applied sciences

MDPI

Article

Preoperative Prediction of Optimal Femoral Implant Size by Regularized Regression on 3D Femoral Bone Shape

Adriaan Lambrechts [1,*], Christophe Van Dijck [1], Roel Wirix-Speetjens [1], Jos Vander Sloten [2], Frederik Maes [3,4] and Sabine Van Huffel [5]

1 Materialise NV, 3001 Leuven, Belgium
2 Biomechanics Section, Mechanical Engineering, KU Leuven, 3001 Leuven, Belgium
3 Processing Speech and Images (PSI), Department of Electrical Engineering (ESAT), KU Leuven, 3001 Leuven, Belgium
4 Medical Imaging Research Center, UZ Leuven, 3000 Leuven, Belgium
5 STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Department of Electrical Engineering (ESAT), KU Leuven, 3001 Leuven, Belgium
* Correspondence: adriaan.lambrechts@materialise.be

Abstract: Preoperative determination of implant size for total knee arthroplasty surgery has numerous clinical and logistical benefits. Currently, surgeons use X-ray-based templating to estimate implant size, but this method has low accuracy. Our study aims to improve accuracy by developing a machine learning approach that predicts the required implant size based on a 3D femoral bone mesh, the key factor in determining the correct implant size. A linear regression framework imposing group sparsity on the 3D bone mesh vertex coordinates was proposed based on a dataset of 446 MRI scans. The group sparse regression method was further regularized based on the connectivity of the bone mesh to enforce neighbouring vertices to have similar importance to the model. Our hypergraph regularized group lasso had an accuracy of 70.1% in predicting femoral implant size while the initial implant size prediction provided by the instrumentation manufacturer to the surgeon has an accuracy of 23.1%. Furthermore, our method was capable of predicting the implant size up to one size smaller or larger with an accuracy of 99.1%, thereby surpassing other state-of-the-art methods. The hypergraph regularized group lasso was able to obtain a significantly higher accuracy compared to the implant size prediction provided by the instrumentation manufacturer.

Keywords: total knee arthroplasty; templating; machine learning; group lasso

Citation: Lambrechts, A.; Van Dijck, C.; Wirix-Speetjens, R.; Vander Sloten, J.; Maes, F.; Van Huffel, S. Preoperative Prediction of Optimal Femoral Implant Size by Regularized Regression on 3D Femoral Bone Shape. *Appl. Sci.* **2023**, *13*, 4344. https://doi.org/10.3390/app13074344

Academic Editor: Valeriu Surlin

Received: 28 February 2023
Revised: 20 March 2023
Accepted: 23 March 2023
Published: 29 March 2023

1. Introduction

Knee osteoarthritis is a degenerative disease affecting the knee cartilage layers. Due to progressive knee cartilage wear, patients can experience pain, instability, loss of flexibility, joint stiffness and swelling. For patients with severe knee osteoarthritis, a total knee arthroplasty (TKA) surgery can help improve quality of life [1]. During a TKA procedure the femur and tibial bones are resected along the joint interface and replaced by metal components. Two-dimensional templating is the most frequently used method to preoperatively determine the required implant size. In 2D templating, anteroposterior and lateral radiographs are overlaid with 2D template shapes of the implants to determine the best fitting size. Despite its broad adoption, 2D templating has relatively low accuracy in determining the required size due to the nature of projection images [2–5]. Three-dimensional templating based on computed tomography (CT) or magnetic resonance imaging (MRI) scans can overcome this limitation of 2D templating resulting in higher templating accuracy [6–8]. Three-dimensional templating allows the surgeon to position the implants on a 3D model of the joint to determine the optimal implant size and position.

There are several benefits to preoperatively planning the implant size. First of all, determining the correct implant sizes is of importance from a clinical point of view. Femoral oversizing might lead to irritation of the soft-tissue structures surrounding the joint, increased patellofemoral pressure or reduced range of motion and, hence, should be avoided [9]. On the other hand, femoral undersizing can lead to femoral fracture due to anterior implant notching, flexion instability, or inferior patellar tracking due to a medialized implant placement [10]. Secondly, the extremely large and small implant sizes need to be specifically ordered as they are not consistently present in the hospital [11]. Furthermore, knowing the implant size up to a size smaller or larger reduces the sterilization cost and operation room setup time [12,13].

Over recent decades, machine learning has been successfully applied to many medical applications, including implant size templating [14–16]. Several studies have investigated implant size prediction based on demographic data [2,11,16–22]. The method by Trainor et al. used only shoe size and gender to obtain the highest reported accuracy of 63% and an accuracy of 99% for predicting up to one size larger or smaller [11]. Finally, Lambrechts et al. created a support vector machine based model that predicts implant size based on manually defined features deduced from 3D bone models, resulting in an absolute accuracy of 82.2%, significantly higher compared to models relying on demographic data [23].

Traditional methods for determining implant size are based on human predefined measurements. However, surgeons base the implant size on the size and morphology of the bone. Therefore, we hypothesize that shape based implant size prediction can yield higher accuracy compared to previously investigated methods. To the best of our knowledge, our method is the first to regress the implant size directly based on the 3D model of the bone. The computational method most similar to the one presented in this study, is an optimal-scoring-based method by Clemmensen et al. that induces sparsity [24]. This algorithm was evaluated on a classification task distinguishing male and female face silhouettes based on the point coordinates. The method by Clemmensen et al. obtains interpretability by introducing sparsity through the use of l_1 regularization on the model coefficients corresponding to single coordinates of certain points. As a result, this method can only find directions of shape variations that are parallel to the coordinate system axes which is a major disadvantage. A second limitation is that proximity information is not incorporated into the model as neighbouring points in space should have similar contribution to the model.

Therefore, we propose a method that extends the method by Clemmensen et al. in multiple ways. Firstly, our method is targeted toward triangular meshes allowing better incorporation of spatial structure compared to point clouds. Secondly, group lasso regularization was used to obtain directions of variation in bone shape which are predictive of our target. As opposed to the method by Clemmensen et al., our method allows these directions of variation to be in any direction, not only parallel to coordinate axes. Thirdly, connectivity of the mesh structure is included in our model as extra prior information.

