Annals of Biomedical Engineering
https://doi.org/10.1007/s10439-025-03863-3

ORIGINAL ARTICLE

Gait-to-Contact (G2C): A Novel Deep Learning Framework to Predict Total Knee Replacement Wear from Gait Patterns

Mattia Perrone[1] · Scott Simmons[1,2] · Philip Malloy[1,3] · Vasili Karas[1] · Catherine Yuh[1] · John Martin[1] · Steven P. Mell[1]

Received: 5 October 2024 / Accepted: 21 September 2025

Abstract

Purpose Total knee replacement (TKR) is the most common inpatient surgery in the US. Studies leveraging finite element analysis (FEA) models have shown that variability of gait patterns can lead to significant variability of wear rates in TKR settings. However, FEA models can be resource-intensive and time-consuming to execute, hindering further research in this area. This study introduces a novel deep learning-based surrogate modeling approach aimed at significantly reducing computational costs and processing time compared to traditional FEA models.

Methods A published method was used to generate 314 variations of ISO14243-3 (2014) anterior/posterior translation, internal/external rotation, flexion/extension, and axial loading time series, and a validated FEA model was used to calculate linear wear distribution on the polyethylene liner. A deep learning model featuring a transformer-CNN based encoder–decoder architecture was trained to predict linear wear distribution using gait pattern time series as input. Model performance was evaluated by comparing the deep learning and FEA model predictions using metrics such as mean absolute percentage error (MAPE) for relevant geometric features of the wear scar, structural similarity index measure (SSIM), and normalized mutual information (NMI).

Results The deep learning model significantly reduced the computational time for generating wear predictions compared to FEA, with the former training and inferring in minutes, and the latter requiring days. Comparisons of deep learning model wear map predictions to FEA results yielded MAPE values below 6% for most of the variables and SSIM and NMI values above 0.88, indicating a high level of agreement.

Conclusion The deep learning approach provides a promising alternative to FEA for predicting wear in TKR, with substantial reductions in computational time and comparable accuracy. Future research will aim to apply this methodology to clinical patient data, which could lead to more personalized and timely interventions in TKR settings.

Keywords Deep learning · Transformers · Total knee replacement · Wear prediction · Finite element analysis

Introduction

Total knee replacement (TKR) is the most common inpatient surgery in the USA [1], with annual procedures anticipated to increase to 3.48 million by 2030 [2]. It is estimated that over 4.2% of adults are living with a TKR in the US alone [3, 4]. Additionally, there is a demographic trend toward younger and more active patients undergoing TKR surgery, raising the concern of component wear [5, 6]. The growing number of individuals undergoing TKR surgeries magnifies the economic and societal burden of even a small percentage of failures, and underscores the importance of comprehending the causes behind the failure of TKR.

Wear of ultra-high molecular weight polyethylene (PE) in TKR devices remains a significant factor for mid-late TKR failure [7–10], with recent studies suggesting little change in wear-related failures over time [11, 12] and a primary revision rate due to wear-related causes of up to 15% [7]. Although modern highly cross-linked polyethylene has markedly reduced clinically recognized wear-related failures in TKR (with the 2024 American Joint Replacement

Associate Editor Joel Stitzel oversaw the review of this article.

✉ Mattia Perrone
mattia_x_perrone@rush.edu

[1] Rush University Medical Center, Chicago, IL, USA

[2] Drury University, Springfield, MO, USA

[3] Arcadia University, Glenside, PA, USA

Published online: 27 September 2025

Registry reporting ~ 4% of revisions coded as wear or osteolysis [13]), these figures likely underestimate the true incidence of these failures, due in part to limitations in registry coding, which have been shown to be error prone [14]. Material damage such as delamination and wear can also occur concomitantly with other failure modes, and can present clinically as instability or loosening. However, material damage itself can often only be identified intraoperatively at revision and not on preoperative imaging where the diagnosis is made and coded [7]. Moreover, particulate debris from even subclinical wear may drive biological responses that manifest as aseptic loosening years later, further obscuring its contribution to failure in registry data [7]. For these reasons, TKR wear could be misclassified under broader failure categories such as 'instability,' 'other mechanical complications,' or 'mechanical loosening' [7]. Hence, studying wear remains vital, as it may be an overlooked determinant of TKR longevity and a key target for improving both implant design and patient outcomes.

Given the challenges in accurately capturing wear-related failures, an enhanced understanding of patient-specific factors that contribute to wear has become essential. One patient-specific factor known to contribute to TKR wear is the patient's movement patterns during walking or knee kinematics and kinetics [15–17]. Identifying how aberrant knee kinematics and kinetics contribute to wear may allow for early intervention in high-risk patients. However, the relationship between how a TKR patient moves versus risks of TKR wear-related failure remains understudied. Therefore, there is a need for comprehensive tools that are capable of linking variability in movement to TKR wear patterns, to improve predictions of TKR failure and outcomes, both clinically and for research.

Finite element analysis (FEA) is a computational tool extensively used to investigate wear profiles in TKR [17, 18] and has demonstrated that variability in gait can lead to variability in implant wear [15]. However, FEA can be resource-intensive and time-consuming to execute, with many studies reporting a computational time of hours or even days [19]. This may hinder further large-scale parametric investigations, prompting the exploration of alternative methodologies to reduce processing time. Surrogate modeling is an approach where a computationally efficient mathematical approximation of a system is built using data selectively generated from a more complex and high-fidelity simulation, such as FEA [20]. We have previously used this approach to create surrogate linear regression models to predict wear rates in TKRs under a variety of conditions [15, 18, 21]. However, these models only predict volumetric wear rates, and neglect the spatial distribution of wear provided by FEA models. This spatial information may be important for understanding failure. For example, a deep wear scar with a smaller area may be more detrimental than a more diffuse wear scar, despite both having the same overall wear volume. Likewise, a wear scar situated posteriorly, or with a greater angle between the medial and lateral compartments, may demonstrate to clinicians that implant malalignment is the source of wear debris and a subsequent foreign body reaction.

Recently, growing interest in deep learning has increased its application within the computational biomechanics domain. Several studies have used deep learning models as surrogate models, presenting a promising alternative to conventional FEA approaches, with a notable advantage in terms of processing time [19]. Such surrogate models have found applications in various areas, including studies of the liver [22, 23], thoracic aorta [24], and lower limb biomechanics [25]. Most of these studies concentrate on assessing deformations in anatomical regions, which is a common output of FEA. However, there is a limited number of investigations employing this methodology in analysis of total joint replacements, with none aimed at predicting mechanical wear in total knee implants. While some studies have attempted to use surrogate machine learning models to predict strains [26] and stresses [27] at the joint interface, no studies have utilized time series data such as gait kinematics and kinetics to predict surface wear patterns.

