



Predicting hip-knee-ankle and femorotibial angles from knee radiographs with deep learning



Jinhong Wang^{a,1}, Thomas A.G. Hall^{a,1}, Omar Musbahi^b, Gareth G. Jones^b, Richard J. van Arkel^{a,*}

^aDepartment of Mechanical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom

^bDepartment of Surgery and Cancer, Imperial College London, White City Campus, London W12 0BZ, United Kingdom

ARTICLE INFO

Article history:

Received 7 October 2022

Revised 25 January 2023

Accepted 9 March 2023

Keywords:

X-ray

Knee Angle

Surgical planning

Artificial Intelligence

Neural Network

Mechanical Alignment

ABSTRACT

Background: Knee alignment affects the development and surgical treatment of knee osteoarthritis. Automating femorotibial angle (FTA) and hip-knee-ankle angle (HKA) measurement from radiographs could improve reliability and save time. Further, if HKA could be predicted from knee-only radiographs then radiation exposure could be reduced and the need for specialist equipment and personnel avoided. The aim of this research was to assess if deep learning methods could predict FTA and HKA angle from posteroanterior (PA) knee radiographs.

Methods: Convolutional neural networks with densely connected final layers were trained to analyse PA knee radiographs from the Osteoarthritis Initiative (OAI) database. The FTA dataset with 6149 radiographs and HKA dataset with 2351 radiographs were split into training, validation, and test datasets in a 70:15:15 ratio. Separate models were developed for the prediction of FTA and HKA and their accuracy was quantified using mean squared error as loss function. Heat maps were used to identify the anatomical features within each image that most contributed to the predicted angles.

Results: High accuracy was achieved for both FTA (mean absolute error 0.8°) and HKA (mean absolute error 1.7°). Heat maps for both models were concentrated on the knee anatomy and could prove a valuable tool for assessing prediction reliability in clinical application.

Conclusion: Deep learning techniques enable fast, reliable and accurate predictions of both FTA and HKA from plain knee radiographs and could lead to cost savings for healthcare providers and reduced radiation exposure for patients.

© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Knee osteoarthritis is a degenerative disease that affects 16% of the global population [1], causing joint pain, stiffness, and reduced range of motion. Knee alignment is highly relevant to the development, progression and distribution of knee osteoarthritis [2–4], and for knee arthroplasty procedures which provide a treatment option for end-stage disease [5–7]. Globally, more than 3.2 million knee arthroplasties are performed each year [8] and for each of these procedures, a knee radiograph is taken to measure coronal knee alignment and plan the procedure.

* Corresponding author.

E-mail address: r.vanarkel@imperial.ac.uk (R.J. van Arkel).

¹ These authors contributed equally to this work.

There are two common measurements of knee angle taken from radiographs: hip-knee-ankle (HKA) and femorotibial angle (FTA). HKA is defined as the angle between the mechanical axes of the femur and the tibia [9]; its measurement requires a full-limb radiograph to identify the hip, knee and ankle joint centres. HKA is the gold standard measure of coronal plane knee alignment as it relates to the load distribution within the knee joint. However, its measurement requires more expensive equipment and higher radiation exposure (15–50 millirems) for full-limb radiographs versus localised knee radiographs (3 millirems) [10]. FTA – the angle between the femoral and tibial anatomical axes [9] – is moderately correlated to HKA [10] and can be taken from a knee-only posterior-anterior (PA) radiograph and is therefore often used as a surrogate measure. Whilst HKA is the better measure, many centres choose to use only FTA for preoperative planning or postoperative follow-up as it is cheaper and more convenient. Enabling HKA measurement from knee-only radiographs would thus be of great clinical use [11]: producing the gold-standard measure with reduced cost and radiation exposure, providing additional information for knee arthroplasty planning and follow-up for surgeons who do not routinely have access to full-limb radiographs.

State-of-the-art for prediction of HKA from knee-only radiographs is based on algorithmic approaches that relate to a clinical understanding of knee anatomy [10–12]. An alternative approach could be to leverage recent developments in convolutional neural networks (CNNs) and growth in computing power [13]. Deep learning methods have shown early promise in measuring HKA on full-limb radiographs [14–16] and in analysing PA knee radiographs for other applications [17–19]. The scope for impact is exciting, for example recent research has demonstrated that neural networks are able to identify arthroplasty implants from plain radiographs with 100% accuracy [20] and grade knee osteoarthritis severity as accurately as a fellowship-trained arthroplasty surgeon [21]. Therefore, the aim of this study was to investigate whether CNNs could be applied to automatically measure FTA and HKA from knee-only PA radiographs.