2. Materials and Methods

2.1. Data Preprocessing

For our model we used a dataset of 446 TKA preoperative plans retrospectively collected. All cases were planned by a single surgeon based on an MRI scan of the patient's knee joint. MRI scans have as benefit the visualization of the knee bone, as well as cartilage which can aid the surgeon during 3D templating. The MRI scan protocol is described in Table 1. The scans were semi-automatically segmented (Mimics, Materialise, Belgium) by experienced conversion engineers. The segmented mask contains both the bone and the cartilage of femur. Next, these masks were converted to a triangulated surface mesh based on the marching cubes algorithm [25]. For each patient the corresponding femoral implant size planned by the surgeon was also available. All patients received a posterior-stabilized Zimmer-Biomet Vanguard implant, which has 10 possible sizes.

Table 1. Description of the MRI scan protocol.

Parameter	Value
Scanner	GE Optima^{TM} MR450w
Field strength	1.5 T
Scan type	3D
Scan direction	Sagittal
Sequence	Fat saturated T1 spoiled gradient echo
Slice thickness	1 mm
Pixel size	0.4 mm

Our group lasso method relies on meshes which have point correspondences and identical adjacency matrices. To achieve this, we employ a previously developed statistical shape model (SSM) which was created based on a training set of 524 MRI scans independent from the dataset used for training the hypergraph regularized group lasso [26]. The SSM was fitted to all bones in the dataset to obtain point correspondence. Next, a non-rigid surface registration is applied to register the fitted model to the target femur [27]. Finally the vertex coordinates of the registered femur meshes are stacked in a matrix $\mathbf{X} \in \mathbb{R}^{N \times 3p}$, where N indicates the number of femur meshes in the dataset, and p the number of points in the mesh. Finally, all columns of the data matrix $\mathbf{X}$ are scaled to have a mean of zero and a standard deviation of 1. The implant sizes are coded as an ordinal variable with a range from 1 to 10, represented by $\mathbf{y}$. Based on a stratified split, 70% of the samples are used for creating the model and the remaining samples form the test set to measure the final accuracy of the model.

2.2. Hypergraph Representation of a Triangular Mesh

A triangular mesh can be represented as a simple undirected graph characterized by a set of vertices and edges: $\mathcal{G} = (V, E)$. Figure 1 represents a femur mesh with edges represented as black lines. Simple graphs can be represented by a Laplacian matrix $\mathbf{L} \in \mathbb{R}^{p \times p}$, defined as the difference of the adjacency and degree matrices. The Laplacian matrix has useful properties for analysing the structure of a graph. However, the simple graph structure of a mesh is not compatible with the regression framework proposed in this study. Therefore, the mesh will be represented by a hypergraph, a generalization of a graph. Hypergraphs have hyperedges which can connect any number of vertices. In a simple graph, each vertex represents a point in 3D space, while for the hypergraph we split each of these points into three vertices: one for each coordinate (Figure 2). All edges of the original graph become hyperedges in the hypergraph connecting 6 vertices, hence it is a 6-uniform hypergraph.

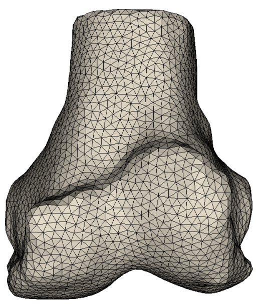
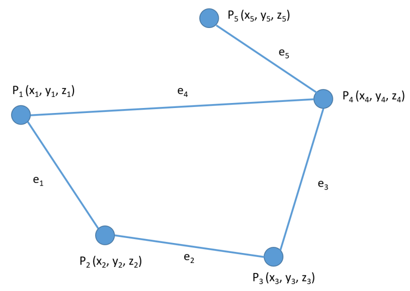
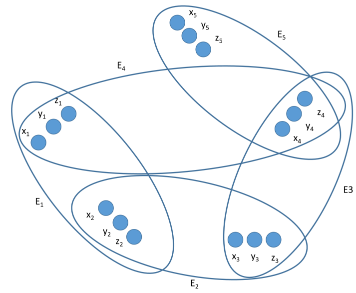


Figure 1. A mesh of the distal femur.



(**a**)



(**b**)

Figure 2. Graph and hypergraph structure for a mesh in 3D space. (**a**) Graph structure. (**b**) Hypergraph structure.

For hypergraphs, the incidence matrix $\mathbf{H} \in \mathbb{R}^{|V| \times |E|}$ identifies which vertices are part of a hyperedge:

$$\mathbf{H}_{ij} = \begin{cases} 1 & \text{if } v_i \in E_j \\ 0 & \text{if } v_i \notin E_j \end{cases} \tag{1}$$

Similarly, two diagonal degree matrices can be defined: the edge degree matrix $\mathbf{D}_e \in \mathbb{R}^{|E| \times |E|}$ and the vertex degree matrix $\mathbf{D}_v \in \mathbb{R}^{|V| \times |V|}$.

$$\mathbf{D}_{\mathbf{e}ij} = \begin{cases} \sum_{k=1}^{|V|} \mathbf{H}_{ki} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad \mathbf{D}_{\mathbf{v}ij} = \begin{cases} \sum_{k=1}^{|E|} \mathbf{H}_{ik} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Based on these matrices, the hypergraph Laplacian can be defined as:

$$\mathbf{L}_H = \mathbf{D}_v - \mathbf{H}\mathbf{D}_e^{-1}\mathbf{H}^T \tag{3}$$

2.3. Hypergraph Regularized Group Lasso

The target of the hypergraph regularized group lasso is to obtain a subset of points on the femur mesh, the coordinates of which can be linearly combined to predict the femur implant size. This can be accomplished by the following optimization problem:

$$\min_{\beta} \quad \frac{1}{2}\|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda_G \sum_{\mathbf{g} \in P} \|\beta_{\mathbf{g}}\|_2 + \frac{\lambda_L}{2}\beta^T \mathbf{L}_H \beta + \frac{\lambda_R}{2}\beta^T \mathbf{I}\beta \tag{4}$$