In this study, we introduce a deep learning framework capable of predicting tibial polyethylene wear in a TKR, starting from input kinematic and kinetic gait patterns, lowering computational costs and processing time compared to FEA. We utilized an existing dataset of 314 FEA simulations where kinematics and kinetics were modified to predict polyethylene wear in TKRs. Unlike previous studies that utilize fully connected neural networks or traditional machine learning algorithms, our approach introduces a novel transformer encoder paired to a CNN decoder architecture, and predict spatial polyethylene wear from gait. This architecture enables the generation of polyethylene wear maps as output data to visually assess implant performance in a clinical setting. To our knowledge, this is the first surrogate deep learning model for FEA in TKR settings.

Methods

Training Dataset

In a previous study, we generated a synthetic dataset of 314 unique variations of the ISO14243-3(2014) standard kinematic and kinetic gait waveforms, including anterior/posterior translation, internal/external rotation, flexion/extension, and axial loading [15], using Latin Hypercube Sampling [28]. Waveforms were all within a range of ± 25% of 9 characteristic peaks of the ISO Standard (Fig. 1). These unique waveforms were used as inputs to a previously published and

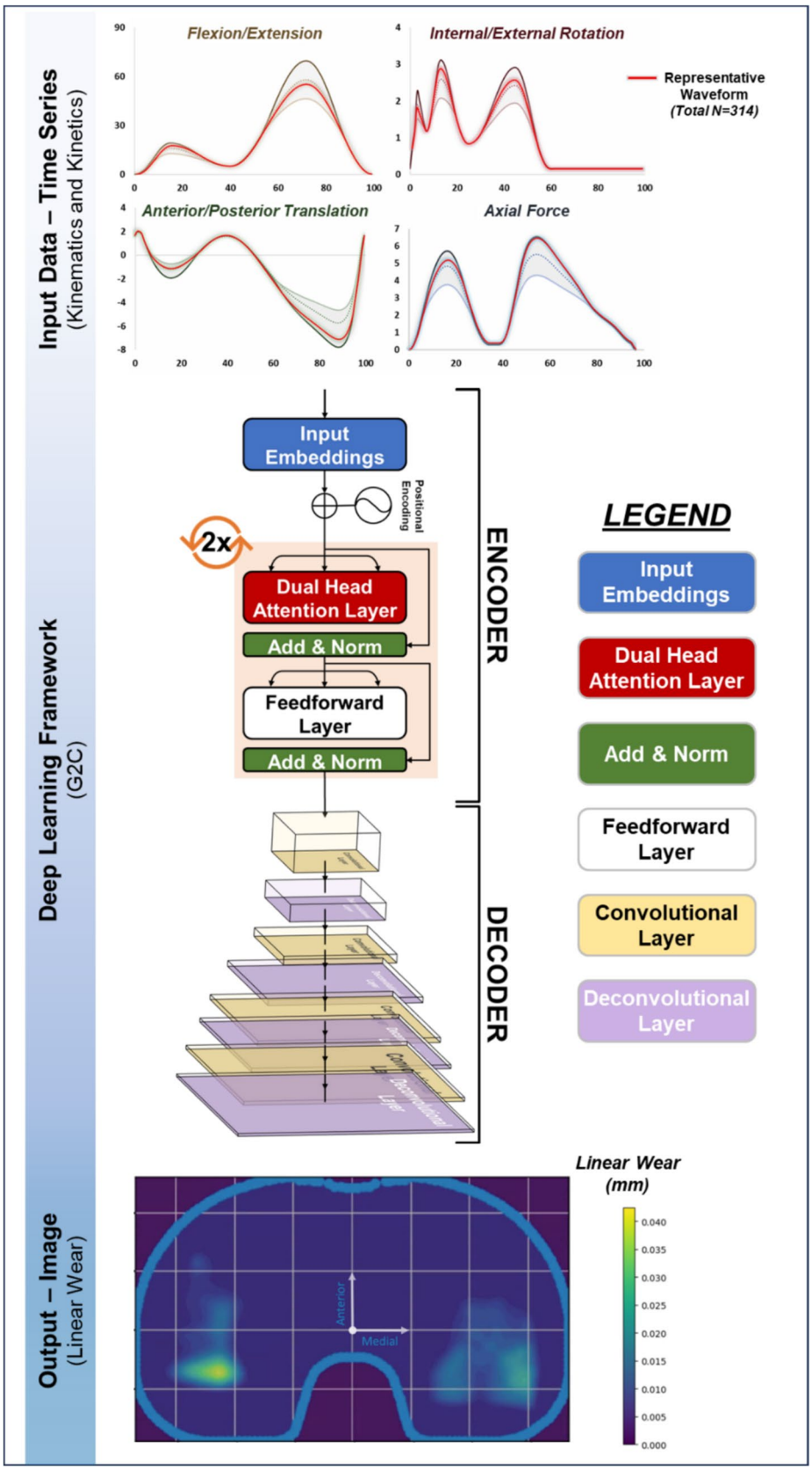

Fig. 1 The proposed model consists of a U-net architecture incorporating a dual-head transformer encoder to discern temporal dependencies within the multivariate time series and deconvolutional layers in the decoder to ensure the reconstruction of image data. 314 input time series were normalized to 100 points, resulting in a total shape of $314 \times 100 \times 4$. The model outputs a 100×100 pixel linear wear map prediction.

validated TKR model [15, 18, 21, 29] and used to calculate linear wear distribution on the polyethylene liner. Our FE model has previously demonstrated an R^2 value of 0.84 for volumetric wear when compared to mechanical knee simulator tests under 3 different loading conditions [21]. We have also extensively demonstrated the model's ability to replicate linear wear distribution as well as distribution of wear and contact area [29]. All time series generated by the previous study were normalized to 100 time steps, resulting in a total shape of 314 × 100 × 4 for our model input data. For our ground truth model output, the mesh of the polyethylene tibial insert was exported from Abaqus v2017/Standard (Dassault Systèmes, Waltham, MA) and interpolated to a 100 × 100 pixel grid using Python (version 3.12.1), so that each pixel represents a specific value of linear wear.

Model Architecture

A deep learning model was developed to predict polyethylene wear on the interpolated image starting from the gait pattern time series (Fig. 1). We implemented an encoder-decoder architecture incorporating a dual-head transformer encoder and transposed convolutional layers in the decoder. Our model uses the transformer encoder to condense gait time series into a compressed but feature-rich latent space, while the decoder employs transposed convolutional layers to reconstruct pixel-wise spatial structure from latent feature representations. We specifically chose to use transformers, as they are often better able to manage long-range dependencies within the time series data compared to other commonly used models, such as long short-term memory models and recurrent neural networks [30]. In designing the model architecture, we thoroughly tuned model hyperparameters on the transformer encoder to enhance feature learning [31] and experimented with the depths and widths of decoder layers.

