A Patient-Specific Model for Predicting Tibia Soft Tissue Insertions From Bony Outlines Using a Spatial Structure Supervised Learning Framework

Cao Xiao, *Member, IEEE*, Shouyi Wang, *Member, IEEE*, Liying Zheng, Xudong Zhang, and Wanpracha Art Chaovalitwongse, *Senior Member, IEEE*

Abstract-Recreating the natural anatomy in ligament reconstruction is crucial to fully restore the knee joint function and reduce impingement on iatrogenic injury to adjacent structures, yet is subject to the difficulties in locating ligament and other associated soft tissues insertion sites intraoperatively and the high interperson morphological variability cross patients. In this study, we present a new quantitative analysis method capable of achieving personalized identification of cruciate ligament and soft tissue insertions. We craft patient-specific features of tibia outline that can be accurately and reliably measured from CT images. In addition, we propose a supervised structure learning and prediction model with special interdimensional and response structure regularization terms to capture relationship between the spatial arrangement of soft tissue insertions and the patient-specific features extracted from the tibia outlines. In the experiment, the proposed model outperforms baseline models and provides an accurate and accessible approach that can be used as the first and the most critical step to achieve personalized surgical planning in cruciate ligament reconstruction.

Index Terms—Cruciate ligament and meniscal insertions, data mining, knee anatomic reconstruction, patient-specific prediction, structured supervised learning, X-ray imaging and computed tomography.

I. INTRODUCTION

T HE knee is a complex joint that supports relatively large loads and great mobility, making it vulnerable to a variety of injuries. The anterior and posterior cruciate ligaments (ACL and PCL) are two primary stabilizers of the human knee. The ACL is commonly injured: an estimated 175 000 ACL reconstructions are performed annually, with a financial impact exceeding 2 billion dollars, in the U.S. alone. PCL injuries occur

Manuscript received March 27, 2015; revised September 30, 2015, December 29, 2015, and March 4, 2016; accepted March 11, 2016. Date of publication April 27, 2016; date of current version September 14, 2016. This paper was recommended by Associate Editor K. Li.

C. Xiao is with the Department of Industrial and Systems Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: xiaoc@uw.edu).

S. Wang is with the Department of Industrial and Manufacturing Systems Engineering, Universit of Texas, Arlington, TX 76019 USA (e-mail: shouyiw@uta.edu).

L. Zheng is with the Department of Orthopaedic Surgery, University of Pittsburgh, Pittsburgh, PA 15203 USA (e-mail: zlyreed@gmail.com).

X. Zhang is with the Departments of Mechanical Engineering and Materials Science, Bioengineering, and Orthopaedic Surgery, University of Pittsburgh, Pittsburgh, PA 15260 USA (e-mail: xuz9@pitt.edu).

W. A. Chaovalitwongse is with the Department of Industrial and Systems Engineering and the Department of Radiology, University of Washington, Seattle, WA 98195 USA (e-mail: artchao@uw.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier

less frequently are believed to be underdiagnosed; yet they affect about 3% of the general population and account for as many as 40% of patients with knee trauma seen in emergency rooms.

The anatomic reconstruction procedure involves creating the bone tunnels and placing the substituting grafts in the exact anatomical positions as the native ligaments. An accurate replication of the natural anatomy in anatomic reconstruction is crucial to fully restore knee joint function and reduce impingement on or iatrogenic injury to adjacent structures [1]–[5]. However, current approaches to treating knee injuries are not quite good in terms of consistency and effectiveness in restoring knee function and preventing the development of osteoarthritis (OA). Analyses of long-term outcome after ACL reconstruction have revealed that only 37% of the patients were normal restored in terms of knee structure and function [6], and 90% of the ACLreconstructed knees exhibited radiographic evidence of OA 3-10 years after injury [7]. A growing body of evidence is suggesting that anatomic reconstruction, performed by creating the bone tunnels and placing the substituting graft at the native ligament insertion site, can better restore the joint function and deter the development of OA. A number of challenges are present in practical anatomic reconstruction of cruciate ligaments.

