we predict the SDFs from the aligned images. The predicted SDF contains two channels, with separate signed distance maps for the outlines of femur and tibia. ③ Each testing sample is then directly fitted to the learned latent shape space via test-time optimisation. ④ Finally, we directly regress the angular measurements with a compact 4-layer neural network (MLP).

2.2 Data

Image Data. Our internal image data consists of anonymised standard anteroposterior (AP) unilateral knee radiographs from patients undergoing TKR. To standardise anatomical orientation, all right knee images were horizontally flipped to appear as left knees. The radiographs were retrospectively collected from Stockport NHS Foundation Trust (approved by the Health Research Authority, IRAS 244130). This internal dataset includes 566 pre-operative images for training and 50 patients for testing.

Our external dataset includes 402 knees, comprising both bilateral and unilateral images randomly selected from the Emory Knee Radiograph (MRKR) Dataset [18]. To ensure consistency with the internal data, bilateral images were split, and all right knee radiographs were flipped to maintain uniform orientation.

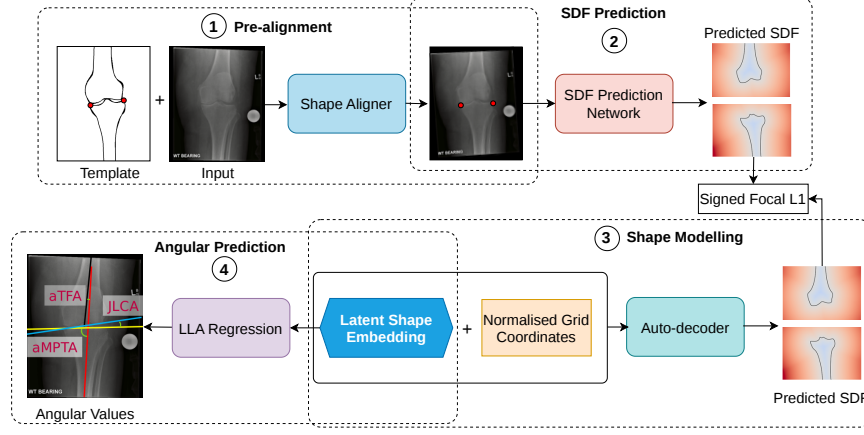


Fig. 2. Overview of test-time shape embedding optimisation and LLA regression. Only latent shape embedding is optimised during test time.

LLA Angular Data. For training, ground-truth angles were derived from manually annotated landmarks. For internal testing, aTFA was independently measured by an orthopaedic surgeon and a radiologist. Each clinician performed two measurements 7–10 days apart, with the second blinded to the first; the mean of the two measurements was used as the final ground truth. The external evaluation included 402 cases from the Emory Knee Radiograph (MRKR) Dataset [18]. Ground-truth measurements were obtained by the same radiologist. The radiologist recorded absolute aTFA and JLCA and measured the acute aMPTA.

2.3 CNN-based SDF Prediction

Following [11,12], we first trained an Hourglass-based [19,13] CNN to detect two reference points from the image for aligning all images into the canonical space. Subsequently, a U-Net [20] was trained to predict the canonical SDFs from the aligned knee images.

As our dataset consists of knee radiographs centered on the knee joint, portions of the femoral or tibial shafts are occasionally truncated before reaching the image boundaries. It has been shown that incorporating a greater extent of shaft anatomy in the model improves the accuracy of aTFA measurements [10]. To preserve anatomically consistent bone geometry and incorporate more shaft anatomy during training, we manually extended truncated shaft contours toward the image margins along estimated shaft tangents. This preprocessing step for the training data enables the U-Net model to learn consistent shaft geometry and to infer anatomically plausible shaft extensions toward the image boundaries at test time.

2.4 Implicit Shape Representation Learning

We adopt SIREN [17] as an auto-decoder to learn compact latent shape representations and to model bone geometries using SDFs, which output the distance from a queried spatial coordinate to the shape surface:

$$f_\theta(x, z) = d : x \in \mathbb{R}^2, d \in [-1, 1] \quad (1)$$

where θ is the network weights, z is the latent shape code, x is the input 2D normalised coordinates, the signed distance value d is negative inside the shape and positive outside, with the surface defined by the zero level set ($d = 0$).