The objective function has four terms. The first term is a least squares term trying to represent the implant size $\mathbf{y}$ as a linear combination of the vertex locations of the mesh, parameterized by β. The parameter vector β indicates the importance of each vertex's coordinates in predicting the femoral implant size. The second term is a group lasso regularization term which results in sparsity in the points selected. P is the set of 3-tuples in which each tuple $\mathbf{g}$ references the column indices corresponding to the group of $\{x, y, z\}$ coordinates of a point on the mesh. $\beta_{\mathbf{g}} \in \mathbb{R}^3$ is a set of 3 of the model parameters corresponding with the $\{x, y, z\}$ coordinates from point $p_{\mathbf{g}}$. This regularization forces either all coordinates of a point to be included in the model or all coordinates of the point are found to be uninformative to the model, in which case they receive a zero-valued $\beta_{\mathbf{g}}$ coefficient. As a result, the group sparsity term creates model interpretability by expressing the femoral size as a linear combination of a small subset of points on the mesh. The third term forces the model coefficients of vertices linked by an edge, to be similar in value. We opted to use this penalty term because our hypothesis is that vertices linked by an edge are nearby in space, hence they should have a similar influence on the model. The last term is a ridge penalty to obtain uniqueness of the solution. The group lasso itself does not necessarily result in a unique solution because it is not strictly convex [28]. The addition of the ridge regularization leads to a strictly convex function guaranteeing a unique solution.

The group lasso problem is a convex non-differentiable problem as it is discontinuous at $\beta_{\mathbf{g}} = \mathbf{0}$. For these types of problems computationally efficient solvers exist based on the proximal gradient method which alternatively optimize each group [29]. Therefore, the objective function should be group separable. However, the hypergraph regularization term is not. Nevertheless, the objective function can be rewritten as a normal group lasso problem which can be solved using the proximal gradient method.

By combining the ridge penalty with the hyper-graph Laplacian regularization, the objective function can be simplified. More specifically, the weighted sum of the identity matrix and the hypergraph Laplacian is a positive definite matrix. Hence, it can be decomposed using a Cholesky decomposition [30]: $\frac{\lambda_R}{\lambda_L}\mathbf{I} + \mathbf{L}_H = \mathbf{U}^T\mathbf{U}$. Using the upper triangular matrix $\mathbf{U}$ we augment the feature matrix $\mathbf{X}$ and target vector $\mathbf{y}$ as follows [31]:

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \sqrt{\lambda_L}\mathbf{U} \end{bmatrix} \in \mathbb{R}^{(N+3p)\times 3p} \tag{5}$$

$$\widetilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0_{3p}} \end{bmatrix} \in \mathbb{R}^{N+3p} \tag{6}$$

$\mathbf{0_{3p}}$ is a column vector with $3p$ zero values. Using this augmented data matrix and target vector, the optimization problem can be reformulated as a standard group lasso, which, in turn, is fully group-separable:

$$\min_{\beta} \quad \frac{1}{2}\|\widetilde{\mathbf{y}} - \sum_{\mathbf{g}\in P} \widetilde{\mathbf{X}}_{\mathbf{g}}\beta_{\mathbf{g}}\|_2^2 + \lambda_G \sum_{\mathbf{g}\in P} \|\beta_{\mathbf{g}}\|_2 \tag{7}$$

Tibshirani et al. proposed a proximal gradient solving method to find the optimal solution for this convex problem [29]. The algorithm cycles over each group $\mathbf{g}$ and iteratively solves three equations until convergence. This inner iterative loop can be accelerated with the Nesterov acceleration method [32]:

$$\begin{aligned} \mathbf{r} &:= \widetilde{\mathbf{y}} - \sum_{\mathbf{i}\neq\mathbf{g}} \widetilde{\mathbf{X}}_{\mathbf{i}}\beta_{\mathbf{i}}^0 \\ \omega_{\mathbf{g}}^{t+1} &= \beta_{\mathbf{g}}^t + \nu\widetilde{\mathbf{X}}_{\mathbf{g}}^T(\mathbf{r}_{\mathbf{g}} - \widetilde{\mathbf{X}}_{\mathbf{g}}\beta_{\mathbf{g}}^t) \\ \beta_{\mathbf{g}}^{t+1} &= \left(1 - \frac{\nu\lambda_G}{\|\omega_{\mathbf{g}}^t\|_2}\right)_+ \omega_{\mathbf{g}}^{t+1} \\ \beta_{\mathbf{g}}^{t+1} &:= \beta_{\mathbf{g}}^{t+1} + \frac{t}{t+3}(\beta_{\mathbf{g}}^{t+1} - \beta_{\mathbf{g}}^t) \end{aligned} \tag{8}$$

The ν parameter is the step size in the optimization and should be carefully set to converge sufficiently fast while avoiding divergence from the optimum. $(q)_+ := max\{0, q\}$ is the positive part function. In Equation (8), the superscripts indicate the iteration count. The complete algorithm is described in Algorithm 1. Mathematical operations, such as additions, norm calculation, and matrix-vector products were performed using compressed sparse column operations to reduce both computational time and memory footprint [33]. The algorithm's accuracy is determined by three parameters λ_G, λ_R, and λ_L, which control the relative strength of the three regularization terms. Larger values of λ_G improve the interpretability by reducing the number of points in the final solution. Larger values of λ_L improve smoothness of the coefficient values across neighbouring points on the mesh. Finally, larger values of λ_R, generally avoid overfitting of the model by reducing the values of the model parameters. To determine the optimal values for these three parameters, a 3D logarithmic grid search approach was combined with five-fold cross validation on the training set. The range of λ_G, λ_R and λ_L started at 10^{-6} and was increased by a factor of 10 until the value of 1000 was reached.

Algorithm 1: Hypergraph regularized group Lasso.