Intraoperative identification of the native cruciate ligament insertion sites, as a requisite for anatomical reconstruction, poses a tremendous challenge. Not all surgeons can maintain an acute awareness of the anatomy: about 85% of ACL reconstructions are done by surgeons who perform fewer than ten cases per year [8] and PCL reconstructions are even less frequently performed by most surgeons; for those who can, factors including the arthroscopic distortion and disappearance of the ligament remnant (naturally or due to a notchplasty procedure) can still cause misidentification of the natural insertion or attachment sites. There is considerable variability of knee anatomy in terms of bone and soft tissue insertion morphology (position, size, and shape) [9]. Sample data from our preliminary study of tibial insertion site morphometry suggest that simplistic cross-referencing or generalization from one patient to another is likely to lead to nonanatomical tunnel drilling and iatrogenic injury to adjacent tissue structures. Although it may be difficult to gauge the incidence and impact of these iatrogenic injuries as complications of ACL or PCL surgeries, the importance of minimizing the risk of such injuries is readily recognized [10], [11].

The key to anatomic cruciate ligament surgery with minimized risk of iatrogenic injury is an accurate, quantitative knowledge base of the tissue morphology, documenting interperson variability and specificity versus uncertainty associated with alternative ways to predict morphometrics. Studies have investigated the quantification of the insertion sites of the cruciate ligaments and other soft tissue components using statistical and quantitative approaches [12]-[14]. However, such quantitative analysis and measures generally cannot fully capture the accurate spatial arrangement of soft tissue insertions. The location and morphological measures cannot account for the interperson variability of cruciate ligament and meniscus insertions, which are mostly characterized by qualitative measures [15], [16]. Advanced imaging techniques, such as 3-D CT or MRI, which can be useful to visualize the major structure outline of the knee of a patient clearly, cannot be applied directly to determine the insertion sites for cruciate ligaments reconstruction. This is because the imaging shows the structure of a knee with serious cruciate ligaments injuries and structure misalignment. While in the surgery, it is crucial to identify the native location of the cruciate ligaments to reconstruct the natural anatomy of the ligament structure. Therefore, one needs inference on the appropriate insertion sites of native cruciate ligaments. Due to the complex anatomy of the knee, the identification of insertion sites of cruciate ligaments in knee reconstruction surgery is still an unsolved problem.

This study aims to develop a new quantitative analysis method to achieve personalized identification of cruciate ligament and meniscal insertions using patient-specific knee morphological features. In particular, the proposed framework first digitalized outlines of tibia from 3-D CT images and aligned the outlines using generalized procrustes analysis (GPA) techniques. It then extracted patient-specific features, trained a supervised structure learning and prediction model, and predicted the centroids of the sites of the cruciate ligament and meniscal insertions using the learned prediction model. The supervised structure learning and prediction model captured the relationship between the spatial arrangement of soft tissue insertions and the patientspecific features extracted from the tibia outlines, which can be easily and reliably measured from CT images. To the best of our knowledge, this is the first supervised machine learning algorithms in knee soft tissue site identification. The proposed learning and prediction framework provides a critical step to achieve the highly demanded personalized surgical planning in cruciate ligament reconstruction.

The rest of the paper is organized as follows. Section II presents the data acquisition and knee imaging data processing including the digitalization of tibia and soft tissues from 3-D CT image, the image alignment and normalization using GPA, the innovative feature engineering via coordinates transformation, and the structure learning and prediction model for the soft tissue insertion centroids. Section III presents the prediction performance of the proposed framework as well as the performance comparison with the available baseline models. Finally, we conclude this paper in Section IV.

II. METHODS

A. Data Collection

Twenty tibia specimens (ten left and ten right unpaired knees; 11 from men and 9 from women; mean age at death: 61 ± 5

Fig. 1. Digitized cartilage and insertion site outlines mapped onto the CTbased 3-D tibia model. The digitized points (asterisks) were spline-fitted, generating 100 equidistant points (circles on the close-up view of ACL insertion outline) on the fitted outlines to facilitate the subsequent analyses [14].