For the decoder, we employed transposed convolutional layers to reconstruct the spatial dimensions from the compressed representations and convolutional layers to enhance feature representations by extracting spatial details. Specifically, we leveraged a strategy of progressively halving the channel dimensions for each transposed convolutional layer. This setup of combining convolutional and transposed convolutional layers is particularly beneficial in tasks requiring generation of high-resolution outputs from lower-resolution inputs [32]. This approach, combined with the use of smaller (3) and larger (6) kernel sizes for transposed convolutional and convolutional layers, respectively, allows for an optimal trade-off between capturing local details and aggregating broader contextual information. Strides of 2 across all deconvolutional layers ensure effective dimensionality reduction and expansion without excessive computational costs. The final layer of the decoder uses a 1x1 convolution to transform the multi-channel feature maps into the desired single-channel output consisting of the wear value for each specific pixel.

Model Training and Evaluation Metrics

Hyperparameters tested in the encoder include the input dimension, the number of transformer layers and transformer attention heads, the number of feedforward layers within each transformer layer, as well as training parameters (e.g., learning rate) (Table 1). The optimal configuration identified for these hyperparameters were an input dimension of 128, 2 transformer heads, 2 transformer layers each with a feedforward layer of 256, and a learning rate of 9.5×10^{-5}. To identify the optimal configuration across a broad range of hyperparameters, we employed Optuna [33], an optimization framework designed to automate hyperparameter tuning (version 3.6.1). Specifically, tree-structured Parzen estimators [34] (TPE) were used to model the probability distributions of hyperparameters and guide the search toward the most effective configurations.

Before training the model, each input feature was individually normalized to a range of 0 to 1 to improve model predictions. After performing a random 60/20/20 training/validation/test split, we trained the model using mean squared error as the loss function, conducted hyperparameter tuning on the validation set and assessed the performance of the best model on the test set. Comparisons were made between predictions from the deep learning model and the gold standard predictions from the FEA model. These were assessed in both the medial and lateral compartments, using the mean absolute percentage error (MAPE) as error metric. MAPE is defined as follows:

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|,$$

where $\hat{y}$ are the predictions from the deep learning model, y are the ground truth from the FEA model, and n is the total number of samples in the dataset. To provide a comprehensive assessment of the predicted wear scar heatmaps, we examined metrics that are particularly relevant to the evaluation of wear scars in TKR settings, along with those commonly used in general heatmap analysis. Wear scar area

Table 1 Overview of the range of values tested for each hyperparameter in the study. The bolded values indicate the hyperparameters that were selected for the final model configuration. Square brackets denote that any value within the specified extremes was eligible for testing.

Hyperparameter	Range of values tested
Input Dimension	**128** - 256 - 512
Number of Transformer Layers	1 - **2** - 3 - 4
Number of Heads per Transformer Layer	1 - **2** - 4 - 8
Feedforward Layer Dimension	128 - **256** - 512
Learning Rate	[10^{-6} - 10^{-4}]

(both weighted by linear penetration and non-weighted) and angle between centroids were evaluated for their specific relevance to TKR wear patterns [21, 35–37] (Fig. 2), where weighted wear scar area was determined by summing the values of linear wear across all pixels, while non-weighted wear scar area was calculated by identifying pixels that fall within the wear scar boundary and simply summing their count. NMI (Normalized Mutual Information) and SSIM (Structural Similarity Index Measure) were utilized for broader pattern identification across heatmaps [38, 39]. Specifically, SSIM assessed the spatial fidelity between FEA and deep learning heatmaps, while NMI measured the shared information between the heatmaps predicted by these two methodologies. We also investigated other metrics, such as width and length of the wear scar (Fig. 2). All these parameters were evaluated both for the medial and lateral compartments of the implant. To assess whether the grid size used for interpolation of the wear scar influenced the results when exporting the mesh from Abaqus, we analyzed different configurations with grid sizes of 100 × 100, 150 × 150, and 200 × 200 pixels. Python libraries used for data preprocessing, model implementation and hyperparameters tuning include Pandas, NumPy, SciPy, Scikit-learn, PyTorch, PyTorch Lightning, and Optuna. The model was trained on an NVIDIA A5000 GPU with 24 GB of VRAM.

Results

Deep learning model predictions were compared to the ground truth data from the FEA simulation results of TKR wear (Fig. 3). MAPE for the non-weighted wear scar area remained approximately 4% for both compartments (sd lateral: 0.28–0.43%; sd medial: 0.38–0.49%), while for the weighted wear scar areas, MAPE values were approximately 12% for both compartments (sd lateral: 1.34–1.43%;