Input: $\mathbf{X}$, $\mathbf{y}$, $\mathbf{L}_H$, λ_G, λ_L, λ_R, ν, tol, max_iter
Output: β
Sparse Cholesky decomposition: $\mathbf{L_H} + \frac{\lambda_R}{\lambda_L}\mathbf{I} \rightarrow \mathbf{U}^T\mathbf{U}$;
$\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \sqrt{\lambda_L}\mathbf{U} \end{bmatrix} \quad \widetilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0_{3p}} \end{bmatrix}$;
Initialize $\beta \leftarrow \mathcal{U}(-\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}})$;
$i \leftarrow 0$;
while *not converged and* $i <$ *max_iter* **do**
    $\mathbf{r} \leftarrow \widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}\beta$;
    for $\mathbf{g} \in P$ **do**
        $\mathbf{r_g} \leftarrow \mathbf{r} + \widetilde{\mathbf{X}}_\mathbf{g}\beta^0_\mathbf{g}$;
        $t \leftarrow 0$;
        while $\|\beta^t_\mathbf{g} - \beta^{t-1}_\mathbf{g}\| < tol$ **do**
            $t \leftarrow t + 1$;
            $\theta^t_\mathbf{g} \leftarrow \beta^{t-1}_\mathbf{g} + \frac{t}{t+3}(\beta^{t-1}_\mathbf{g} - \beta^{t-2}_\mathbf{g})$;
            $\omega^t_\mathbf{g} \leftarrow \theta^t_\mathbf{g} + \nu\widetilde{\mathbf{X}}^T_\mathbf{g}(\mathbf{r_g} - \widetilde{\mathbf{X}}_\mathbf{g}\theta^t_\mathbf{g})$;
            $\beta^t_\mathbf{g} \leftarrow \left(1 - \frac{\nu\lambda_G}{\|\omega^t_\mathbf{g}\|_2}\right)_+ \omega^t_\mathbf{g}$;
        end
        $\mathbf{r} \leftarrow \mathbf{r_g} - \widetilde{\mathbf{X}}_\mathbf{g}(\beta^t_\mathbf{g} - \beta^0_\mathbf{g})$;
        $i \leftarrow i + 1$;
    end
end

2.4. Baseline Method

To establish a baseline, our method was compared against statistical shape mode coefficient-based linear regression. While fitting the SSM to each of the femurs in the dataset, the shape mode coefficients are obtained. These coefficients quantify the presence of the uncorrelated modes of shape variation. For example, the first shape mode variation is generally a measure for bone volume and, hence, should correlate with the required implant size. Using an elastic net, the shape mode coefficients were regressed to the femoral implant size [31]. The hyperparameters of the elastic net were tuned using grid search during a five-fold cross validation procedure.

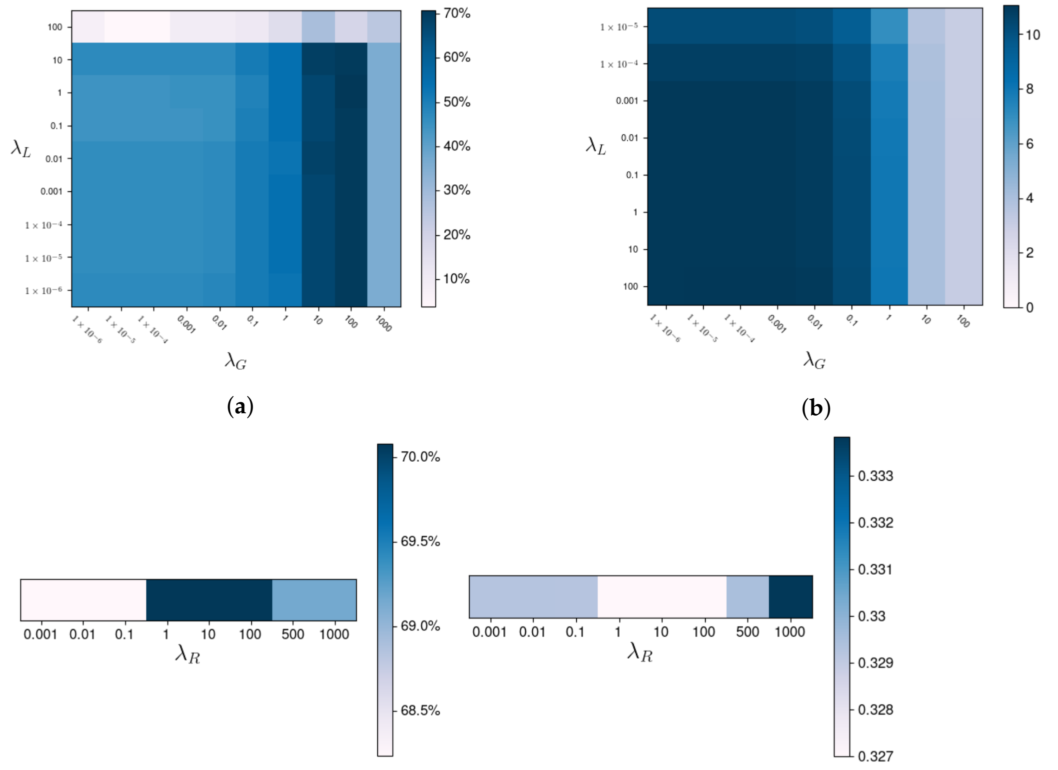
3. Results

All meshes in the dataset had 22,476 vertices after SSM fitting and non-linear registration resulting in a data matrix $\mathbf{X} \in \mathbb{R}^{312\times67428}$. The data matrix is augmented with the upper triangular matrix from the Cholesky decomposition resulting in a data matrix of size $\widetilde{\mathbf{X}} \in \mathbb{R}^{67.740\times67.428}$ with 88% sparsity. Owing to the sparsity of the problem, it becomes feasible to solve the hypergraph regularized group lasso efficiently through the use of sparse matrix and vector calculations.

The algorithm performance in terms of average accuracy of femoral implant size prediction $\hat{\mathbf{y}}$ and average mean squared error (MSE) over all folds, during the cross-validation procedure, is displayed in Figure 3. Figure 3a,b visualize the effect of λ_G and λ_L on the accuracy and MSE, respectively, with constant $\lambda_R = 1$. Figure 3c,d visualize the effect of λ_R on the accuracy and MSE, respectively, for constant $\lambda_G = 100$ and $\lambda_L = 10$. During the cross validation stage the average calculation time for training with one combination of hyperparameters was 28.7 min on an Intel® Xeon® E5-1620 v3 CPU and 64 GB RAM memory.

$\lambda_G = 100$, $\lambda_L = 10$, and $\lambda_R = 1$ resulted in optimal performance during cross-validation with a test set accuracy and accuracy up to one size smaller or larger of 70.08% and

99.11%, respectively. The final solution has a sparsity of 99.8%, which represent 43 vertices from the bone mesh with non-zero model coefficients (Figure 4). Gray zones identify vertices with zero coefficients resulting from the model's sparsity. The vertices have a colour on the scale from yellow to red indicating low to high importance. Figure 5 provides an overview of the model's performance per iteration of training. The convergence towards the final model coefficients β is defined as $\|\beta^t - \beta^{t-1}\|_2^2$. One can observe the clear link between sparsity of the solution and the model's performance in terms of MSE and accuracy. On the other hand, the baseline method obtained a severely lower test set accuracy of 58.9%.