2. Materials and methods

Two deep learning models to measure knee alignment from knee-only posterior-anterior radiographs were developed: the first measured FTA and the second predicted HKA. Clinically reported measurements from the Osteoarthritis Initiative (OAI) were used as ground truth data for training and evaluation.

2.1. Data

Bilateral knee radiographs with the X-ray beam passing posterior to anterior (PA) were sourced from the baseline dataset of the OAI: a mixed-sex, multi-centre, ten-year observational study. The dataset contained 4,795 bilateral knee radiographs and 2,723 full-limb radiographs. FTA measurements – calculated from knee radiographs using a landmark-based method [12] – were supplied as part of the dataset but were not available for radiographs with <10 cm of tibial shaft visible. HKA was also included in the dataset, having been measured from full-limb radiographs using OASYS Horizon Surveyor software [22–24]. All images that had a corresponding knee alignment measurement were used, including those for patients with knee deformity, a history of prior surgery (trauma, replacement, etc.) on or around the knee joint, and those where clinicians might perceive the image quality to be low. The bilateral radiographs were split into unilateral images, and right knees were mirrored with no further pre-processing or annotation. The FTA and HKA datasets were divided in a ratio of 70:15:15 into training, validation, and test sets. Ultimately, 6,149 unilateral radiographs were used to train the FTA prediction models and 2,351 were used for HKA training. There was no patient crossover between datasets.

2.2. Architecture

Separate convolutional models of identical architecture (Figure 1) were developed to predict FTA and HKA from knee radiographs. Two commonly used CNNs were investigated as the backbone of the model: DenseNet-121 [25] and Inception-ResNet V2 [26]. Outputs of the base CNN were flattened using global average pooling before being fed through two densely connected layers (1024 and 512 neurons with ReLU activation functions) and into the final regression layer.

2.3. Training

The Adam optimizer [27] was used to train the models with mean squared error (MSE) between the predicted and OAI-reported knee angle as a loss function. Ground truth data for training the FTA model were measured from the same knee-only image using a semi-automatic landmark-based method [12]. Ground truth data for HKA were measured on separate full-limb radiographs using OASYS Horizon Surveyor Software. Networks were initialised with base CNN weights pretrained on the ImageNet classification dataset [28] and randomly initialised final layer weights. To encourage the model to learn from the knee anatomy rather than the protocols of image acquisition, images were randomly rotated ($\pm 10^\circ$) and cropped (85–100%) during training. Models were trained on NVIDIA Quadro RTX 6000 GPUs in batch sizes of 64 for 100 epochs. The model was evaluated on the validation set after each epoch, with the best performing model selected for testing.

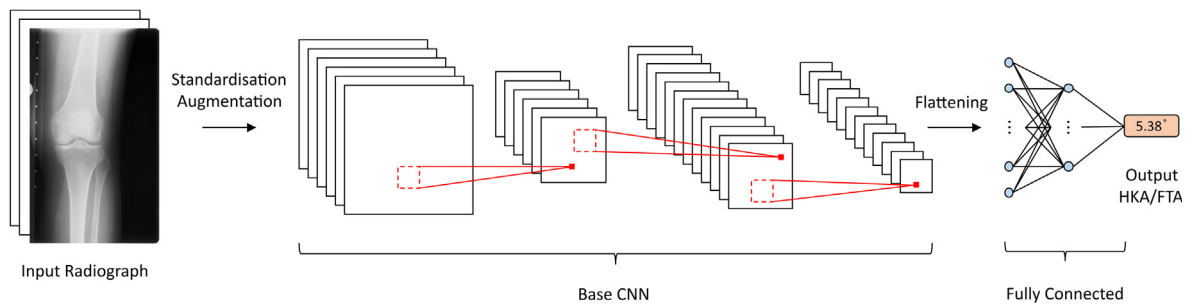


Figure 1. The architecture for predicting alignment angles (HKA/FTA) from knee radiographs consisted of a base convolutional neural network (DenseNet/Inception-ResNet-V2) and two fully connected final layers (1024 and 512 neurons).