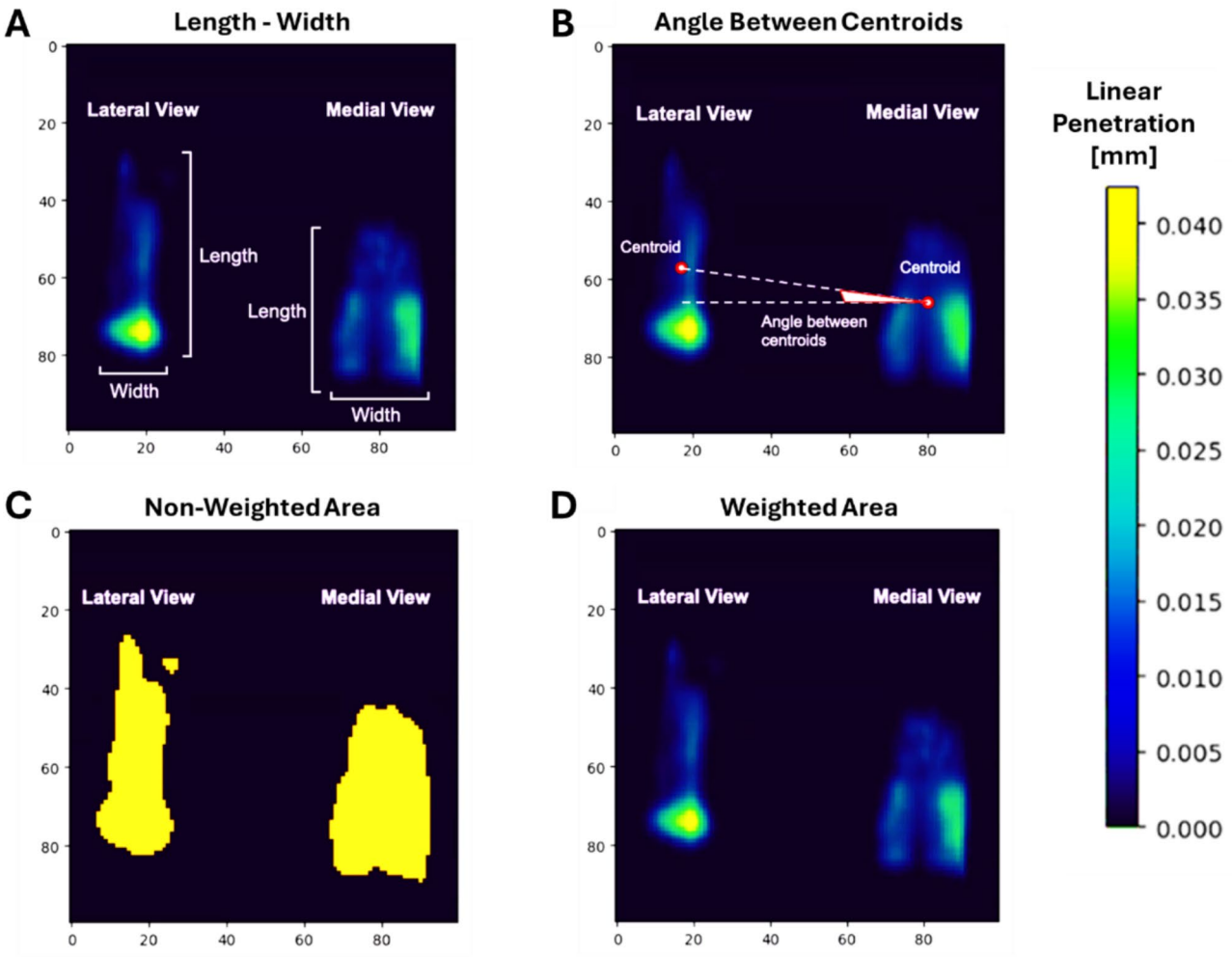

Fig. 2 Metrics utilized in the current study, including the length and width of the wear (**a**), the angle between centroids (**b**), the non-weighted area (**c**), and the weighted area (**d**). Results for both the lateral and medial compartments are reported.

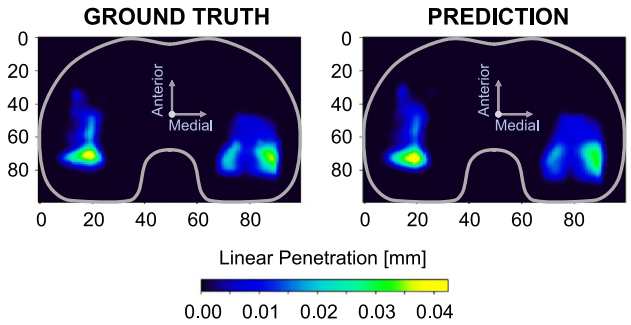

Fig. 3 Representative sample (closest to the average mean squared error) comparing the deep learning model prediction to the FEA model ground truth.

sd medial: 1.05–1.13%) (Table 2). The MAPE for the angle between centroids was around 13% (sd: 0.89–1.42%) for the weighted areas and 15% (sd: 1.15–1.47%) for the non-weighted areas. Width and length measurements of the wear scar showed errors consistently below 6% (sd: 0.29–0.47%) for both compartments, except for the width of the lateral compartment at a 100x100 grid size, which had an error of 7.69% (sd: 0.77%) (Table 2). Additionally, for both compartments, SSIM and NMI values remained above 0.88 and 0.89, respectively (Table 3). The model training times for grid sizes of 100 × 100, 150 × 150, and 200 × 200 were 672 seconds, 716 seconds, and 946 seconds, respectively, while inference times were less than two seconds for all configurations.

Discussion

We developed a surrogate modeling framework to predict TKR component wear using a deep learning model trained on FEA simulations. The deep learning model requires only kinematic and kinetic variables as input data, reducing the resources and time to calculate TKR wear. Specifically, all input variables can be acquired using existing motion analysis techniques via wearable sensors, cameras, or force plates [40–42], removing the dependence on invasive methods such as embedded sensors within the implant [43–45] or from complex musculoskeletal models [46–48]. With the rise of markerless motion capture and portable force plates, all input data can be easily obtained in a non-invasive manner in a clinical setting [49–51].

Table 2 Mean and standard deviation of absolute percentage error (APE) across different interpolation grids (100 × 100, 150 × 150, and 200 × 200). The comparison includes metrics for wear scar width, length, area, and angle between centroids (both non-weighted and weighted) for both lateral and medial compartments.

Parameter	Absolute Percentage Error (APE) – 100x100		Absolute Percentage Error (APE) – 150x150		Absolute Percentage Error (APE) – 200x200	
	Mean	Std dev	Mean	Std dev	Mean	Std dev
Width (Lateral)	7.69%	0.77%	3.57%	0.39%	5.58%	0.47%
Length (Lateral)	5.28%	0.46%	5.06%	0.46%	4.69%	0.41%
Area – non-weighted (Lateral)	3.51%	0.43%	4.05%	0.37%	3.42%	0.28%
Area – weighted (Lateral)	12.33%	1.34%	12.46%	1.40%	13.29%	1.43%
Width (Medial)	1.93%	0.29%	2.52%	0.29%	1.43%	0.16%
Length (Medial)	4.37%	0.41%	4.31%	0.40%	3.90%	0.34%
Area – non-weighted (Medial)	4.75%	0.49%	4.61%	0.38%	4.80%	0.42%
Area – weighted (Medial)	12.54%	1.07%	11.66%	1.05%	13.46%	1.13%
Angle between centroids – non-weighted	15.12%	1.47%	13.14%	1.15%	13.78%	1.28%
Angle between centroids – weighted	11.58%	1.42%	13.65%	1.41%	9.41%	0.89%

Table 3 Structural similarity index measure (SSIM) and normalized mutual information (NMI) values for each grid size with confidence intervals.

Parameter	Interpolation grid: 100x100	Interpolation grid: 150x150	Interpolation grid: 200x200
SSIM (lateral)	0.89 (0.88, 0.90)	0.90 (0.89, 0.91)	0.95 (0.94, 0.96)
SSIM (medial)	0.88 (0.87, 0.89)	0.94 (0.94, 0.95)	0.96 (0.95, 0.96)
NMI (lateral)	0.94 (0.94, 0.95)	0.95 (0.94, 0.95)	0.89 (0.88, 0.91)
NMI (medial)	0.96 (0.96, 0.97)	0.96 (0.95, 0.96)	0.96 (0.95, 0.96)

This novel approach may have exciting future clinical implications to help identify patients at higher risk of implant failure. Markerless gait analysis workflows could measure the necessary kinematic and kinetics in the clinic and then be used to generate a resulting wear pattern [52]. This enhanced patient monitoring approach could help extend the lifespan of implants and improve patient outcomes. In this context, the role of the deep learning model is to capture the impact of gait patterns on wear distributions within the knee implant, which is a relationship that has been underscored in several studies [16, 17], and may help promote timely intervention to address patients at risk for wear-related failure.