(**a**) (**b**) (**c**) (**d**)

Figure 3. The average model performance over the cross-validation folds for different values of λ_G, λ_L, and λ_R in terms of accuracy and MSE. (**a**) Accuracy as a function of λ_L and λ_G with $\lambda_R = 1$. (**b**) MSE as a function of λ_L and λ_G with $\lambda_R = 1$. (**c**) Accuracy as a function of λ_R with $\lambda_G = 100$ and $\lambda_L = 10$. (**d**) MSE as a function of λ_R with $\lambda_G = 100$ and $\lambda_L = 10$.

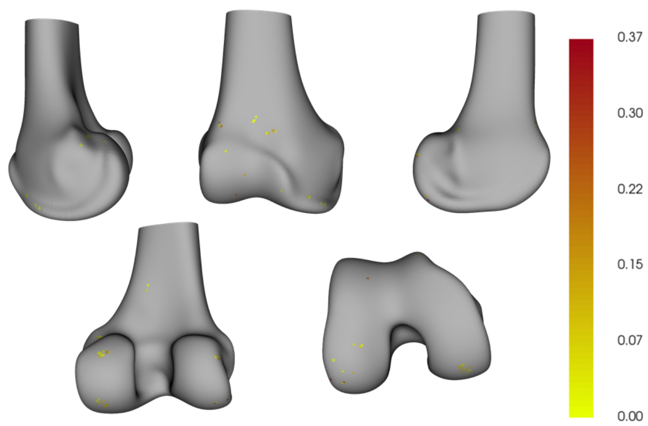


Figure 4. A medial, anterior, lateral, posterior, and distal view of the vertex importance for the model.

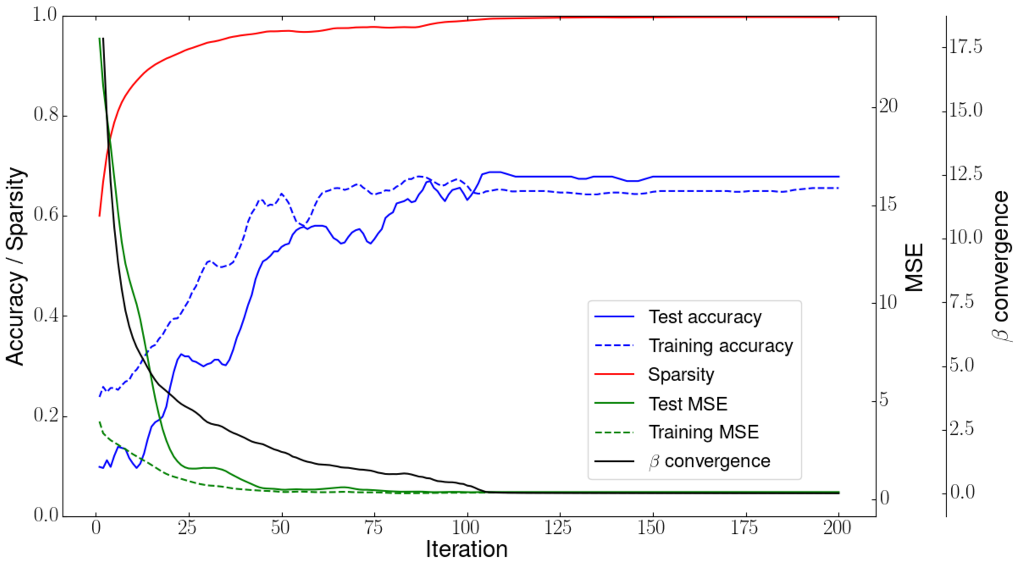


Figure 5. The learning curve obtained from training the model with the optimal hyperparameters.

From the model coefficients β, the importance of each point p can be derived by the norm of its corresponding coordinates' model coefficients: $||\beta_{\mathbf{p}}||_2$. Figure 4 visualizes the vertices importance values of the final model coefficients on different views of the femur. The coefficient vector $\beta_{\mathbf{g}}$ can also be interpreted by the direction it points to, representing the direction of variation which is most sensitive to variation in the implant size changes. In Figure 6, the directions of the model coefficients have been displayed by arrows drawn from vertex location with non-zero $||\beta_{\mathbf{g}}||_2$ in the direction of $\beta_{\mathbf{g}}$. All arrows were normalized to unit length.

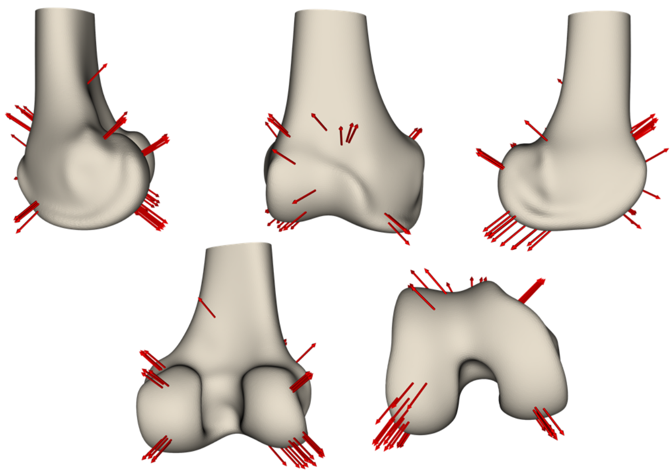


Figure 6. A medial, anterior, lateral, posterior, and distal view of the directions of the $\beta_{\mathbf{g}}$ vectors for the vertices with non-zero coefficients.

4. Discussion

The group sparsity regularization has the largest impact on the model accuracy as shown in Figure 3. For $\lambda_L > 10$ model performance starts to decrease significantly due to the diminishing effect of sparsity resulting in decreasing generalization. Nevertheless, the effect of the hypergraph regularization cannot be neglected, as can be seen in Figure 3, where the MSE starts to decrease for values of $\lambda_L > 0.1$. The L_2 regularization results in the desirable property of uniqueness of the solution. However, at higher values of λ_R the accuracy and MSE start to decrease as a result of reduced sparsity.