years) were used to acquire the morphometric data [17]. All epithelial, subcutaneous, and muscular tissues were removed from the specimens. High-resolution CT scans of the tibias were taken with slice spacing of 0.625 mm and 3-D bone models of the tibias were created in Mimics (Materialise Inc., Belgium). A Polaris Spectra optical tracking system (Northern Digital Inc., Ontario, Canada), with a manufacturer-reported accuracy of ± 0.25 mm, was used to digitize the outlines of the ACL, PCL, the medial cartilage (mcart), the lateral cartilage (lcart), anterior and posterior medial meniscal root (ammr and pmmr), and anteriorlateral and posterior-lateral meniscal root (almr and plmr). The digitization was performed by the same experimenter with the repeatability, as assessed by intraclass correlation coefficients, ranging from 0.94 to 0.99. The digitized outlines were mapped onto the CT-based 3-D tibia models with a fiducial registration error smaller than 2% [14]. A closed spline was then fitted to each outline, resulting in 100 equidistant discrete points to represent the outline, as shown in Fig. 1.

A 3-D coordinate system was defined on each tibia based on its digitized and mapped cartilage outlines (see Fig. 1). First, the origin of the coordinate system was determined as the midpoint of the medial and lateral cartilage centroids. The principal component analysis (PCA) was then performed on the equidistant discrete points representing the cartilage outlines (200 points in total). The X-axis was the first principal component axis passing the origin and pointing laterally. The Y-axis was orthogonal to the X-axis, passing the origin and pointing anteriorly. To make the Z-axis point proximally, the coordinate system was designed as a right-handed system for the right tibia and a lefthanded system for the left tibia.

B. Image Alignment and Normalization Using Generalized Procrustes Analysis

Cartilage outlines for all 20 tibias were optimally aligned using GPA, which is an iterative process of applying procrustes superimposition to all possible pairs of configurations; a configuration here refers to a set of cartilage outline landmark coordinates in a predefined order. For each cartilage configuration





Fig. 2. Effect of procrustes superimposition illustrated by one pair of tibias. One cartilage configuration served as the base (thick) and another as the target (thin). Six tissue structure insertions before (left column) and after (right column) superimposition are also shown in three orthogonal views [14].



Fig. 3. Outlines of tibial cartilage and six insertion sites from 20 subjects before (left plot) and after (right plot) cartilage-based GPA [14].

pair, one configuration served as the base and the other as the target. Procrustes superimposition matches the target configuration onto the base, centering, rotating, and uniformly scaling the target configuration to minimize the shape difference (see Fig. 2). For multiple (20 in this study) configurations, GPA identified the reference or overall base configuration as the one with the smallest overall procrustes distance to all others (i.e., the 19 remaining tibial cartilage configurations). The 19 remaining configurations were then procrustes-superimposed onto this selected reference and their insertion sites transformed accordingly by the same translation, rotation, and scaling rules, without any shape distortion. Fig. 3 shows the outlines of tibial cartilage and six insertion sites from 20 subjects before and after cartilage-based GPA.

With the ultimate goal of 3-D prediction directly, we first need to investigate and evaluate the feasibility if the internal structure of cruciate ligaments can be possibly learned from the outline of tibia. Thus, we first limit our scope in a 2-D plane and investigate if the structure relation can be learned. Conceptually, if certain 3-D relationship between tibia outline and locations of cruciate ligaments exists, then the structure should also exhibit in the projected 2-D plane. In other words, if the structural prediction performance on a 2-D plane is promising, then it will confirm that the knee outline information is useful in prediction of native locations of soft tissues. Most importantly, it will also indicate structural relation between tibia outline and soft tissue locations may also be valid in the 3-D space.

C. Feature Engineering Using Coordinates Transformation

The outline of tibia can be easily and reliably measured, making it a feature candidate to predict the locations of intangible soft tissues. From examining the digitalized 3-D data, we observed that the length and width dimensions did not change significantly along with the change of the depth dimension. Hence, the depth dimension can be considered a noise dimension to be filtered from digital images. After preprocessing, the digitalized image data in Cartesian coordinates can represent each point in 2-D space (x, y). In particular, we employed the polar coordinates to reduce feature dimension. The feature dimension for Cartesian coordinates for each 2-D point on the tibia boundary is 2, however, for polar coordinates, since we take points from equal intervals and drop the angle, and the feature dimension becomes 1.