In addition to TKR wear prediction, the presented G2C framework is broadly applicable to human articulating joint mechanics, whether artificial or native. TKR wear is used in this study as a first demonstration of the capability of the G2C framework to translate motion into a predicted *in vivo* 'joint contact exposure profile.' While TKR wear patterns are the focus of this particular study, wear can also be useful as a proxy metric of TKR contact mechanics and contact area. Quantifying TKR contact mechanics and contact area has broad clinical implications for TKR function and could aid in studying instability (a failure of joint *function*), which is a major cause of revision surgery [13]. Evaluating joint contact mechanics *in vivo* could provide insight into why an implant is not functioning properly, due to factors, including (1) TKR design [53, 54], (2) surgical technique (TKR alignment, ligament balance) [18, 55–59], (3) anatomical changes (bone remodeling, osteophyte formation) [60–63], (4) patient lifestyle (body mass index, activity) [64–71], and (5) other mechanical contributors [72–77]. The G2C framework therefore seeks to provide a clinically viable method to generate gait-derived contact-exposure profiles that can guide targeted interventions (e.g., physiotherapy, alignment correction) and surveillance of high-risk patients before failure occurs.

As this is the first study utilizing deep learning to predict spatial wear distribution in total joint replacement settings, comparing our outcomes with existing studies is challenging. Nonetheless, analyzing the differences in metrics between the lateral and medial compartments still remains informative. Table 2 shows that for all grid setups, MAPE for the lateral compartment's width consistently exceeds that of the medial compartment. This is likely due to lower values of wear scar width for the lateral compartment, where even minor errors result in a significant relative increase in MAPE. MAPE for the length of the wear scar is comparable between the two compartments, while the angle between centroids is slightly greater for the non-weighted compared to the weighted wear scar. It is worth noting that such angles are quite small, with the average angle between centroids being respectively around 5° and 10° for the non-weighted and weighted wear scars. Therefore, errors around 15% in terms of MAPE correspond to slightly less than 1° for the non-weighted and slightly more than 1° for weighted wear scars, which may not be clinically relevant. MAPE values for both non-weighted and weighted areas of the wear scars are consistent across the compartments. However, the error for the weighted area is approximately 3×higher than that for the non-weighted area. These disparities may be linked to small differences in predicted linear wear between deep learning and FEA predictions rather than in the gross wear scar geometry, and may not represent meaningful differences. Therefore, we also considered metrics [38, 39] that specifically focus on heatmap patterns to directly assess the spatial fidelity between FEA and deep learning heatmaps (SSIM) and the shared information between the heatmaps predicted by these two methodologies (NMI). When evaluating both metrics, images were cropped to isolate either the lateral or medial compartment. This approach ensures that the metrics are not affected by regions of the heatmaps with zero values. Given that the values for both metrics ranged between 0.88 and 0.96, we can consider the predictive performance of the models to be satisfactory [38, 39].

The primary advantage of using deep learning over traditional FEA is the reduction in computational time required to determine the output linear wear distribution. Training of deep learning models required approximately 7 minutes for the 100x100 model and 11 minutes for the 200 × 200 model, and inference took just a few seconds on an A5000 GPU with 24 GB of VRAM. In contrast, the FEA model required significantly more time, taking approximately 8 hours to process each batch of up to 4 samples. With 314 samples in the dataset distributed across 79 batches, the FEA model required roughly 26 days on a workstation with two Xeon E5-2680 CPUs running at 2.40 GHz and 256 GB of RAM. Although our proposed methodology has not been applied to a clinical population, based on the results of this study, the proposed surrogate modeling framework appears to be a viable efficient approach for predicting wear patterns in clinical settings, and even in virtual clinical trials [78].

The current study is not without limitations. First, the proposed G2C deep learning methodology was applied to a previously published synthetic dataset [15], generated by modifying the ISO Standard waveforms [79], rather than on data from patients. Given the novelty of the current approach, we chose to test it on synthetic data generated through an extensively validated framework [15] as a first attempt. Having achieved promising results, next steps will involve applying and evaluating this methodology on a clinical dataset of TKR patients [80], utilizing a previously published deep learning approach to generate representative gait waveforms [81]. Second, the absence of component alignment data represents another limitation of this study. This issue could be addressed in future research by generating our training

dataset at different component alignments, and accounting for these additional inputs in our deep learning framework. Clinically, component alignment can be approximated from post-operative X-rays [82–84]. Third, although we propose an alternative approach to FEA for predicting wear distribution in TKR settings, a well validated FEA model is still initially needed to generate a high-fidelity training set for the deep learning model. Finally, while our model currently only accounts for level walking, other motion tasks beyond gait should be investigated to ensure that our deep learning framework accurately predicts wear maps across different activities of daily living. Some studies have explored the effects of different motion tasks on wear in TKR [16, 21], and employing our deep learning framework to these various tasks will be important for clinical application.

This study introduces a novel deep learning approach for predicting wear distribution in TKRs, utilizing kinetic and kinematic gait patterns. The dataset is comprised of synthetic data featuring variations of ISO14243-3(2014) anterior/posterior translation, internal/external rotation, flexion/extension, and axial loading time series waveforms. A transformer-CNN architecture was implemented to capture temporal dependencies within the input multivariate time series and to accurately reconstruct output image data. The deep learning model's predictions of length, width, area, and angle between wear scar centroids were evaluated using MAPE, while SSIM and NMI were employed to assess the overall similarity of wear maps between the deep learning and FEA predictions. Future work will involve applying and evaluating this methodology on patient data.

Funding No funding was received to assist with the preparation of this manuscript.

Declarations

Conflict of interest The authors have the following conflicts of interests to declare: MP—none; SS—none; PM—none; VK—consultant for Stryker and Corin, stock and/or royalties in Corin, receiving research and fellowship support from Corin and Stryker; CY—none; JM—none; SPM—none.