2.4. Evaluation

The accuracy of the FTA and HKA models were evaluated with respect to the clinically reported OAI measurements using mean square error (MSE), mean absolute error (MAE), and root mean square error (RMSE). Additionally, the percentage of predictions within 3° and 1° of the OAI-reported angles was evaluated as a clinically relevant performance metric. Intraclass correlation coefficients (ICC) – two-way mixed, single measures for absolute agreement – between the predicted angles for each model and the OAI-reported angles were computed in SPSS Statistics software as a measure of reliability. Graphically, accuracy was evaluated with density plots of the predicted values of the testing dataset against the reported values. Bland-Altman plots were also plotted for direct comparison to the state-of-art semi-automatic methods of calculating FTA and HKA from knee radiographs [12] with 95% limits of agreement (95% LoA).

2.5. Attention analysis

Attention heat maps were used to identify the focus of the models during inference (Figure 2). Blocks of the input radiographs were systematically occluded [29] with a black square and passed through the model, whereupon a new prediction of knee alignment was generated. The predicted alignment for each occluded image, y_{ij} , was compared to the original prediction on the unoccluded image, \tilde{y} , to generate a sensitivity matrix: $\epsilon_{ij} = |y_{ij} - \tilde{y}|$, which was then normalised (min-max) for each image. Large deviations in the predicted alignment indicate that the occluded region was significant to the model's prediction, thereby allowing its focus to be determined when superposed onto the original radiograph.

3. Results

For both FTA and HKA, the best performing models were built upon the DenseNet architecture (Table 1, Figure 3), with FTA being better predicted than HKA – as would be expected for local PA radiographs. The best performing computer-assisted approach [12] identified in literature, which was used to generate the ground-truth FTA measures in this study, reported a 95% limit of agreement of 1.1° in an inter-reader variability study. The 95% limit of agreement between our fully automatic FTA predictions and those of the reference method was worse (1.9°) but comfortably within the 3° threshold for clinical use (Table 1, Figure 4).

HKA predictions were correlated (ICC: 0.80, Figure 3) to the clinically reported HKA measures taken from full-limb radiographs with 83.9% of predictions within a 3° difference. The 95% limit of agreement between the predictions and the reference measures was 4.3° , which outperforms the state-of-the-art semi-automatic method for HKA prediction [11] by 0.1° (Figure 4).

Heat maps representative of excellent FTA prediction focused on the femoral and tibial condyles with moderate attention on the tibial spine and a distinguishable gap between the femoral and tibial foci (Figure 5), which corresponds to the anatomical landmarks used to estimate FTA in the reference method [12]. By contrast, heat maps with large absolute prediction errors ($>3^\circ$) only focused on the medial side of the tibia. Attention during highly accurate HKA predictions also focused on the knee's bony anatomy, albeit more broadly, with the lateral focus being closer towards the trochlear groove than it was during FTA predictions. Conversely in poor HKA predictions, the model did not establish distinct bilateral femoral/tibial foci and focused on soft tissue features around the knee. Interestingly, the model proved capable of predicting alignment for both native knees and knees with a history of prior surgery (Figure 5).

4. Discussion

This is the first study to our knowledge to present a fully automatic method for measuring coronal knee alignment using knee-only PA radiographs, demonstrating high accuracy for FTA (MAE: 0.8°) and outperforming state-of-the-art [11] for HKA