Following the auto-decoder paradigm in [14,15], we construct K aligned canonical SDF training samples, $\{\mathcal{X}_k\}_{k=1}^K$, and associate each sample with a corresponding latent vector, $\{z_k\}_{k=1}^K$. The MLP decoder takes both spatial coordinates and the latent vector as input. During training, the network parameters and the latent shape embedding are optimised jointly. To better capture high-frequency geometric details, such as highly curved surface, we employ the Signed Focal L1 loss proposed in [21]. The overall optimisation objective is:

$$\arg \min_{\theta, z} \sum_{k=1}^K \left(\sum_{i=1}^{|\mathcal{X}_k|} \frac{1}{|\Omega|} \sum_{i \in \Omega} |S_i - P_{i,k}| \frac{|S_i - P_{i,k}|^\gamma \mathbb{I}(S_i P_{i,k} \geq 0)}{\max(|S_i|, |P_{i,k}|) + \epsilon} \right) \quad (2)$$

where $P_{i,k}$ includes the use of the tanh function to convert the network output into the range $[-1, 1]$ ($P_{i,k} = \tanh(f_\theta(x_i, z_k))$), S is the ground-truth SDF value, ϵ is a positive constant to avoid numerical issues, Ω denotes the set of spatial coordinates sampled in the image domain for SDF supervision, γ is a positive hyperparameter, and $\mathbb{I}(\cdot)$ is the indicator function. The summations are nested: outer over training samples k , inner averaging focal-L1 loss over sampled coordinates.

2.5 Test-time Optimisation and LLA Regression

Prior work typically fits each sample to the implicit shape model by training an additional network to predict the latent code [15]. We use a simpler approach. We use a U-Net to predict the SDF for each image. We then perform an optimisation to find the latent vector, z , which minimises the focal loss between the generated SDF and that from the U-Net.

Unlike traditional methods for LLA assessment, which explicitly compute geometric measurements from localised anatomical landmarks, we directly regress the angular values that quantify LLA using a simple MLP. The MLP takes the latent shape codes as input and outputs the corresponding angle measurements.

3 Results

Shape Modelling. We constructed a latent shape space represented by 128-dimensional vectors. To investigate the relationship between each latent dimension and LLA, we computed the Pearson correlation coefficient ($|r|$) between the

latent codes and the angular values in the training set. The dimensions were then ranked according to their absolute correlation coefficients. All three angles exhibited at least moderate correlations ($|r| > 0.6$) with their respective most strongly associated latent dimensions. Notably, signed aTFA and JLCA demonstrated strong correlations ($|r| \approx 0.8$) with their most correlated latent dimensions (Fig. 3a and 3c). Latent dimension 116 exhibited the strongest correlation with both aTFA and JLCA, whereas latent dimension 112 was most strongly associated with aMPTA. The latent dimensions are learned in an unsupervised manner and are not explicitly constrained to correspond to predefined anatomical variables. Consequently, the numerical indices (e.g., 116 and 112) have no inherent meaning and are not fixed: the identity and ordering of the dimensions can change across training runs. A given dimension becomes interpretable only through its observed statistical association with the angular measurements.

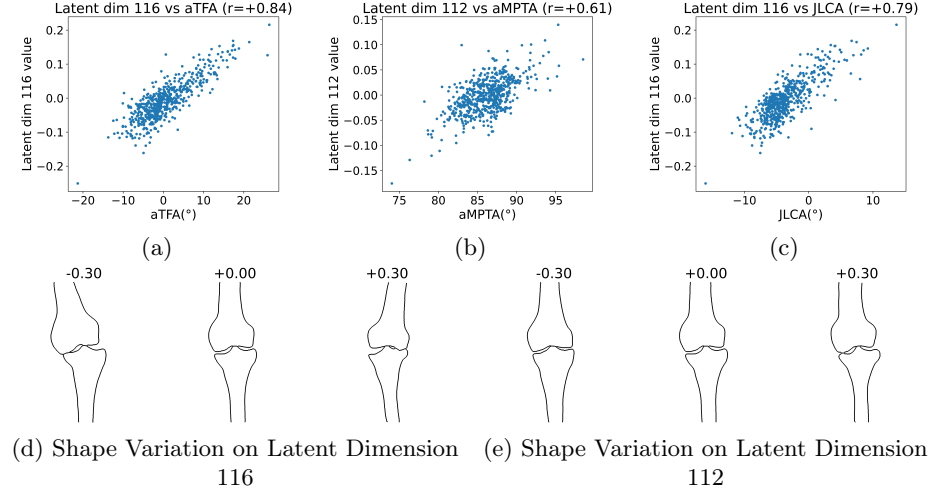


Fig. 3. (a-c) Correlation analysis between selected latent dimensions and angles, and (d-e) visualisation of shape variation on the most correlated dimensions.

