EPiC Series in Health Sciences
Volume 8, 2026, Pages 109–113
Proceedings of The 25th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery

Machine learning-based automatic implant size prediction for total knee arthroplasty using bone dimensions

Sandeep Katragadda[1] and Kevin de Souza[2]

[1] Stryker UK, sandeep.katragadda@stryker.com
[2] Stryker UK, kevin.desouza@stryker.com

Abstract

Predicting suitable implant sizes from 3D radiographic images of joint anatomy can be accomplished using templating methods. Automatic templating, which eliminates the need for manual intervention, is especially valuable for speeding up the creation of computer- or robot-assisted surgical plans. In our previous work, automatic templating for total knee arthroplasty was achieved through automatic bone segmentation, followed by matching a set of anatomical landmarks with corresponding points on candidate implants of various sizes. This paper introduces a novel approach that eliminates the reliance on point correspondences and matching, instead leveraging bone dimensions for implant size prediction. An experimental analysis on a dataset of 3261 knee cases demonstrates that the proposed method improves the performance of implant size prediction.

1 Introduction

Incorrect selection of femoral and tibial implants in total knee arthroplasty (TKA) can lead to various post-operative complications [1]. Moreover, precise and early prediction of implant sizes offers several advantages to operative field preparation, inventory control, and resource optimization [2]. Templating-based approaches use implants templates (of all sizes) along with bone radiographs [3, 4, 1]. A limitation of existing templating-based methods is the reliance on manual intervention to complete the process. They often manually compute bone dimensions such as the anterior-posterior (AP) width and the medial-lateral (ML) width [4, 5]. The "auto-knee" algorithm in TraumaCAD, a commercial digital templating system, automates the process but achieves significantly lower accuracy in size prediction compared to manual templating [6]. Our previous works [7, 8] achieve automated digital templating by matching anatomical landmarks on segmented bones with their corresponding points on candidate implants. However, these approaches rely on manually predefined correspondences between points on the bone model and those on all candidate implants, limiting their scalability and introduces potential for human error.

In this paper, we propose a novel approach that eliminates the need for predefined landmark correspondences. Instead, the method automatically extracts key bone dimensions from the segmented bone model. Only one dimension per bone is used to predict the size of each implant component. Intraoperatively, surgeons typically assess femoral size using the AP dimension.

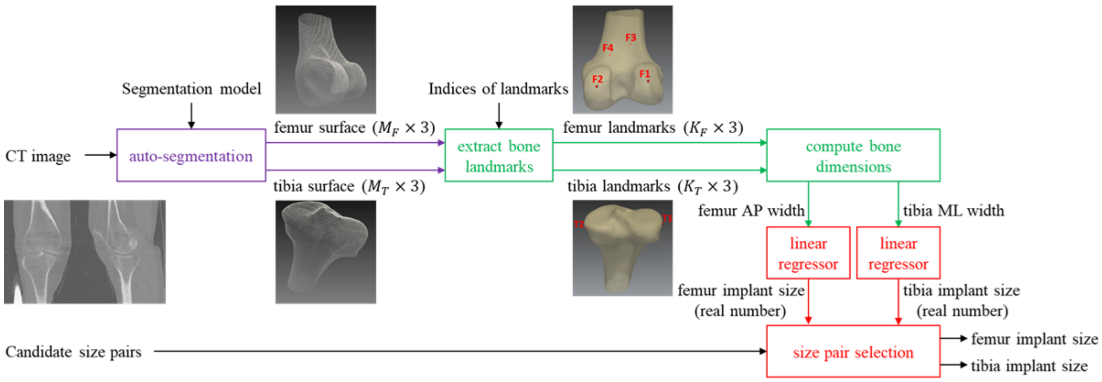

Figure 1: Block diagram of the proposed algorithm, with the three main steps distinguished by color: purple, green, and red.

Although some manufacturers provide narrow implant options, the ML measurement of the femur is generally less reliable for accurate sizing. In contrast, the ML dimension of the tibia is the most commonly used metric for intraoperative tibial sizing [5]. Based on these clinical practices, two linear regressors are trained using the femoral AP and tibial ML dimensions to predict the corresponding implant sizes. These regressors leverage the correlation between the extracted bone measures and the implant sizes selected by surgeons, resulting in a fully automated and scalable prediction pipeline.

2 Proposed approach

Given a 3D radiographic image of the knee (e.g. CT image), the objective is to identify the most suitable femur-tibia implant size pair from the set of available candidate pairs. Figure 1 shows the block diagram of the proposed approach. It consists of three main steps.

(1) Auto-segmentation: Similar to our previous works [7, 8], each given knee image is processed using a trained segmentation model to produce osteophyte-free (OF) surfaces of the femur and tibia bones. The model is built via active appearance modelling [9]. The surfaces are 3D meshes having a constant number (M_b) of corresponding vertices meaning that the same anatomical region on the bone surfaces of different patients has vertices with the same indices. Here, $b \in \{F, T\}$ (F: Femur and T: Tibia).