References

1. Feng, J. E., D. Novikov, A. A. Anoushiravani, and R. Schwarzkopf. Total knee arthroplasty: improving outcomes with a multidisciplinary approach. *J. Multidiscip. Healthc.* 11:63–73, 2018. https://doi.org/10.2147/JMDH.S140550.
2. Kurtz, S. M., K. L. Ong, E. Lau, and K. J. Bozic. Impact of the economic downturn on total joint replacement demand in the United States: Updated projections to 2021. *J. Bone Joint Surg. Am.* 96(8):624–630, 2014. https://doi.org/10.2106/JBJS.M.00285.
3. Weinstein, A. M., et al. Estimating the burden of total knee replacement in the United States. *J. Bone Joint Surg. Am.* 95(5):385–392, 2013. https://doi.org/10.2106/JBJS.L.00206.
4. Trieu, J., D. J. Gould, C. Schilling, T. Spelman, M. M. Dowsey, and P. F. Choong. Patient-reported outcomes following total knee replacement in patients <65 years of age—a systematic review and meta-analysis. *J. Clin. Med.* 9(10):1–17, 2020. https://doi.org/10.3390/jcm9103150.
5. Kurtz, S. M., E. Lau, K. Ong, K. Zhao, M. Kelly, and K. J. Bozic. Future young patient demand for primary and revision joint replacement: National projections from 2010 to 2030. *Clin. Orthop. Relat. Res.* 467(10):2606–2612, 2009. https://doi.org/10.1007/s11999-009-0834-6.
6. Hawker, G. A., E. Bohm, M. J. Dunbar, C. A. Jones, T. Noseworthy, and D. A. Marshall. The effect of patient age and surgical appropriateness and their influence on surgeon recommendations for primary TKA: a cross-sectional study of 2,037 patients. *J. Bone Joint Surg. Am.* 104(8):700–708, 2022. https://doi.org/10.2106/JBJS.21.00597.
7. Asher, D. P., et al. Is wear still a concern in total knee arthroplasty with contemporary conventional and highly crosslinked polyethylene tibial inserts in the mid- to long-term? *Arthroplast Today*.30:101550, 2024.
8. Siqueira, M. B. P., A. K. Klika, C. A. Higuera, and W. K. Barsoum. Modes of failure of total knee arthroplasty: registries and realities. *J. Knee Surg.* 28(2):127–138, 2015. https://doi.org/10.1055/s-0034-1396014.
9. Schroer, W. C., et al. Why are total knees failing today? Etiology of total knee revision in 2010 and 2011. *J. Arthroplasty*. 28(8 Suppl):116–119, 2013. https://doi.org/10.1016/j.arth.2013.04.056.
10. Koh, I. J., W. S. Cho, N. Y. Choi, and T. K. Kim. Causes, risk factors, and trends in failures after TKA in Korea over the past 5 years: A multicenter study. *Clin. Orthop. Relat. Res.* 472(1):316–326, 2014. https://doi.org/10.1007/s11999-013-3252-8.
11. Kerzner, B., K. N. Kunze, M. B. O'Sullivan, K. Pandher, and B. R. Levine. Temporal trends of revision etiologies in total knee arthroplasty at a single high-volume institution: an epidemiological analysis. *Arthroplast Today*. 9:68–72, 2021.
12. Sabah, S. A., et al. No exponential rise in revision knee replacement surgery over the past 15 years: an analysis from the National Joint Registry. *Osteoarthr. Cartil.* 30:1670–1679, 2022.
13. American Joint Replacement Registry. *American Joint Replacement Registry: 2024 Annual Report.* Rosemont, IL: American Academy of Orthopaedic Surgeons; 2024.
14. Wilson, J. M., et al. Is the American Joint Replacement Registry able to correctly classify revision total knee arthroplasty procedural diagnoses? *J. Arthroplast.* 38(Suppl):S32-S35.e3, 2023.
15. Mell, S. P., M. A. Wimmer, and H. J. Lundberg. Sensitivity of total knee replacement wear to variability in motion and load input: A parametric finite element analysis study. *J. Orthop. Res.* 38(7):1538–1549, 2020. https://doi.org/10.1002/jor.24755.
16. Abdel-Jaber, S., C. Belvedere, A. Leardini, and S. Affatato. Wear simulation of total knee prostheses using load and kinematics waveforms from stair climbing. *J. Biomech.* 48(14):3830–3836, 2015. https://doi.org/10.1016/j.jbiomech.2015.09.007.
17. Shu, L., S. Hashimoto, and N. Sugita. Enhanced in-silico polyethylene wear simulation of total knee replacements during daily activities. *Ann. Biomed. Eng.* 49(1):322–333, 2021. https://doi.org/10.1007/s10439-020-02555-4.
18. Mell, S. P., M. A. Wimmer, J. J. Jacobs, and H. J. Lundberg. Optimal surgical component alignment minimizes TKR wear—an in silico study with nine alignment parameters. *J. Mech. Behav. Biomed. Mater.*125:104939, 2022. https://doi.org/10.1016/j.jmbbm.2021.104939.
19. Phellan, R., B. Hachem, J. Clin, J. M. Mac-Thiong, and L. Duong. Real-time biomechanics using the finite element method and machine learning: Review and perspective. *Med. Phys.* 48(1):7–18, 2021. https://doi.org/10.1002/mp.14602.