A 2-D point with Cartesian coordinates (x, y) can be transformed into polar coordinates (ρ, θ) following the rule: $x = \rho \cos \theta$, $y = \rho \sin \theta$. For the shape of tibia, we consider the aligned origin as its center and transform all points (including all boundary and all inner points) into polar coordinates according to the above rule. To represent the outline of tibia with a limited number of features, we implemented a discretization and boundary detection algorithm. The algorithm first divides a complete cycle into N = 36 equal intervals, creating a series of 10° intervals. Then, the algorithm finds the maximal ρ in each interval and records it as ρ_t , where t = 1, ..., 36, resulting a functional



Fig. 4. Patient-specific feature extraction procedure from CT image of the Tibia. First, the tibia outline in Cartesian coordinates is converted into polar coordinates. We divide a complete cycle into 36 equal intervals and, then, find the maximal magnitude in each interval to generate a 36-D functional data series as the input predictive variable to train the prediction model of soft tissue insertion sites.

data series of tibia shape (see Fig. 4). This way, each outline of tibia was represented by a 36-D vector $(\rho_1, \rho_2, ..., \rho_{36})$.

Such transformation characterizes the shape of tibia by a functional data series for each individual patient. Such patient-specific tibia features will be used to predict the sites of ligament and other soft tissue insertions.

There is high variability in ligament and meniscus insertions. In a surgical procedure, the centroids of the insertions are of particular importance to tunnel locations. Thus, we focused on the prediction of the centroids of the ligament and meniscus insertions instead of the complete morphology of the insertions.

For each subject $i \in \{1, ..., 20\}$, the centroid of an insertion site j, (a_{ij}, b_{ij}) , is defined as

$$\left(\frac{\max(x_{ij}) + \min(x_{ij})}{2}, \frac{\max(y_{ij}) + \min(y_{ij})}{2}\right). \quad (1)$$

D. Structure Supervised Learning Model for Soft Tissue Centroid Prediction

Regardless of interperson variability existing in morphology, the spatial arrangement of the eight soft tissues basically follow an intrinsic pattern of intertissue structure. For example, in 2-D CT scan, ACL is above PCL, AMMR is always adjacent to ACL, and no soft tissue can be out of the tibia outline. To correctly retain such a structure when predicting the centroid of each soft tissue, we also need to consider the correlations among soft tissue centroids. Since centroids are the responses in our model, we adopt the following strategy to address such consideration.

In multivariate regression, we learn the relationship between m response variables $\{y^{(1)}, y^{(2)}, \ldots, y^{(m)}\}$ and p predictors $\{x^{(1)}, x^{(2)}, \ldots, x^{(p)}\}$. Each $y^{(j)}$ has its own regression model

$$y^{(j)} = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 + \dots + \beta_{pj}x_p + \epsilon_j.$$
 (2)

The above problem can be generalized as sparse multitask learning [18], which minimizes overall errors among m responses instead of the error of each individual response with a proper regularization on matrix β :

$$\min \sum_{j=1}^{m} ||y^{(j)} - X\beta^{(j)}|| + ||\beta||.$$
(3)

To capture the spatial arrangement of the soft tissues, based on our previous studies on prediction model development for medical problems [19]–[21], we modeled the spatial structure in a simplified linear relation as follows:

$$(y^{(1)}, \dots, y^{(m)})A^{(j)} = 0, \quad j \in \{1, 2, \dots, m\}, \ a_{jj} \neq 0.$$
 (4)

It shows that each response variable can be linearly represented by the other seven response variables. The optimization objective should not only consider how well the prediction $\hat{y}^{(j)}$ fit the response, but also how well the predictions $\{\hat{y}^{(1)}, \hat{y}^{(2)}, \ldots, \hat{y}^{(m)}\}$ fit the structure that the responses would fit.

To predict the centroids of the eight soft tissues within the tibia, we have developed a structure supervised learning model to construct a regression model for the centroids of all eight soft tissues. In the model, the independent variables (features) are the functional data series of tibial shape and the response variables are the distance from soft tissue centroids to the tibia centroid (0, 0).