We further analysed how variations on the most strongly correlated dimension of each angle (116, 112) affected the shape contours, similar to visualisations in point-based shape models [22,23]. We visualised the mean shape and systematically varied the corresponding latent parameter from -0.3 to 0.3 while keeping all other dimensions fixed (Fig. 3d and 3e). For latent dimension 116, increasing the parameter produced progressive rightward shifts of the femoral and tibial shafts, indicating that the learned shape model automatically captures pose-related information that influences aTFA and JLCA, thereby affecting LLA.

LLA Assessment. Intraclass correlation coefficient (ICC) and mean absolute difference (MAD) were used to evaluate the agreement between the measure-

ments assessed in various ways. We first compared our approach with the methods proposed in [10,12], where the anatomical landmark positions of the knee joint are predicted first and angular measurements are subsequently derived from selected landmark subsets. We also compared our method against intra- and inter-rater agreements, using the aTFA measurements obtained by both clinicians on the internal dataset ($n = 50$). When comparing automated methods with the ground-truth aTFA values measured by the orthopaedic surgeon, our method achieved comparable results ($MAD = 1.2^\circ$; $ICC = 0.97$) with both [10] ($MAD = 1.2^\circ$; $ICC = 0.97$) and [12] ($MAD = 1.1^\circ$; $ICC = 0.97$). However, all automated methods performed marginally worse than the intra-rater agreement. For comparison with ground-truth aTFA values measured by the radiologist, automated methods were comparable to intra- or inter-rater agreements ($MAD \approx 1.0^\circ$; $ICC \approx 0.95$).

When stratified by shaft length, our auto-shaft extension approach outperformed the same method without shaft extension for aTFA measurement. Specifically, when compared with the orthopaedic surgeon’s measurements, the MAD decreased from 1.8° to 1.2° , and the ICC improved from 0.94 to 0.97. Similarly, when compared with the radiologist’s measurements, the MAD decreased from 1.6° to 1.0° , and the ICC increased from 0.87 to 0.94. These findings demonstrate that incorporating shaft extension enhances both measurement accuracy and agreement with expert assessments.

We additionally evaluated our method on a subset ($n = 402$) of the external MRKR dataset [18], comparing it with the approach in [12] and manual intra-rater agreement. In addition to aTFA, we included aMPTA and JLCA in the generalisation experiment to assess extendability. On this dataset, our method achieved performance comparable to [12] for aTFA, with no statistically significant difference observed. For the other angles, our method showed marginally lower numerical performance. Both automated approaches underperformed relative to manual intra-rater agreement.

The comparison of our method, [12], along with the intra-rater agreement, is presented in Table 1. Statistical testing was performed to determine whether the MAD exceeded $\pm 1^\circ$, using a two one-sided tests scheme with Bonferroni–Holm adjustment. All p -values were < 0.0001 and are therefore not reported. Overall, the statistical analysis indicates that the evaluated methods perform equivalently within a $\pm 1^\circ$ difference threshold.

4 Discussions and Conclusions

In this paper, we have shown that implicit neural representations capture variations associated with LLA. The proposed LLA assessment workflow achieved performance comparable to landmark-based methods and to manual intra- or inter-rater agreement on the internal dataset, but declined slightly on the external dataset, where it was slightly inferior to the landmark-based method and manual agreement. This reduction in generalisability may reflect population differences. The internal dataset predominantly comprised White patients, whereas

Table 1. Agreement analysis: mean absolute difference (MAD; °) and intraclass correlation coefficient (ICC) for internal and external datasets. Statistical testing was done to verify if the MAD exceeds $\pm 1^\circ$. All p -values were < 0.0001 . Here, OS indicates orthopaedic surgeon and RA indicates board-certified radiologist.