20. Kudela, J., and R. Matousek. Recent advances and applications of surrogate models for finite element method computations: a review. *Soft Comput.* 26(24):13709–13733, 2022. https://doi.org/10.1007/s00500-022-07362-8.
21. Mell, S. P., M. A. Wimmer, and H. J. Lundberg. The choice of the femoral center of rotation affects material loss in total knee replacement wear testing—a parametric finite element study of ISO 14243–3. *J. Biomech.* 88:104–112, 2019. https://doi.org/10.1016/j.jbiomech.2019.03.027.
22. Pfeiffer, M., C. Riediger, J. Weitz, and S. Speidel. Learning soft tissue behavior of organs for surgical navigation with convolutional neural networks. *Int. J. Comput. Assist. Radiol. Surg.* 14(7):1147–1155, 2019. https://doi.org/10.1007/s11548-019-01965-7.
23. Mendizabal, A., P. Márquez-Neila, and S. Cotin. Simulation of hyperelastic materials in real-time using deep learning. *Med. Image Anal.*59:101569, 2020. https://doi.org/10.1016/j.media.2019.101569.
24. Liu, M., L. Liang, and W. Sun. Estimation of in vivo constitutive parameters of the aortic wall using a machine learning approach. *Comput. Methods Appl. Mech. Eng.* 347:201–217, 2019. https://doi.org/10.1016/j.cma.2018.12.030.
25. Maag, C., C. K. Fitzpatrick, and P. J. Rullkoetter. Evaluation of machine learning techniques for real-time prediction of implanted lower limb mechanics. *Front. Bioeng. Biotechnol.* 12:1461768, 2025.
26. Nimmal Haribabu, G. N., and B. Basu. Implementing machine learning approaches for accelerated prediction of bone strain in acetabulum of a hip joint. *J. Mech. Behav. Biomed. Mater.*153:106495, 2024. https://doi.org/10.1016/j.jmbbm.2024.106495.
27. Lu, Y., P. R. Pulasani, R. Derakhshani, and T. M. Guess. Application of neural networks for the prediction of cartilage stress in a musculoskeletal system. *Biomed. Signal Process Control.* 8(6):475–482, 2013. https://doi.org/10.1016/j.bspc.2013.04.004.
28. McKay, M. D., R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 42(1):55–61, 2000. https://doi.org/10.1080/00401706.2000.10485979.
29. Mell SP, Fullam S, Wimmer MA, Lundberg HJ. Computational parametric studies for preclinical evaluation of total knee replacements. In: *Computer Methods, Imaging and Visualization in Biomechanics and Biomedical Engineering: Selected Papers from the 16th International Symposium CMBBE and 4th Conference on Imaging and Visualization,* New York City, USA, Aug 14–16, 2019. Cham: Springer; 2020. pp. 60–85.
30. Vaswani A. Attention is all you need. *Adv. Neural Inf. Process. Syst.*; 2017.
31. Pan, S., X. Liu, N. Xie, and Y. Chong. EG-TransUNet: a transformer-based U-Net with enhanced and guided models for biomedical image segmentation. *BMC Bioinform.* 24(1):1–22, 2023. https://doi.org/10.1186/s12859-023-05196-1.
32. He Y, Nath V, Yang D, Tang Y, Myronenko A, Xu D. SwinUNETR-V2: Stronger Swin Transformers with stagewise convolutions for 3D medical image segmentation. In: *Lect Notes Comput Sci* 14223. Springer Nature Switzerland; 2023.
33. Akiba, T., S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: a next-generation hyperparameter optimization framework. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 2019. https://doi.org/10.1145/3292500.3330701.
34. Bergstra, J., R. Bardenet, Y. Bengio, and B. Kégl. Algorithms for hyper-parameter optimization. *Adv. Neural Inf. Process Syst.* 24:1–9, 2011.
35. Dammer, R. H., C. Zietz, and R. Bader. A comparison of wear patterns on retrieved and simulator-tested total knee replacements. *J. Funct. Biomater.* 2022. https://doi.org/10.3390/jfb13040256.
36. Knowlton, C. B., P. Bhutani, and M. A. Wimmer. Relationship of surface damage appearance and volumetric wear in retrieved TKR polyethylene liners. *J. Biomed. Mater. Res. B Appl. Biomater.* 105(7):2053–2059, 2017. https://doi.org/10.1002/jbm.b.33684.
37. Grupp, T. M., C. Kaddick, J. Schwiesau, A. Maas, and S. D. Stulberg. Fixed and mobile bearing total knee arthroplasty—influence on wear generation, corresponding wear areas, knee kinematics and particle composition. *Clin. Biomech.* 24(2):210–217, 2009. https://doi.org/10.1016/j.clinbiomech.2008.11.006.
38. Sara, U., M. Akter, and M. S. Uddin. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. *J. Comput. Commun.* 7(3):8–18, 2019. https://doi.org/10.4236/jcc.2019.73002.
39. Sebastian, A. P., G. Rohith, and L. S. Kumar. Significant full reference image segmentation evaluation: a survey in remote sensing field. *Multimed. Tools Appl.* 81(13):17959–17987, 2022. https://doi.org/10.1007/s11042-022-12769-4.
40. Parati, M., et al. Video-based goniometer applications for measuring knee joint angles during walking in neurological patients: a validity, reliability and usability study. *Sensors (Basel)*. 2023. https://doi.org/10.3390/s23042232.
41. Versteyhe, M., H. De Vroey, F. Debrouwere, H. Hallez, and K. Claeys. A novel method to estimate the full knee joint kinematics using low cost IMU sensors for easy to implement low cost diagnostics. *Sensors (Basel)*. 2020. https://doi.org/10.3390/s20061683.
42. Hullfish, T. J., F. Qu, B. D. Stoeckl, P. M. Gebhard, R. L. Mauck, and J. R. Baxter. Measuring clinically relevant knee motion with a self-calibrated wearable sensor. *J. Biomech.* 89:105–109, 2019. https://doi.org/10.1016/j.jbiomech.2019.04.003.
43. Bergmann, G., et al. Standardized loads acting in knee implants. *PLoS One*. 2014. https://doi.org/10.1371/journal.pone.0086035.
44. Heinlein, B., F. Graichen, A. Bender, A. Rohlmann, and G. Bergmann. Design, calibration and pre-clinical testing of an instrumented tibial tray. *J. Biomech.* 40(Suppl 1):S4, 2007. https://doi.org/10.1016/j.jbiomech.2007.02.014.
45. D'Lima, D. D., S. Patil, N. Steklov, S. Chien, and C. W. Colwell. In vivo knee moments and shear after total knee arthroplasty. *J. Biomech.* 2007. https://doi.org/10.1016/j.jbiomech.2007.03.004.
46. Kia, M., A. P. Stylianou, and T. M. Guess. Evaluation of a musculoskeletal model with prosthetic knee through six experimental gait trials. *Med. Eng. Phys.* 36(3):335–344, 2014. https://doi.org/10.1016/j.medengphy.2013.12.007.
47. Navacchia, A., P. J. Rullkoetter, P. Schütz, R. B. List, C. K. Fitzpatrick, and K. B. Shelburne. Subject-specific modeling of muscle force and knee contact in total knee arthroplasty. *J. Orthop. Res.* 34(9):1576–1587, 2016. https://doi.org/10.1002/jor.23171.
48. Perrone, M., M. Guidetti, M. Galli, S. J. Nho, M. A. Wimmer, and P. Malloy. Hip joint contact forces are lower in people with femoroacetabular impingement syndrome during squat tasks. *J. Orthop. Res.* 2023. https://doi.org/10.1002/jor.25744.
49. D'Antonio, E., J. Taborri, E. Palermo, S. Rossi, and F. Patane. A markerless system for gait analysis based on OpenPose library. *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (12MTC)*. 2020. https://doi.org/10.1109/I2MTC43012.2020.9128918.
50. D'Antonio, E., J. Taborri, I. Mileti, S. Rossi, and F. Patane. Validation of a 3D markerless system for gait analysis based on OpenPose and two RGB webcams. *IEEE Sens. J.* 21(15):17064–17075, 2021. https://doi.org/10.1109/JSEN.2021.3081188.
51. Moro, M., G. Marchesi, F. Hesse, F. Odone, and M. Casadio. Markerless vs marker-based gait analysis: a proof of concept study. *Sensors (Basel)*. 22(5):1–15, 2022. https://doi.org/10.3390/s22052011.