Let x_i , i = 1, ..., 20, be features extracted from the tibia outline, where x_i is a vector of 36 dimensions that represent the tibia functional data series. Let y_i be response variables (soft tissue centroids), each is a vector of 16 dimensions (xy coordinates off the centroids of all eight soft tissues.) Let A be a matrix describing linear relations between each components of y_i , and β be the matrix indicating linear relation between y_i and x_i . Let α_1 , α_2 , and α_3 be the tuning parameters. The objective function of our structure-based regression model is given by

$$\min_{A,\beta} \sum_{i=1}^{20} ||y_i - x_i\beta|| + \alpha_1 \sum_{i=1}^{20} ||y_iA|| + \alpha_2 \sum_{i,j} |A_{i,j}| + \alpha_3 \sum_{i,j} |\beta_{i,j}|$$
(5)

where the first term is the sum of Euclidean norm of the error terms based on linear regression between y_i and x_i ("Euclidean norm error term"). The second term is the sum of Euclidean norm of the error terms based on linear regression between different dimensions of y_i ("interdimensional regularization term"). The last two terms are l_1 regularization terms, where the third regularizes on the structures on response ("response structure regularization term"), and the last one is the l_1 regularization term on β (" $\beta - l_1$ term"). With the three regularization components, our framework can predict on response variables with learned spatial structures.

E. Soft Tissue Centroid Prediction

The trained model balances the fitness between feature vector x and response vector y and the spatial structure embedded in the response vector. With the learned model parameters β , A, the prediction step is to find a \hat{y} that minimizes the objective function as follows:

$$\min_{y} ||y - x\beta|| + \alpha_1 ||yA||.$$
 (6)

Given a feature vector of a new patient, the optimization framework ensures that the predicted centroid locations follows the learned spatial structure and also minimizes the fitting errors.

III. EVALUATIONS

In this section, we test the proposed model on the knee data with 20 subjects and also compare the model with several alternative methods. Since the sample size is smaller than the feature size, we try lasso, population mean, and k-nearest neighbor approaches.

A. Performance Measures

The leave-one-out-cross-validation was employed to obtain an unbiased estimate of the generalization ability of the prediction models. For a dataset with n subjects, the procedure consists of n trials. In each trial, a model is trained from n - 1subjects to predict the remaining one and calculate the squared error against its actual value. The procedure repeats n times until all subjects have been tested once. We used the mean squared error (MSE) over the 20 subjects as the performance metric to evaluate the accuracy of centroid prediction to compare different approaches.

B. Effect of Spatial Structure Learning

First, we examine the proposed formulations on the knee data of 20 subjects. In particular, we compared our proposed framework with the following three variations: 1) Model A: minimize Euclidean norm error term with only l_1 regularization term on β ; 2) Model B: structural learning with linear predictor (no optimization); 3) Model C: minimize Euclidean norm error term with only interdimensional error term and l_1 regularization term on β , and 4) Model D: the population mean method. The comparison will be based on prediction error for each predictor variables and the total number of variables that outperform population mean method.

The comparison results are summarized in Fig. 5. One can observe that with only l_1 regularization term on β (Model A), the prediction suffered from high prediction error due to ignorance of variance and structural properties in the response. With only interdimensional regularization term and l_1 regularization term on β (Model C), prediction still cannot outperform population mean method due to lack of consideration of structural properties in the response. Structural learning without optimization procedure (Model B) does not predict well either due to ignoring the correlation among the aforementioned terms. Our proposed model, with adding interdimensional regularization term, response structure regularization term, l_1 regularization term on β , and optimization procedure, can outperform population mean method in terms of total residual sum of errors.

C. Comparison With Baseline Methods

To show the advantages of the proposed method, we also compared the model with several alternative methods, including lasso, population mean, and k-nearest neighbor approach described as follows.

1) Linear Regression With Regularization: Since the number of features is greater than sample size, we apply lasso [22] to each coordinate of eight soft tissues. This approach does not consider the interrelationship of the positions.