Dataset	#	Angle	Method	MAD (°)	ICC
Internal	50	aTFA	Ours vs. OS	1.19 _(0.92, 1.54)	0.97 _(0.94, 0.98)
			Hu <i>et al.</i> [12] vs. OS	1.09 _(0.84, 1.40)	0.97 _(0.95, 0.98)
			Ours vs. RA	1.03 _(0.80, 1.32)	0.94 _(0.90, 0.96)
			Hu <i>et al.</i> [12] vs. RA	0.95 _(0.74, 1.22)	0.96 _(0.92, 0.97)
			OS (intra-rater)	0.89 _(0.69, 1.14)	0.99 _(0.98, 0.99)
			RA (intra-rater)	0.94 _(0.73, 1.20)	0.95 _(0.92, 0.97)
			OS vs. RA (inter-rater)	0.97 _(0.75, 1.26)	0.95 _(0.90, 0.97)
External	402	aTFA	Ours	1.08 _(0.98, 1.19)	0.84 _(0.81, 0.87)
			Hu <i>et al.</i> [12]	0.91 _(0.83, 1.01)	0.87 _(0.85, 0.90)
			RA (intra-rater)	0.65 _(0.59, 0.72)	0.93 _(0.92, 0.94)
		aMPTA	Ours	1.29 _(1.17, 1.42)	0.58 _(0.30, 0.74)
			Hu <i>et al.</i> [12]	0.88 _(0.80, 0.97)	0.72 _(0.64, 0.78)
			RA (intra-rater)	0.77 _(0.70, 0.85)	0.79 _(0.76, 0.83)
		JLCA	Ours	0.94 _(0.85, 1.03)	0.67 _(0.61, 0.72)
			Hu <i>et al.</i> [12]	0.75 _(0.68, 0.83)	0.80 _(0.72, 0.86)
			RA (intra-rater)	0.49 _(0.45, 0.54)	0.89 _(0.86, 0.91)

the external MRKR dataset [18] included approximately 40% White and 40% African American patients. In addition, the mean age of the internal dataset was approximately ten years higher than that of the MRKR cohort. These findings suggest that further refinement may improve robustness across diverse populations.

Performance differences were generally small between methods, and variability in clinician annotations and measurement protocols may have influenced the results. We observed no significant differences between automated approaches, as performance is inherently dependent on the underlying clinical annotations. Despite minor performance discrepancies, none of the observed effects was statistically or clinically significant. The methods were found equivalent to the ground truth measurements, as well as to each other ($p < 0.0001$).

When it comes to the interpretability of the latent space, its correlations with aTFA or JLCA were strong, whereas the association with aMPTA was comparatively weaker. This may reflect the model’s more accurate representation of relative bone positioning rather than displacement of individual bones.

A key limitation of this study is that we utilised the pre-processing step from [11,13] for shape pre-alignment out of convenience. This, however, can easily be incorporated within one UNet model. In general, our training targets were derived from landmark-based annotations, which may inherently favor landmark-driven approaches and constrain the evaluation of the full advantages offered by

a landmark-free representation. Future work should focus on directly regressing measurements that cannot be readily derived from landmark positions. This would enable a more comprehensive assessment of the proposed approach and more convincingly demonstrate its extendability.

Nevertheless, our framework offers important advantages. By encoding global anatomical shape information into latent representations, it learns LLA-related features without predefined landmarks. This enables compact shape representation in latent vectors and facilitates extension to new clinical tasks. Such flexibility is particularly valuable when landmarks are difficult to detect or inconsistently defined, making the framework well suited to evolving clinical definitions where traditional point-based models are impractical.

Acknowledgments. ZH is funded by European Laboratory for Learning and Intelligent Systems (ELLIS) Unit Manchester. CL is funded by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (223267/Z/21/Z). This work was supported by the European Union’s Horizon Programme under projects STAGE (grant 101137146; HHN) and UltraSense (grant 101130192; EP and HHN), as well as by the Research Council of Finland (project 336449; EP and AT).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Steinmetz, J.D., Culbreth, G.T., Haile, L.M., Rafferty, Q., Lo, J., Fukutaki, K.G., Cruz, J.A., Smith, A.E., Vollset, S.E., Brooks, P.M. and Cross, M.: Global, regional, and national burden of osteoarthritis, 1990–2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet Rheumatology* **5**(9), e508–e522 (2023)
2. Özden, V.E., Osman, W.S., Morii, T., Pastor, J.C.M., Abdelaal, A.M. and Younis, A.S.: What percentage of patients are dissatisfied post-primary total hip and total knee arthroplasty?. *The Journal of Arthroplasty* **40**(2), S55–S56 (2025)
3. DeFrance, M.J. and Scuderi, G.R.: Are 20% of patients actually dissatisfied following total knee arthroplasty? A systematic review of the literature. *The Journal of Arthroplasty* **38**(3), 594–599 (2023)
4. Ritter, M.A., Davis, K.E., Meding, J.B., Pierson, J.L., Berend, M.E. and Malinzak, R.A.: The effect of alignment and BMI on failure of total knee replacement. *JBJS* **93**(17), 1588–1596 (2011)
5. Ritter, M.A., Davis, K.E., Davis, P., Farris, A., Malinzak, R.A., Berend, M.E. and Meding, J.B.: Preoperative malalignment increases risk of failure after total knee arthroplasty. *JBJS* **95**(2), 126–131 (2013)
6. Zhang, Z., Keles, E., Durak, G., Taktak, Y., Susladkar, O., Gorade, V., Jha, D., Ormeci, A., Medetalibeyoglu, A., Yao, L. and others: Large-scale multi-center CT and MRI segmentation of pancreas with deep learning. *Medical image analysis* **99**, 103328 (2025)
7. Lin, Y., Wang, L., Hagemann, I.S., Kuroki, L.M., Sanders, B.E., Hagemann, A.R., Siegel, C., Powell, M.A. and Zhu, Q.: Vascular graph network for ovarian lesion classification using optical-resolution photoacoustic microscopy. *Photoacoustics* 100794 (2025)