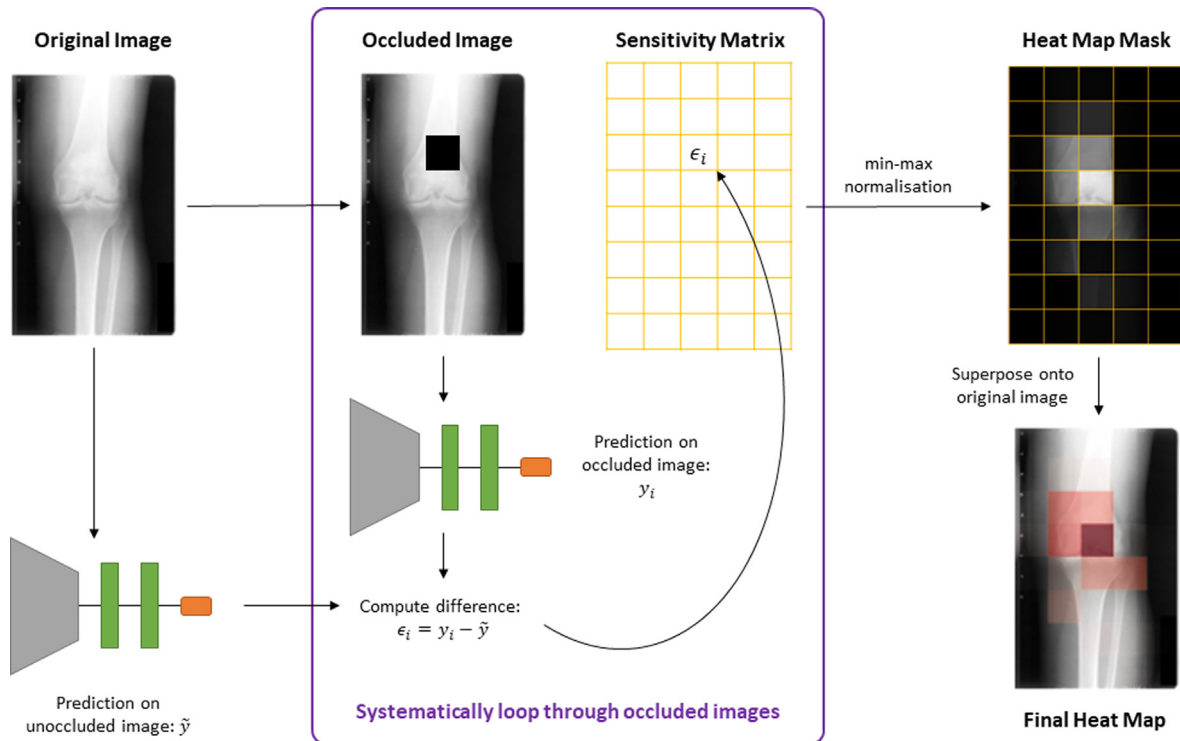


Figure 2. The attention of the model during angle prediction was determined by systematically occluding regions of the original image and recording sensitivity.

Table 1

Summary of performance for trained neural networks on testing dataset with comparison to the state-of-the-art computer-assisted methods.

Angle	Base Models	MSE (° ²)	MAE (°)	RMSE (°)	<3° Difference (%)	<1° Difference (%)	ICC (-)	95% Limit of Agreement (°)
FTA	Inception-ResNet	1.32	0.89	1.15	98.5	66.1	0.90	2.2
	DenseNet	1.19	0.82	1.09	98.7	69.5	0.91	1.9
	Iranpour-Boroujeni et al. (2014)	-	-	0.37	100.0	-	0.98	1.1
HKA	Inception-ResNet	6.87	2.02	2.62	77.7	30.5	0.70	5.1
	DenseNet	5.06	1.73	2.25	83.9	37.9	0.80	4.3
	Gielis et al. (2020)	-	1.8	-	-	-	0.90	4.4

MSE: Mean Squared Error; MAE: Mean Absolute Error; RMSE: Root Mean Square Error; ICC: Intraclass Correlation Coefficient.

(MAE: 1.7°). This finding is particularly impressive considering no restrictions were imposed on the training dataset and no anatomical annotation was required. Using deep learning, the model was able to predict alignment for knees in various conditions: native, osteotomised, partially and totally replaced (Figure 5). FTA prediction was more accurate than HKA prediction: an intuitive finding given that all the information required to calculate FTA is contained within the knee-only radiograph. Further, the state-of-the-art method used to provide reference measures for FTA [12] avoids any scan-rescan reproducibility errors. The 95% limit of agreement for our method was higher than its inter-reader variability (1.9° vs. 1.1°), which is to be expected as the model cannot outperform its input data.

Historically, an accuracy of 3° in HKA measurement has been regarded as sufficient for clinical use in orthopaedic surgery [30], though recent advances in computer-assisted and robotic-assisted surgery have prompted a drive for more accurate surgery. Contemporary robotic systems have been reported to achieve alignments of 89–100% within 3° of the surgical plan [31–35], 73% within 2° [31], and 75–78% within 1° [33,35]. On the testing set, 83.9% of automatic HKA predictions were within 3° of the reference measurement and the mean absolute difference was 1.7°. The reference measures were acquired on separate full-limb radiographs and were therefore susceptible to two sources of error: intra/inter-reader variability and scan-rescan reproducibility, which limit the achievable accuracy. In an analysis of inter-reader variability, the reference method exhibited a mean difference of 0.1° with a 95% limit of agreement of ±1.2° [24]. For reference, a deep learning model for HKA prediction on OAI full-limb radiographs, wherein the hip-knee-ankle centres are visible, achieved mean difference and 95% limit of agreement results of 0.09° ± 0.73° [14]. Data on the scan-rescan reproducibility of HKA measurement is