2) Population Mean: To predict soft tissue centroids using population-based prediction, in our case, we consider our 20 subjects as a target population. When we make prediction, we need to ensure that the predicted soft tissue centroids are consistent with the community's average characteristics.



Fig. 5. Flowchart of the training and testing of the proposed patient-specific prediction model of soft tissue insertion sites using the 36-D features extracted from tibia outlines.

 TABLE I

 KNEE SOFT TISSUE CENTROID PREDICTION MSE (MM) (LEAVE-ONE-OUT PREDICTION) IN CARTESIAN COORDINATES

Soft Tissue	Proposed Model	lasso Model	Population Mean	KNN (Euclidean)	KNN (t-stat)	KNN (DTW=1)	KNN (DTW=2)
lcart (x,y)	(17.49, 21.65)	(28.44, 21.18)	(70.01, 21.73)	(19.64, 21.88)	(34.82, 22.36)	(19.19, 20.98)	(22.09, 21.47)
mcart (x,y)	(22.32, 11.63)	(25.60, 11.63)	(69.50, 11.63)	(23.84, 15.72)	(44.60, 12.90)	(27.63, 13.54)	(32.18, 12.53)
acl (x,y)	(73.60, 99.78)	(77.69, 108.33)	(70.77 101.50)	(51.41, 137.32)	(62.32, 103.00)	(56.72, 120.75)	(57.61, 126.24)
almr (x,y)	(25.58, 76.84)	(34.40, 107.26)	(28.93, 82.60)	(28.00, 98.17)	(28.09, 85.07)	(29.68, 87.77)	(29.98, 93.07)
ammr (x,y)	(224.71, 57.37)	(222.64, 40.95)	(214.33, 33.68)	(355.41, 41.10)	(228.54, 31.43)	(271.12, 31.34)	(261.33, 30.49)
pcl (x,y)	(56.70, 36.39)	(49.15, 49.79)	(58.60, 40.15)	(62.19, 34.91)	(71.78, 36.23)	(67.77, 35.39)	(63.49, 36.68)
plmr(x,y)	(66.20, 64.78)	(78.47, 82.59)	(68.85, 68.54)	(74.36, 67.75)	(65.78, 66.02)	(83.66, 69.17)	(76.27, 69.64)
pmmr (x,y)	(65.51, 88.69)	(68.42, 102.61)	(74.99, 102.85)	(73.18, 110.79)	(78.15, 95.08)	(68.63, 105.83)	(58.84, 106.00)

3) K Nearest Neighbor Algorithm With Three Distance Measures: K nearest neighbor [23] can be used both in classification and in regression. In classification, we assign class label to a subject according to the majority vote of its k closest training examples. In regression, we compute the value of a subject by taking average of its k closest training examples. From its k most nearest training samples, we can learn the knowledge of a subject. The k nearest neighbor algorithm provides us with more patient-specific information than population-based approach.

In the k-nearest-neighbor algorithm, distances between two subjects can be considered similarity between them. In our analysis, we compare three different distance measures in calculating such similarities.

- Euclidean distance measure is the distance on Euclidean space. Intuitively, it is the length of a line segment connecting two data points.
- The two sample t-test measure. Lower p-value indicates bigger difference.
- 3) In time-series analysis, dynamic time warping (DTW) [24] measures similarity between two temporal sequences, which may vary in time or speed. When we compare two sequences, instead of making point-to-point comparison, we create time windows and consider the minimal distance in a time window as the distance. We try two difference

time windows. Window parameter is the unit of time shift we allow.

4) Comparison Results With Baseline Models: Table I summarizes the prediction performance of the eight soft tissue centroids by the proposed model and the baseline models aforementioned. For each soft tissue centroid, the average residual sum of squares of x- and y-axis over the 20 subjects are reported. The MSE of the proposed model is equal to or better than that of population mean-based method in 13 out of 16 axes (see Table I) and is the best (or tie) in ten out of 16 axes among all the methods. Table II summarizes the Euclidean distances between the predicted and actual soft tissue centroids. Again, the proposed method achieved the best prediction performance in six out of the eight soft tissue centroids. These experimental results confirmed that the proposed supervised learning model was effective to learn the spatial structure of the eight soft tissues to improve prediction performance. Fig. 6 illustrates the centroid predictions on one subject. The proposed method generated predictions closer to the actual centroids in six out of the eight soft tissues.