During the optimizer's iterations the amount of sparsity increases while the model's MSE decreases, resulting in model interpretability (Figure 5). The final solution has small clusters of points above the patellofemoral articular cartilage, on the femoral posterior condyles, on the femoral distal condylar area, and around the medial and lateral rim of the femoral distal cartilage layer are selected by the model. These regions also hold important anatomical information for determining femoral implant size. For example, the points selected on the anterior femoral shaft above the patellofemoral articular surface determine whether femoral implant notching might occur due to undersizing, where the implant undercuts the femoral cortical bone surface. Femoral notching is associated with an increased likelihood of supracondular fracture [34]. Therefore, these regions are likely taken into account by the model.

One interesting observation is that vertices in the model are very often present in small clusters of neighbours. This is most likely due to the hypergraph regularization which has a smoothing effect, trying to force neighbouring vertices to have similar model coefficients. These clusters seem to increase the model's generalization. Our hypothesis is that they make the algorithm more robust against errors in the non-linear registration performed after the SSM fit.

In Figure 6, the directions of bone shape variation which are predictive of the femoral implant size are visualized. All of these directions point outward from the bone indicating, that larger bones result in larger implant sizes, as expected. At first glance, these arrows seem to point normal to the surface of the bone mesh. However, at closer inspection, some β_g directions deviate up to 67.8° from the direction of the surface normal. For example, the vertices selected by the model on the distal lateral condyle point significantly more posterior in comparison with the direction of the surface normal at these points. This indicates that the anteroposterior variation in bone shape at the femoral condyles causes a change in femoral implant size, which is conform with clinical knowledge.

Practical implementation of our method could be through a software application which starts from the patient's MRI scan, and subsequently automatically segments it using previously investigated methods [35]. Next, the application could propose implant size based on the prediction from our proposed model. Alternatively, the model can be integrated in 3D templating software from an implant manufacturer. Our method is compatible with current generation implants with fixed increments between sizes.

Several studies have investigated the accuracy of 2D templating (Table 2) [3–7]. In the table, bold values indicate the best performance per metric. Although 2D templating is frequently performed, it has relatively low accuracy (48–64%). Furthermore, 2D templating is a manual procedure. On the other hand, manual 3D templating shows far greater accuracy of 93.9–100% [4,7,8]. Implant size selection based on manual 3D templating with MRI scans has been shown to have excellent intra- and interclass correlation coefficients amongst surgeons [36]. The intra-class correlation ranged from 96.6% to 99%, and the inter-class correction was 97%. Hence, it is a suitable medium for preoperative assessment of the required implant size. To the best of our knowledge there is only one study which investigated an automatic 2D templating method with 19.2% accuracy [37].

For our dataset, we can compare the accuracy of the model with that of the femoral implant size proposed by the instrumentation manufacturer. The test set accuracy of the femoral implant size proposed by the manufacturer is 23.1%, while our model obtained an accuracy of 70.08%. This accuracy improvement will result in a more efficient planning experience for the surgeon by reducing planning time. However, both the manufacturer's default and our method obtain the same test set accuracy up to one size larger or smaller of 99.11%. Furthermore, our proposed method also significantly surpasses the baseline method in terms of accuracy.

Our method outperforms other methods relying on demographic data in terms of both the absolute accuracy and the accuracy up to 1 size smaller or larger (Table 3). Therefore, we conclude that geometric features extracted from the 3D anatomical model provide additional information that improves implant size prediction accuracy. Furthermore, all other investigated methods rely on features defined based on clinical expertise. In comparison, our method automatically extracts the relevant features which determine the femoral implant size. The study by Lambrechts et al., (2022) obtained a higher absolute accuracy compared to our currently proposed method, however, it relied on a significant amount of clinical knowledge to extract relevant features from the 3D bone models and uses non-linear models [23]. One of the limitations of this comparison is the difference in focus: our method predicts the preoperatively planned implant size, which differs slightly from the intra-operatively used implant size predicted by the other studies relying on demographic data.

Table 2. Comparison of manual methods for predicting the required femoral implant size in terms of their accuracy. Bold values indicate the best performance per metric.

Study	Absolute Accuracy	+1/−1 Size Accuracy	Modality
Trickett et al. 2009 [3]	48%	98%	2D: X-ray
Miller et al. 2012 [4]	64%	**100%**	2D: X-ray
Unnanuntana et al. 2007 [5]	50.4%	97.3%	2D: X-ray
Pietrzak et al. 2019 [6]	52.9%	-	2D: X-Ray
Ettinger et al. 2016 [7]	59.6%	97.9%	2D: X-ray
Pietrzak et al. 2019 [6]	96.6%	-	3D: CT
Ettinger et al. 2016 [7]	**100%**	**100%**	3D: MRI
Schotanus et al. 2016 [8]	93.9%	-	3D: MRI

Table 3. Comparison of automatic methods for predicting femoral implant size in terms of their accuracy. Bold values indicate the best performance per metric.

Study	Absolute Accuracy	+1/−1 Size Accuracy	Modality
Seaver et al. 2020 [37]	19.2%	51.2%	2D: X-ray
Trainor et al. 2018 [11]	56%	99%	Shoe size
Sershon et al. 2017 [17]	-	85–95% (implant dependent)	Demographics
Bhowmik-Stoker et al. 2018 [19]	-	94%	Demographics
Sershon et al. 2019 [18]	-	76%	Demographics
Blevins et al. 2020 [22]	-	94.4%	Demographics
Wallace et al. 2020 [2]	43.7%	90.1%	Demographics
Kunze et al. 2021 [16]	48.4%	95%	Demographics
Naylor et al. 2022 [21]	-	83.09%	Demographics
Lambrechts et al. 2022 [23]	**82.2%**	-	3D: MRI
Manufacturer’s default plan	23.1%	**99.11%**	3D: MRI
Shape coefficient regression	58.93%	98.21%	3D: MRI
Hypergraph regularized group lasso	70.08%	**99.11%**	3D: MRI

More accurate implant size prediction using our method might positively impact total knee arthroplasty surgery through improved logistics requiring reduced inventory, less trays to sterilize and reduced operating room setup times. These measures might reduce the cost associated with the treatment. Furthermore, accurate templating helps the surgeon to be better prepared for surgery and can potentially reduce the operative times. Nevertheless, these claims need to be further validated.