(2) Computation of bone dimensions: Indices of a set of K_b landmarks on the bone model are identified in advance. The landmarks are chosen such that they well represent the bone dimensions. $K_F = 4$ and $K_T = 2$. Two of the femur landmarks are on the posterior side (F1 and F2) and the other two landmarks are on the anterior side (F3 and F4). The two tibial landmarks are on the medial and lateral sides of the tibial plateau (T1 and T2). Based on these indices, coordinates of the landmarks are extracted from the OF surfaces. The femur AP width is represented by the average of the F1-F3 distance and F2-F4 distance. The tibial ML width is represented by the T1-T2 distance.

(3) Linear regression: The dimensions undergo linear regression that models the correlation between the bone dimensions and the surgeons' implant size choices. The regression model for each bone b is: implant size $= m_b \times$ bone dimension $+ c_b$, where m_b and c_b are regression parameters (scalars) that are computed via supervised learning in advance. Finally, the candidate

size pair that is close to the regression output pair is selected using the Euclidean distance.

3 Experimental analysis

The performance of the proposed approach is evaluated by training the linear regressors using a dataset comprising 292 knee CT images, including 152 right and 140 left cases. Multiple surgeons performed the TKA procedure on these cases using Stryker Triathlon PS and CR implants, which are available in 8 discrete sizes. The implant sizes selected by the surgeon are used as the ground truth for this evaluation. The dataset covers a comprehensive range of implant sizes. In line with established clinical practice, when using Stryker Triathlon implants, the size difference between femoral and tibial implants should not exceed one size. This recommendation is grounded in medical expertise rather than being dictated by algorithmic constraints or technical limitations. Hence, the candidate size pairs of femur-tibia implants used in the dataset are (1,1), (1,2), (2,1), (2,2), (2,3), (3,2), (3,3), (3,4), (4,3), (4,4), (4,5), (5,4), (5,5), (5,6), (6,5), (6,6), (6,7), (7,6), (7,7), (7,8), (8,7) and (8,8).

Three metrics are used to analyze the performance. (1) Mean Absolute Error (MAE): the mean absolute difference between the predicted and selected implant sizes. Smaller MAE values indicate better prediction accuracy. (2) P_0: the percentage of cases in which the predicted implant size matches the size selected by the surgeon. (3) P_1: the percentage of cases in which the predicted implant size differs from the surgeon's selection by at most one size. This metric is commonly used because it accommodates both inter- and intra-observer reliability. Higher values of P_0 and P_1 indicate better prediction accuracy.

To analyze the performance of the linear regression models, 5-fold cross-validation is used. The data (containing the computed bone dimensions and the corresponding ground truth sizes) is divided into five subsets, namely $F_1 - F_5$ and $T_1 - T_5$ for each of the femur and tibia bones respectively, ensuring that each subset includes a representative distribution of implant sizes used by the surgeon. During each cross-validation iteration, one subset is reserved for testing, while the remaining four subsets are combined to train the regressor. The learned parameters (m_b and c_b) are then applied to predict implant sizes for the test set samples. These predictions are compared with the ground truth to compute the MAE, P_0 and P_1 values on the test set. Additionally, the correlation coefficient (r_b) between the independent variable (regressor input) and the dependent variable (regressor output) is calculated. Higher r_b values signify a better fit of the regression model to bone b data. Table 1 presents the learned parameters for each iteration and the corresponding performance metrics on the test sets for femur and tibia data. The higher r_F and r_T values highlight the suitability of the regression models for this dataset.

The performance of the trained model is evaluated using an independent dataset (used in [7]) comprising 3261 knee CT images, including 1714 right and 1547 left cases on both the proposed approach and our previous works [7, 8]. The regression parameters used are $m_F = 0.27$, $c_F = -14.0$, $m_T = 0.23$ and $c_F = -12.8$. The observations are as follows.

- The MAE values for the proposed approach are 0.51 and 0.56 for the femur and tibia, respectively. In comparison, the previous approach [8] yields 0.62 and 0.57 for the femur and tibia, respectively. Another prior approach [7] results in MAE values of 0.8 and 1.23 for the femur and tibia, respectively.

- For the femur bones, the proposed approach achieves P_0 and P_1 accuracies of 53.33% and 95.86%, respectively, compared to 46.18% and 92.27% for the previous approach [8], and 35.08% and 85.65% for another previous approach [7].