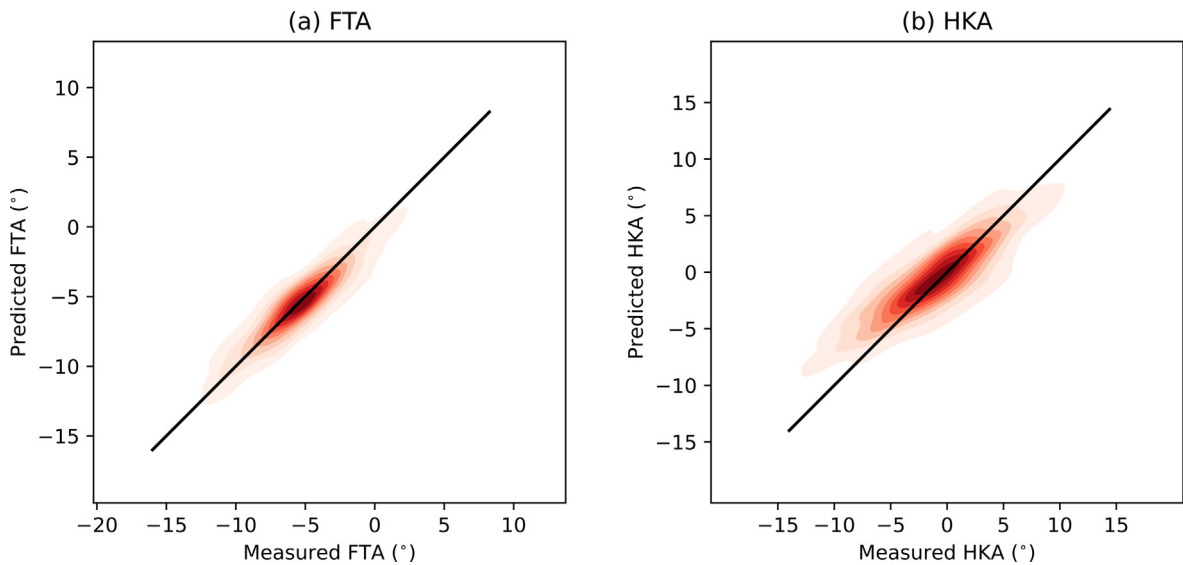


Figure 3. Results on the test dataset for the DenseNet architecture demonstrate a strong correlation between predicted and OAI-reported values for (a) FTA (ICC: 0.91) and (b) HKA (ICC: 0.80), where the intensity of red shading indicates the density of points.