The advantages of our model are its fast calculation time and the ability to produce accurate results when limited data are available. However, there are also points of improvement. The proposed method was validated for a single surgeon. Hence, further research can be conducted to investigate generalization to a larger population of surgeons. Additional research could look into combining the proposed mesh-based features with relevant demographic data to improve model accuracy. Another limitation is the linearity implied by our model. Possibly, non-linearities could provide higher accuracy through interaction of the vertex coordinates, such as distances or angles. Finally, it could be investigated if this method yields equal improvement for predicting tibial implant sizes. Tibial and femoral implant sizes are correlated and, thus, a method combining both bones could help increase the accuracy.

5. Conclusions

We presented a femoral implant sizing algorithm based on a regression framework mapping the bone mesh vertices locations to the implant size. For this under-determined problem, group sparsity ensured selection of the most predictive vertices on the mesh. Extra prior information was added, through hypergraph regularization, to the optimization problem by including the mesh's graph structure to constrain neighbouring to have similar impact on the regressed implant size. From our experiments we found that the group sparsity had the largest impact on the model performance, while the hypergraph-based regularization mainly improved robustness. In total, 43 anatomical locations were identified which could be used to predict the implant size. Our method was successful at predicting the femoral implant size, as preoperatively planned by a surgeon. Our method was accurate in 70.08% of test set cases, in determining the femoral implant size in comparison with an accuracy of 23.1% for the implant size provided in the preoperative plan from the instrumentation manufacturer. Towards future work, an extension of the method could include simultaneous prediction of the femoral and tibial implant sizes, since they are correlated. Furthermore, a non-linear model might, such as graph convolutional networks, might improve model performance.

Author Contributions: Conceptualization, A.L., C.V.D., R.W.-S., J.V.S., F.M. and S.V.H.; formal analysis A.L. and C.V.D.; methodology, A.L.; software, A.L. and C.V.D.; validation, A.L.; data curation, A.L.; writing—original draft preparation, A.L.; writing—review and editing, A.L., C.V.D., R.W.-S., J.V.S., F.M. and S.V.H.; supervision, R.W.-S., J.V.S., S.V.H. and F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study consist of restricted data which are not publicly available.

Acknowledgments: Sabine Van Huffel and Frederik Maes received funding from the Flemish Government (AI Research Program)and are both affiliated to Leuven. AI-KU Leuven institute for AI, B-3000, Leuven, Belgium.

Conflicts of Interest: Adriaan Lambrechts, Christophe Van Dijck and Roel Wirix-Speetjens are paid employees of Materialise NV. Jos Vander Sloten is member of the board of directors of Materialise NV.

References

1. Bourne, R.B.; Chesworth, B.M.; Davis, A.M.; Mahomed, N.N.; Charron, K.D.J. Patient Satisfaction after Total Knee Arthroplasty: Who is Satisfied and Who is Not? *Clin. Orthop. Relat. Res.* **2010**, *468*, 57–63. [CrossRef] [PubMed]
2. Wallace, S.J.; Murphy, M.P.; Schiffman, C.J.; Hopkinson, W.J.; Brown, N.M. Demographic data is more predictive of component size than digital radiographic templating in total knee arthroplasty. *Knee Surg. Relat. Res.* **2020**, *32*, 63. [CrossRef] [PubMed]
3. Trickett, R.W.; Hodgson, P.; Forster, M.C.; Robertson, A. The reliability and accuracy of digital templating in total knee replacement. *J. Bone Jt. Surg. Br. Vol.* **2009**, *91*, 903–906. [CrossRef] [PubMed]
4. Miller, A.G.; Purtill, J.J. Accuracy of digital templating in total knee arthroplasty. *Am. J. Orthop.* **2012**, *41*, 510–512. [PubMed]
5. Unnanuntana, A.; Arunakul, M.; Unnanuntana, A. The accuracy of preoperative templating in total knee arthroplasty. *J. Med. Assoc. Thail. Chotmaihet Thangphaet* **2007**, *90*, 2338–2343.
6. Pietrzak, J.R.T.; Rowan, F.E.; Kayani, B.; Donaldson, M.J.; Huq, S.S.; Haddad, F.S. Preoperative CT-Based Three-Dimensional Templating in Robot-Assisted Total Knee Arthroplasty More Accurately Predicts Implant Sizes than Two-Dimensional Templating. *J. Knee Surg.* **2019**, *32*, 642–648. [CrossRef]
7. Ettinger, M.; Claassen, L.; Paes, P.; Calliess, T. 2D versus 3D templating in total knee arthroplasty. *Knee* **2016**, *23*, 149–151. [CrossRef]
8. Schotanus, M.G.M.; Schoenmakers, D.A.L.; Sollie, R.; Kort, N.P. Patient-specific instruments for total knee arthroplasty can accurately predict the component size as used peroperative. *Knee Surg. Sport. Traumatol. Arthrosc.* **2017**, *25*, 3844–3848. [CrossRef]
9. Mahoney, O.M.; Kinsey, T. Overhang of the Femoral Component in Total Knee Arthroplasty: Risk Factors and Clinical Consequences. *J. Bone Jt.-Surg.-Am. Vol.* **2010**, *92*, 1115–1121. [CrossRef]