Training data		Testing data		Learned parameters			No. of cases with absolute error e				Prediction accuracy		
Sets	Size	Sets	Size	m_b	c_b	r_b	$e=0$	$e=1$	$e=2$	$e=3$	MAE ↓	$P0$ ↑	$P1$ ↑
F_1, F_2, F_3, F_4	213	F_5	59	0.27	-13.6	0.87	27	30	2	0	0.58	45.2	96.6
F_1, F_2, F_3, F_5	214	F_4	58	0.26	-13.5	0.87	33	24	0	1	0.46	56.9	98.3
F_1, F_2, F_4, F_5	213	F_3	59	0.26	-13.5	0.86	32	26	1	0	0.47	54.2	98.3
F_1, F_3, F_4, F_5	214	F_2	58	0.26	-13.5	0.86	29	29	0	0	0.5	50.0	100.0
F_2, F_3, F_4, F_5	214	F_1	58	0.26	-13.5	0.87	34	20	3	1	0.5	58.6	93.1
T_1, T_2, T_3, T_4	213	T_5	58	0.23	-12.75	0.92	37	20	1	0	0.38	63.8	98.3
T_1, T_2, T_3, T_5	213	T_4	58	0.23	-12.9	0.92	35	23	0	0	0.39	60.3	100.0
T_1, T_2, T_4, T_5	214	T_3	58	0.23	-12.7	0.92	36	21	1	0	0.39	62.1	98.3
T_1, T_3, T_4, T_5	214	T_2	59	0.23	-12.8	0.93	36	22	0	1	0.42	61.0	98.3
T_2, T_3, T_4, T_5	214	T_1	59	0.23	-12.7	0.92	40	18	1	0	0.34	67.8	98.3

Table 1: Cross-validation results. m_b, c_b and r_b are regression parameters for bone $b \in \{F, T\}$.

- For the tibial bones, the proposed approach achieves P_0 and P_1 accuracies of 49.19% and 95.09%, respectively, compared to 48.48% and 94.76% for the previous approach [8], and 15.36% and 65.53% for another previous approach [7].

It can be observed that the proposed approach results in higher P_0 and P_1 and smaller MAE for both the femur and tibia and thereby demonstrates an improved implant size prediction accuracy compared to previous works.

4 Conclusion

This paper proposes a machine learning-based knee implant size prediction approach. Femur and tibia bone dimensions are estimated from the auto-segmented bone meshes, and two linear regressors utilize these dimensions to predict the implant sizes for femur and tibia bones. The prediction process is fully automated, requiring no manual intervention. The approach demonstrates an improved prediction performance compared to our previous works by predicting implant sizes in above 95% of cases (both femurs and tibias) with at most 1-size error.

References

[1] R. W. Trickett, P. Hodgson, M. C. Forster, and A. Robertson. The reliability and accuracy of digital templating in total knee replacement. The Journal of Bone & Joint Surgery British Volume, Vol. 91-B, No. 7, Pages: 903-906, 2009.

[2] Kyle Kunze, Evan Polce, Arpan Patel, P. Courtney, and Brett Levine. Validation and performance of a machine-learning derived prediction guide for total knee arthroplasty component sizing. Archives of Orthopaedic and Trauma Surgery, Vol. 141, Dec 2021.

[3] Bertram The, Ron L Diercks, Peter M A van Ooijen, and Jim R van Horn. Comparison of analog and digital preoperative planning in total hip and knee arthroplasties. Acta Orthopaedica, Vol. 76, No. 1, Pages: 78-84, 2005.

[4] Siwadol Wongsak, Viroj Kawinwonggowit, Pornchai Mulpruek, Thanaphot Channoom, and Patarawan Woratanarat. Accuracy of knee implants sizing predicted by digital images. Journal of the Medical Association of Thailand, Vol. 92 Suppl 6, Pages: S85-90, Dec 2009.

[5] Robert A. Sershon, Paul Maxwell Courtney, Brett D. Rosenthal, Scott M. Sporer, and Brett R. Levine. Can demographic variables accurately predict component sizing in primary total knee arthroplasty? The Journal of Arthroplasty, Vol. 32, No. 10, Pages: 3004-3008, 2017.

[6] Thomas Seaver, Kenneth McAlpine, Enrique Garcia, Ruijia Niu, and Eric L. Smith. Algorithm based automatic templating is less accurate than manual digital templating in total knee arthroplasty. Journal of Orthopaedic Research, Vol. 38, No. 7, Pages: 1472-1476, 2020.

[7] Sandeep Katragadda and Kevin de Souza. Implant size prediction for total knee arthroplasty using anatomical landmarks. Proceedings of the 12th international conference on New Technologies for Computer and Robot Assisted Surgery (CRAS), September 2023.

[8] Sandeep Katragadda and Kevin de Souza. Machine learning-based automatic implant size prediction from CT images in total knee arthroplasty. Proceedings of the IEEE RAS EMBS 10th International Conference on Biomedical Robotics and Biomechatronics (BioRob), September 2024.

[9] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3D medical image segmentation: A review. Medical Image Analysis, Vol. 13, No. 4, Pages: 543-563, 2009.