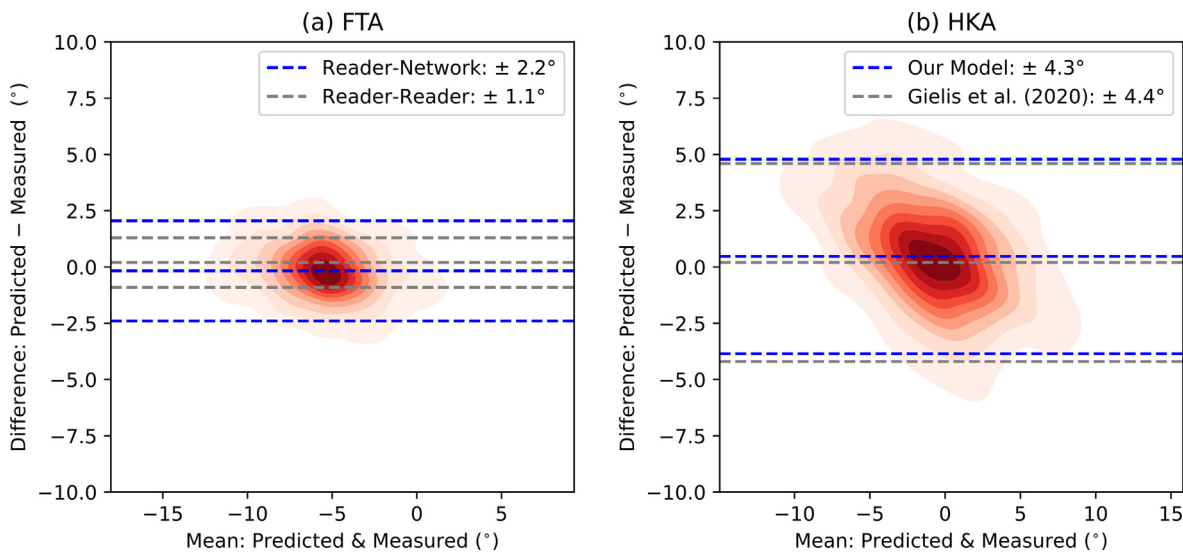


Figure 4. Bland-Altman plots of neural network prediction reliability versus OAI-reported measures for the DenseNet architecture: (a) FTA prediction with comparison to state-of-the-art manual measurement ($\pm 1.1^\circ$ in Iranpour-Boroujeni et al., 2014) and (b) HKA with comparison to state-of-the-art semi-automatic prediction ($\pm 4.4^\circ$ in Gielis et al., 2020). The intensity of red shading indicates point density.

scarce, though a small eight-patient study observed scan-to-scan differences of 0–2° (mean: 1.3°) in HKA measures [30]. Such magnitudes of error are similar to the mean absolute difference for our model (1.7°).

Whilst HKA measurement from full-limb radiographs remains the gold-standard measurement for knee alignment accuracy, this advance in automatic HKA prediction from knee radiographs is clinically important for centres that do not have the expensive specialist equipment to perform long-leg radiographs, such as non-specialist centres and those in developing nations. Compared to measurement on full-limb radiographs, HKA predictions from knee-only radiographs reduces costs for the healthcare provides, time for clinicians, and radiation exposure for patients. This fully automatic method is faster than the state-of-the-art semi-automatic method for HKA prediction [11], which requires clinicians to check and manually correct annotated anatomical landmarks (~1 minute processing time) before calculating the alignment, and required exclusion of patients with prior surgery or deformity. In practice, clinicians must be able to trust the predicted alignments; the concomitant heat maps could therefore be a promising tool to explain and distinguish between good and bad predictions.



Figure 5. Six representative heatmaps for FTA (left) and HKA (right) models using the DenseNet architecture annotated with difference in prediction versus the reference measure. Top row) native knee predictions with near-perfect agreement (<0.01°) to the reference measure. Middle row) native knee predictions with low agreement (>3°). Bottom row) predictions for knees with a history of prior surgery exhibiting high agreement (<1°).

These models provide a strong foundation for predicting coronal knee alignment from knee radiographs using deep learning, with potential improvements to be found by fine-tuning the neural network architecture. Further, as all data in this study came from a single source (OAI), additional datasets with more diverse populations and measurement protocols could improve training and provide external validation. Consensus voting, wherein several measures on the same image are aggregated for a more reliable ground truth, could also be used to improve predictions by reducing any error owing to intra/inter-reader variability in the reference measures.

5. Conclusion

Deep learning enables fully automated HKA prediction from knee radiographs with accuracy greater than a state-of-the-art semi-automatic approach. The approach enabled image analysis even for patients with knee deformity or prior knee surgery. With attention heat maps to gauge prediction accuracy, this advance could reduce the need for full-limb radiographs when planning orthopaedic procedures, democratising access to this important measurement, reducing cost for healthcare providers and reducing radiation exposure for patients.

Ethics Committee

Specific ethics committee approval was not required for this article as no new patient data was generated. This article was prepared using an Osteoarthritis Initiative (OAI) public use data set.

CRedit authorship contribution statement

Jinhong Wang: Methodology, Software, Formal analysis, Investigation, Writing – original draft, Visualization. **Thomas A. G. Hall:** Conceptualization, Methodology, Writing – original draft, Visualization. **Omar Musbahi:** Conceptualization, Resources, Writing – review & editing. **Gareth G. Jones:** Writing – review & editing, Supervision. **Richard J. van Arkel:** Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Thomas Hall was funded by UK Research and Innovation through an Engineering and Physical Sciences Research Council Doctoral Training Account. This article was prepared using an Osteoarthritis Initiative (OAI) public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners. The authors wish to thank the participants, principal investigators, co-investigators and staff of all the hospitals who have contributed data to the OAI. The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health.