52. Uhlrich, S. D., et al. OpenCap: human movement dynamics from smartphone videos. *PLoS Comput. Biol.* 19(10):1–26, 2023. https://doi.org/10.1371/journal.pcbi.1011462.
53. Ardestani, M. M., M. Moazen, and Z. Jin. Contribution of geometric design parameters to knee implant performance: conflicting impact of conformity on kinematics and contact mechanics. *Knee*. 22:217–224, 2015.
54. Bei, Y., B. J. Fregly, W. G. Sawyer, S. A. Banks, and N. H. Kim. The relationship between contact pressure, insert thickness, and mild wear in total knee replacements. *Comput. Model Eng. Sci.* 6:145–152, 2004.
55. Werner, F. W., D. C. Ayers, L. P. Maletsky, and P. J. Rullkoetter. The effect of valgus/varus malalignment on load distribution in total knee replacements. *J. Biomech.* 38:349–355, 2005.
56. Zheng, Z., et al. Medial–lateral translational malalignment of the prosthesis on tibial stress distribution in total knee arthroplasty: a finite element analysis. *Front. Bioeng. Biotechnol.* 11:1119204, 2023.
57. Babazadeh, S., J. D. Stoney, K. Lim, and P. F. M. Choong. The relevance of ligament balancing in total knee arthroplasty: how important is it? A systematic review of the literature. *Orthop. Rev. (Pavia)*. 1:126, 2009.
58. Roche, M., L. Elson, and C. Anderson. Dynamic soft-tissue balancing in total knee arthroplasty. *Orthop. Clin. North Am.* 45:157–165, 2014.
59. Siddiqi, A., et al. Soft-tissue balancing technology for total knee arthroplasty. *JBJS Rev.*8:e0050, 2020.
60. Martin, J. R., C. D. Watts, D. L. Levy, and R. H. Kim. Medial tibial stress shielding: a limitation of cobalt chromium tibial baseplates. *J. Arthroplast.* 32:558–562, 2017.
61. Petersen, M. M., P. T. Nielsen, J. B. Lauritzen, and B. Lund. Changes in bone mineral density of the proximal tibia after uncemented total knee arthroplasty: a 3-year follow-up of 25 knees. *Acta Orthop. Scand.* 66:513–516, 1995.
62. Jaroma, A., T. Soininvaara, and H. Kröger. Periprosthetic tibial bone mineral density changes after total knee arthroplasty. *Acta Orthop.* 87:268–273, 2016.
63. Anijs, T., et al. Computational tibial bone remodeling over a population after total knee arthroplasty: a comparative study. *J. Biomed. Mater. Res. B Appl. Biomater.* 110:776–786, 2022.
64. Varadarajan, K. M., A. L. Moynihan, D. D'Lima, C. W. Colwell, and G. Li. In vivo contact kinematics and contact forces of the knee after total knee arthroplasty during dynamic weight-bearing activities. *J. Biomech.* 41:2159–2168, 2008.
65. Hamilton, M. A., M. C. Sucec, B. J. Fregly, S. A. Banks, and W. G. Sawyer. Quantifying multidirectional sliding motions in total knee replacements. *J. Tribol.* 127:280–286, 2005.
66. Arauz, P., Y. Peng, S. An, and Y.-M. Kwon. In vivo analysis of sliding distance and cross-shear in bicruciate-retaining total knee arthroplasty. *J. Biomech.* 77:8–15, 2018.
67. Lundberg, H. J., A. Swanson, C. Knowlton, N. Inoue, and M. A. Wimmer. Methods for locating the tibio-femoral contact pathway in total knee replacements using marker-based gait analysis and standard radiography. *Iowa Orthop. J.* 34:94, 2014.
68. Waldman, S. D., and J. T. Bryant. Dynamic contact stress and rolling resistance model for total knee arthroplasties. *J. Biomech. Eng.* 119:254–260, 1997.
69. Xie, X., et al. Effect of rotational prosthetic alignment variation on tibiofemoral contact pressure distribution and joint kinematics in total knee replacement. *Proc. Inst. Mech. Eng. H.* 231:1034–1047, 2017.
70. Smith, C. R., M. F. Vignos, R. L. Lenhart, J. Kaiser, and D. G. Thelen. The influence of component alignment and ligament properties on tibiofemoral contact forces in total knee replacement. *J. Biomech. Eng.* 138:021017, 2016.
71. Rothammer, B., et al. Subject-specific tribo-contact conditions in total knee replacements: a simulation framework across scales. *Biomech. Model Mechanobiol.* 22:1395–1410, 2023.
72. Wilson, C. J., A. Theodoulou, R. A. Damarell, and J. Krishnan. Knee instability as the primary cause of failure following total knee arthroplasty: a systematic review on the patient, surgical and implant characteristics of revised TKA patients. *Knee*. 24:1271–1281, 2017.
73. Meehan, J. P., B. Danielsen, S. H. Kim, A. A. Jamali, and R. H. White. Younger age is associated with a higher risk of early periprosthetic joint infection and aseptic mechanical failure after total knee arthroplasty. *J. Bone Joint Surg. Am.* 96:529–535, 2014.
74. Torino, D., et al. Tibial baseplate–cement interface debonding in the ATTUNE total knee arthroplasty system. *Arthroplast. Today*. 17:165–171, 2022.
75. Keohane, D., G. A. Sheridan, and E. Masterson. High rate of tibial debonding and failure in a popular knee replacement: a follow-up review. *Bone Joint Open*. 3:495–501, 2022.
76. Sadauskas, A., C. Engh, M. Mehta, and B. Levine. Implant interface debonding after total knee arthroplasty: a new cause for concern? *Arthroplast. Today*. 6:972–975, 2020.
77. Keohane, D., F. Power, E. Cullen, A. O'Neill, and E. Masterson. High rate of tibial debonding and failure in a popular knee replacement: a cause for concern. *Knee*. 27:459–468, 2020.
78. Mell, S. P., A. L. Hornung, C. Yuh, and D. Samartzis. Virtual clinical trials—implications of computer simulations and artificial intelligence for musculoskeletal research. *J. Bone Joint Surg. Am.* 2024. https://doi.org/10.2106/JBJS.23.01236.
79. International Organization for Standardization. *ISO 14243-3:2014—Implants for surgery—Wear of total knee-joint prostheses. Part 3: Loading and displacement parameters for wear-testing machines with displacement control and corresponding environmental conditions for test.* Geneva: ISO; 2014.
80. Perrone, M., S. P. Mell, J. T. Martin, S. J. Nho, S. Simmons, and P. Malloy. Synthetic data generation in motion analysis: a generative deep learning framework. *Proc. Inst. Mech. Eng. H*. 239:202–211, 2025.
81. Ngai, V., and M. A. Wimmer. Variability of TKR knee kinematics and relationship with gait kinetics: Implications for total knee wear. *Biomed. Res. Int.* 2015. https://doi.org/10.1155/2015/284513.
82. Hirschmann, M. T., P. Konala, F. Amsler, F. Iranpour, N. F. Friederich, and J. P. Cobb. The position and orientation of total knee replacement components: A comparison of conventional radiographs, transverse 2D-CT slices and 3D-CT reconstruction. *J. Bone Joint Surg. Br.* 93-B(5):629–633, 2011. https://doi.org/10.1302/0301-620X.93B5.25893.
83. Kumar, N., C. Yadav, R. Raj, and S. Anand. How to interpret postoperative X-rays after total knee arthroplasty. *Orthop. Surg.* 6(3):179–186, 2014. https://doi.org/10.1111/os.12123.
84. Petersen, T. L., and G. A. Engh. Radiographic assessment of knee alignment after total knee arthroplasty. *J. Arthroplast.* 3(1):67–72, 1988.