10. Tibesku, C.O. 15 The Problem of Under-or Oversizing of Total Knee Replacement. In *The Unhappy Total Knee Replacement: A Comprehensive Review and Management Guide*; Springer: Cham, Switzerland, 2015; pp. 175–184.
11. Trainor, S.; Collins, J.; Mulvey, H.; Fitz, W. Total Knee Replacement Sizing: Shoe Size Is a Better Predictor for Implant Size than Body Height. *Arch. Bone Jt. Surg.* **2018**, *6*, 100–104.
12. Lachiewicz, P.F.; Henderson, R.A. Patient-specific Instruments for Total Knee Arthroplasty. *J. Am. Acad. Orthop. Surg.* **2013**, *21*, 513–518. [CrossRef]
13. MacDessi, S.J.; Jang, B.; Harris, I.A.; Wheatley, E.; Bryant, C.; Chen, D.B. A comparison of alignment using patient specific guides, computer navigation and conventional instrumentation in total knee arthroplasty. *Knee* **2014**, *21*, 406–409. [CrossRef]
14. Liu, Z.; Wu, S.; Jin, S.; Ji, S.; Liu, Q.; Lu, S.; Cheng, L. Investigating Pose Representations and Motion Contexts Modeling for 3D Motion Prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 681–697.
15. Liu, H.; Liu, M.; Li, D.; Zheng, W.; Yin, L.; Wang, R. Recent Advances in Pulse-Coupled Neural Networks with Applications in Image Processing. *Electronics* **2022**, *11*, 3264. [CrossRef]
16. Kunze, K.N.; Polce, E.M.; Patel, A.; Courtney, P.M.; Levine, B.R. Validation and performance of a machine-learning derived prediction guide for total knee arthroplasty component sizing. *Arch. Orthop. Trauma Surg.* **2021**, *141*, 2235–2244. [CrossRef]
17. Sershon, R.A.; Courtney, P.M.; Rosenthal, B.D.; Sporer, S.M.; Levine, B.R. Can Demographic Variables Accurately Predict Component Sizing in Primary Total Knee Arthroplasty? *J. Arthroplast.* **2017**, *32*, 3004–3008. [CrossRef]
18. Sershon, R.A.; Li, J.; Calkins, T.E.; Courtney, P.M.; Nam, D.; Gerlinger, T.L.; Sporer, S.M.; Levine, B.R. Prospective Validation of a Demographically Based Primary Total Knee Arthroplasty Size Calculator. *J. Arthroplast.* **2019**, *34*, 1369–1373. [CrossRef]
19. Bhowmik-Stoker, M.; Scholl, L.Y.; Khlopas, A.; Sultan, A.A.; Sodhi, N.; Moskal, J.T.; Mont, M.A.; Teeny, S.M. Accurately Predicting Total Knee Component Size without Preoperative Radiographs. *Surg. Technol. Int.* **2018**, *33*, 337–342.
20. Ren, A.N.; Neher, R.E.; Bell, T.; Grimm, J. Using Patient Demographics and Statistical Modeling to Predict Knee Tibia Component Sizing in Total Knee Arthroplasty. *J. Arthroplast.* **2018**, *33*, 1732–1736. [CrossRef]
21. Naylor, B.H.; Butler, J.T.; Kuczynski, B.; Bohm, A.R.; Scuderi, G.R. Can Component Size in Total Knee Arthroplasty Be Predicted Preoperatively?—An Analysis of Patient Characteristics. *J. Knee Surg.* **2022**. [CrossRef]
22. Blevins, J.L.; Rao, V.; Chiu, Y.f.; Lyman, S.; Westrich, G.H. Predicting implant size in total knee arthroplasty using demographic variables. *Bone Jt. J.* **2020**, *102*, 85–90. [CrossRef] [PubMed]
23. Lambrechts, A.; Wirix-Speetjens, R.; Maes, F.; Van Huffel, S. Artificial Intelligence Based Patient-Specific Preoperative Planning Algorithm for Total Knee Arthroplasty. *Front. Robot. AI* **2022**, *9*, 899349. [CrossRef] [PubMed]
24. Clemmensen, L.; Hastie, T.; Witten, D.; Ersbøll, B. Sparse Discriminant Analysis. *Technometrics* **2011**, *53*, 406–413. [CrossRef]
25. Lorensen, W.E.; Cline, H.E. Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Comput. Graph.* **1987**, *21*, 163–169. [CrossRef]
26. Van Dijck, C.; Wirix-Speetjens, R.; Jonkers, I.; Vander Sloten, J. Statistical shape model-based prediction of tibiofemoral cartilage. *Comput. Methods Biomech. Biomed. Eng.* **2018**, *21*, 568–578. [CrossRef]
27. Amberg, B.; Romdhani, S.; Vetter, T. Optimal Step Nonrigid ICP Algorithms for Surface Registration. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MI, USA, 17–22 June 2007; pp. 1–8. [CrossRef]
28. Roth, V.; Fischer, B. The Group-Lasso for generalized linear models: Uniqueness of solutions and efficient algorithms. In Proceedings of the 25th International Conference on Machine Learning, ICML '08, Helsinki, Finland, 5–9 July 2008; pp. 848–855. [CrossRef]
29. Hastie, T.R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2015.
30. Ma, X.; Liu, W.; Li, S.; Tao, D.; Zhou, Y. Hypergraph p-Laplacian Regularization for Remotely Sensed Image Recognition. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1585–1595. [CrossRef]
31. Zou, H.; Hastie, T. Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc. Ser. (Stat. Methodol.)* **2005**, *67*, 301–320. [CrossRef]
32. Nesterov, Y.E. A method for solving the convex programming problem with convergence rate O (1/k^ 2). *Dokl. Akad. Nauk SSSR* **1983**, *269*, 543–547.
33. Duff, I.S.; Grimes, R.G.; Lewis, J.G. Sparse matrix test problems. *ACM Trans. Math. Softw.* **1989**, *15*, 62043. [CrossRef]
34. Stamiris, D.; Gkekas, N.K.; Asteriadis, K.; Stamiris, S.; Anagnostis, P.; Poultsides, L.; Sarris, I.; Potoupnis, M.; Kenanidis, E.; Tsiridis, E. Anterior femoral notching $\leq$ 3 mm is associated with increased risk for supracondylar periprosthetic femoral fracture after total knee arthroplasty: A systematic review and meta-analysis. *Eur. J. Orthop. Surg. Traumatol.* **2022**, *32*, 383–393. [CrossRef]
35. Ambellan, F.; Tack, A.; Ehlke, M.; Zachow, S. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative. *Med. Image Anal.* **2019**, *52*, 109–118. [CrossRef]

36. Schoenmakers, D.A.L.; Theeuwen, D.M.J.; Schotanus, M.G.M.; Jansen, E.J.P.; van Haaren, E.H.; Hendrickx, R.P.M.; Kort, N.P. High intra- and inter-observer reliability of planning implant size in MRI-based patient-specific instrumentation for total knee arthroplasty. *Knee Surg. Sport. Traumatol. Arthrosc.* **2021**, *29*, 573–578. [CrossRef]
37. Seaver, T.; McAlpine, K.; Garcia, E.; Niu, R.; Smith, E.L. Algorithm based automatic templating is less accurate than manual digital templating in total knee arthroplasty. *J. Orthop. Res.* **2020**, *38*, 1472–1476. [CrossRef]