References

- [1] Cui A, Li H, Wang D, Zhong J, Chen Y, Lu H. Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. *EClinicalMedicine* 2020;29–30. doi: <https://doi.org/10.1016/j.eclim.2020.100587>.
- [2] Brouwer GM, Van TAW, Bergink AP, Belo JN, Bernsen RMD, et al. Association between valgus and varus alignment and the development and progression of radiographic osteoarthritis of the knee. *Arthritis Rheum* 2007;56:1204–11. doi: <https://doi.org/10.1002/art.22515>.
- [3] Stoddart JC, Dandridge O, Garner A, Cobb J, van Arkel RJ. The compartmental distribution of knee osteoarthritis – a systematic review and meta-analysis. *Osteoarthr Cartil* 2021;29:445–55. doi: <https://doi.org/10.1016/j.joca.2020.10.011>.
- [4] Sharma L, Song J, Dunlop D, Felson D, Lewis CE, Segal N, et al. Varus and valgus alignment and incident and progressive knee osteoarthritis. *Ann Rheum Dis* 2010;69:1940–5. doi: <https://doi.org/10.1136/ard.2010.129742>.
- [5] Fang DM, Ritter MA, Davis KE. Coronal Alignment in Total Knee Arthroplasty: Just How Important is it? *J Arthroplasty* 2009;24:39–43. doi: <https://doi.org/10.1016/j.arth.2009.04.034>.
- [6] Garner A, van Arkel RJ, Cobb J. Classification of combined partial knee arthroplasty. *Bone Joint J* 2019;101 B:922–8. <https://doi.org/10.1302/0301-620X.101B8.Bjj-2019-0125.R1>.
- [7] Jones GG, Clarke S, Jaere M, Cobb JP. Failed high tibial osteotomy: A joint preserving alternative to total knee arthroplasty. *Orthop Traumatol Surg Res* 2019;105:85–8. doi: <https://doi.org/10.1016/j.otsr.2018.11.004>.
- [8] Harmon M. 2020 Knee Arthroplasty Market Report: A Global Analysis for 2019 to 2025; 2020.
- [9] Paley D. *Normal Lower Limb Alignment and Joint Orientation*. Princ. Deform. Correct. Berlin Heidelberg: Springer; 2002. p. 1–18. doi: https://doi.org/10.1007/978-3-642-59373-4_1.
- [10] McDaniel G, Mitchell KL, Charles C, Kraus VB. A comparison of five approaches to measurement of anatomic knee alignment from radiographs. *Osteoarthr Cartil* 2010;18:273–7. doi: <https://doi.org/10.1016/j.joca.2009.10.005>.
- [11] Giellis WP, Rayegan H, Arbabi V, Brooghani SYA, Lindner C, Cootes TF, et al. Predicting the mechanical hip–knee–ankle angle accurately from standard knee radiographs: a cross-validation experiment in 100 patients. *Acta Orthop* 2020;91:732–7. doi: <https://doi.org/10.1080/17453674.2020.1779516>.