8. Lindner, C., Thiagarajah, S., Wilkinson, J.M., arcOGEN Consortium, Wallis, G.A. and Cootes, T.F.: Accurate bone segmentation in 2D radiographs using fully automatic shape model matching based on regression-voting. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 181–189. Springer Berlin Heidelberg (2013).
9. Tiulpin, A., Thevenot, J., Rahtu, E., Lehenkari, P. and Saarakkala, S.: Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. *Scientific reports* **8**(1), 1727 (2018)
10. Cullen, D., Thompson, P., Johnson, D. and Lindner, C.: An AI-based system for fully automated knee alignment assessment in standard knee AP radiographs. *The Knee* **54**, 99–110 (2025)
11. Hu, Z., Cullen, D., Thompson, P., Johnson, D., Tiulpin, A., Cootes, T.F. and Lindner, C.: Automated measurements of knee alignment with deep learning: Accuracy and reliability. *Osteoarthritis and Cartilage* **33**, S100–S101 (2025)
12. Hu, Z., Cullen, D., Thompson, P., Johnson, D., Bian, C., Tiulpin, A., Cootes, T.F. and Lindner, C.: Deep learning-based alignment measurement in knee radiographs. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 121–130. Cham: Springer Nature Switzerland (2025).
13. Tiulpin, A., Melekhov, I. and Saarakkala, S.: KNEEL: Knee anatomical landmark localization using hourglass networks. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 352–361. IEEE (2019).
14. Park, J.J., Florence, P., Straub, J., Newcombe, R. and Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 165–174. (2019).
15. Raju, A., Miao, S., Jin, D., Lu, L., Huang, J. and Harrison, A.P.: Deep implicit statistical shape models for 3d medical image delineation. In *proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 2, pp. 2135–2143. (2022).
16. Pai, S.A., Black, M., Young, K., Sherman, S., Chu, C., Williams, A., Gold, G., Kogan, F., Hargreaves, B., Chaudhari, A. and Gatti, A.: Neural Shape model quantifies early and progressive bone shape changes after aclr. *Osteoarthritis Imaging* **5**, 100342 (2025)
17. Sitzmann, V., Martel, J., Bergman, A., Lindell, D. and Wetzstein, G.: Implicit neural representations with periodic activation functions. *Advances in neural information processing systems* **33**, 7462–7473 (2020)
18. Price, B., Adleberg, J., Thomas, K., Zaiman, Z., Mansuri, A., Brown-Mulry, B., Okecheukwu, C., Gichoya, J. and Trivedi, H.: Emory Knee Radiograph (MRKR) Dataset. *arXiv preprint arXiv:2411.00866*. (2024)
19. Newell, A., Yang, K. and Deng, J.: Stacked hourglass networks for human pose estimation. *arXiv preprint arXiv:1603.06937*. (2016)
20. Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv preprint arXiv:1505.04597*. (2015)
21. Dang, T., Nguyen, H.H. and Tiulpin, A.: Singr: Brain tumor segmentation via signed normalized geodesic transform regression. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 593–603. Cham: Springer Nature Switzerland (2024).
22. Cootes, T.F., Taylor, C.J., Cooper, D.H. and Graham, J.: Active shape models—their training and application. *Computer vision and image understanding* **61**(1), 38–59 (1995)

23. Cootes, T.F., Ionita, M.C., Lindner, C. and Sauer, P.: Robust and accurate shape model fitting using random forest regression voting. In European conference on computer vision, pp. 278–291. Springer Berlin Heidelberg (2012).