- [12] Iranpour-Boroujeni T, Li J, Lynch JA, Nevitt M, Duryea J. A new method to measure anatomic knee alignment for large studies of OA: data from the Osteoarthritis Initiative. *Osteoarthr Cartil* 2014;22:1668–74. doi: <https://doi.org/10.1016/j.joca.2014.06.011>.
- [13] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nat* 2015;521:436–44. doi: <https://doi.org/10.1038/nature14539>. 2015 5217553.
- [14] Tack A, Preim B, Zachow S. Fully automated Assessment of Knee Alignment from Full-Leg X-Rays employing a “YOLOv4 And Resnet Landmark regression Algorithm” (YARLA): Data from the Osteoarthritis Initiative. *Comput Methods Programs Biomed* 2021;205:. doi: <https://doi.org/10.1016/j.cmpb.2021.106080>106080.
- [15] Pei Y, Yang W, Wei S, Cai R, Li J, Guo S, et al. Automated measurement of hip-knee-ankle angle on the unilateral lower limb X-rays using deep learning. *Phys Eng Sci Med* 2020.53–62.;2020(441):44. doi: <https://doi.org/10.1007/S13246-020-00951-7>.
- [16] Nguyen TP, Chae DS, Park SJ, Kang KY, Lee WS, Yoon J. Intelligent analysis of coronal alignment in lower limbs based on radiographic image with convolutional neural network. *Comput Biol Med* 2020;120:. doi: <https://doi.org/10.1016/j.compbiomed.2020.103732>103732.
- [17] Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach. *Sci Rep* 2018;8:1–10. doi: <https://doi.org/10.1038/s41598-018-20132-7>.
- [18] Norman B, Padoia V, Noworolski A, Link TM, Majumdar S. Applying Densely Connected Convolutional Neural Networks for Staging Osteoarthritis Severity from Plain Radiographs. *J Digit Imaging* 2018.471–7.;2018(323):32. doi: <https://doi.org/10.1007/S10278-018-0098-3>.
- [19] Leung K, Zhang B, Tan J, Shen Y, Geras KJ, Babb JS, et al. Prediction of Total Knee Replacement and Diagnosis of Osteoarthritis by Using Deep Learning on Knee Radiographs: Data from the Osteoarthritis Initiative. *Radiology* 2020;296:584–93. doi: <https://doi.org/10.1148/radiol.2020192091>.
- [20] Borjali A, Chen AF, Muratoglu OK, Morid MA, Varadarajan KM. Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural network. *J Orthop Res* 2020;jor.24617. <https://doi.org/10.1002/jor.24617>.
- [21] Schwartz AJ, Clarke HD, Spangehl MJ, Bingham JS, Etzioni DA, Neville MR. Can a Convolutional Neural Network Classify Knee Osteoarthritis on Plain Radiographs as Accurately as Fellowship-Trained Knee Arthroplasty Surgeons? *J Arthroplasty* 2020;35:2423–8. doi: <https://doi.org/10.1016/j.arth.2020.04.059>.
- [22] Cooke TDV, Sled EA, Scudamore RA. Frontal plane knee alignment: a call for standardized measurement. *J Rheumatol* 2007;34:1796–801.
- [23] Cooke DT, Harrison L, Khan B, Scudamore A, Chaudhary AM. Analysis of limb alignment in the pathogenesis of osteoarthritis: a comparison of Saudi Arabian and Canadian cases. *Rheumatol Int* 2002.160–4.;2002(224):22. doi: <https://doi.org/10.1007/S00296-002-0218-7>.
- [24] Sled EA, Sheehy LM, Felson DT, Costigan PA, Lam M, Cooke TDV. Reliability of lower limb alignment measures using an established landmark-based method with a customized computer software program. *Rheumatol Int* 2009.71–7.;2009(311):31. doi: <https://doi.org/10.1007/S00296-009-1236-5>.
- [25] Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. *Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017* 2016;2017-January:2261–9.
- [26] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *Thirty-First AAAI Conf Artif Intell*; 2017.
- [27] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc*; 2014.
- [28] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis* 2015.211–52.;2015(1153):115. doi: <https://doi.org/10.1007/S11263-015-0816-Y>.
- [29] Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Comput Vis - ECCV 2014*. EECV 2014. Lect Notes Comput Sci, vol. 8689 LNCS, Springer, Cham; 2014. p. 818–33. https://doi.org/10.1007/978-3-319-10590-1_53.
- [30] Odenbring S, Berggren A-M, Peil L. Roentgenographic Assessment of the Hip-Knee-Ankle Axis in Medial Gonarthrosis: A Study of Reproducibility. *Clin Orthop Relat Res* 1993;289:195–6.
- [31] Shin C, Crovetti C, Huo E, Lionberger D. Unsatisfactory accuracy of recent robotic assisting system ROSA for total knee arthroplasty. *J Exp Orthop* 2022;9:1–6. doi: <https://doi.org/10.1186/S40634-022-00522-7/TABLES/2>.
- [32] Doan GW, Courtis RP, Wyss JG, Green EW, Clary CW. Image-Free Robotic-Assisted Total Knee Arthroplasty Improves Implant Alignment Accuracy: A Cadaveric Study. *J Arthroplasty* 2022;37:795–801. doi: <https://doi.org/10.1016/j.arth.2021.12.035>.
- [33] Sires JD, Craik JD, Wilson CJ. Accuracy of Bone Resection in MAKO Total Knee Robotic-Assisted Surgery. *J Knee Surg* 2021;34:745–8. doi: <https://doi.org/10.1055/S-0039-1700570/ID/IR190204OA-12>.
- [34] Baier C, Maderbacher G, Springorum HR, Zeman F, Fitz W, Schaumburger J, et al. No difference in accuracy between pinless and conventional computer-assisted surgery in total knee arthroplasty. *Knee Surgery, Sport Traumatol Arthrosc* 2014;22:1819–26. doi: <https://doi.org/10.1007/S00167-013-2430-2/FIGURES/7>.
- [35] Li C, Zhang Z, Wang G, Rong C, Zhu W, Lu X, et al. Accuracies of bone resection, implant position, and limb alignment in robotic-arm-assisted total knee arthroplasty: a prospective single-centre study. *J Orthop Surg Res* 2022;17:61. doi: <https://doi.org/10.1186/S13018-022-02957-1>.