In addition, we also compared a simple linear model with the proposed model without regularization, and the prediction results are summarized in Table III. One can observe that the prediction performance of the linear model has a high variance

TABLE II KNEE SOFT TISSUE CENTROID PREDICTION MSE (MM) (LEAVE-ONE-OUT PREDICTION) IN POLAR COORDINATES

Soft Tissue	Proposed Model	lasso Model	Population Mean	KNN (Euclidean)	KNN (t-stat)	KNN (DTW=1)	KNN (DTW=2)
lcart (r)	39.14	49.62	91.74	41.52	57.18	40.17	43.56
mcart (r)	33.95	37.23	81.13	39.56	57.5	41.17	44.71
acl (r)	173.38	186.02	172.27	188.73	165.32	177.47	183.85
almr (r)	102.42	141.66	111.53	126.17	131.16	117.45	123.05
ammr (r)	282.08	263.59	248.01	396.51	259.97	302.46	291.82
pcl (r)	93.09	98.94	98.75	97.1	108.01	103.16	100.17
plmr (r)	130.98	161.06	137.39	142.11	131.8	152.83	145.91
pmmr (r)	154.2	171.03	177.83	183.97	173.23	174.46	164.84



Fig. 6. Outline of tibia and the centroids of soft tissues are in red. The predicted centroids (green) by our model are in general closer to the actual locations than the predicted centroids (blue) by population mean.

 TABLE III

 KNEE SOFT TISSUE CENTROID PREDICTION MSE (MM) (OVERFITTING)

Soft Tissue	Linear Model	lasso Model		
lcart (x, y)	(107.84, 102.42)	(51.44, 34.55)		
mcart(x, y)	(56.66, 112.55)	(32.08, 35.74)		
$\operatorname{acl}(x, y)$	(769.07, 321.92)	(192.03, 225.45)		
$\operatorname{almr}(x, y)$	(326.28, 727.48)	(75.46, 208.19)		
$\operatorname{ammr}(x, y)$	(670.52, 720.83)	(418.54, 88.61)		
pcl(x, y)	(332.68, 109.64)	(148.62, 27.27)		
plmr(x, y)	(570.84, 520.73)	(180.58, 159.46)		
pmmr (x, y)	(459.02, 1219.24)	(286.87, 235.75)		

and is among the worst compared with other baseline models. Also without the structural learning term, the proposed method is reduced to the regular lasso model, and the prediction performance was significantly deteriorated. This observation also indicated that the structural learning term is indeed useful to improve the prediction performance by considering the learned spatial structure of the eight soft tissues.

IV. DISCUSSION AND CONCLUSION

The motivation of this study is to provide an accurate quantitative approach to aid soft tissue insertion localization using patient-specific measures that can be reliably and accurately acquired from knee imaging data. An extensive quantitative analysis of the location and the interrelationship of soft tissue insertions on the tibial plateau has been performed, including digitalization of tibia outlines from 3-D CT images, the imaging alignment using GPA, patient-specific morphological feature extraction from tibia outlines, integration of a spatialstructure learning in a regularized regression-based model training framework, and a patient-specific prediction framework with the learned spatial structure of the soft tissue insertions. In particular, we demonstrated the possibility of using outlines of tibia to predict the centroids of eight soft tissue insertions that are crucial in anatomical reconstruction of cruciate ligaments. The proposed methodology yielded the best prediction accuracies compared with other baseline models for the eight soft tissue locations. The integration of the proposed spatialstructure learning framework has demonstrated to improve the prediction performance significantly.

With the limited data size of 20 subjects, the primary goal of this study was to investigate the feasibility of establishing a quantitative prediction methodology of soft tissue insertions. The proposed structure learning and prediction framework could potentially be strengthened or complemented if more data samples become available. In this study, we only restricted our analysis of tibia outlines and soft tissue centroids in a 2-D space using the GPA aligned tibia outlines and a PCA-transformed coordinates. A natural extension of this work is to construct a full 3-D prediction model using 3-D outlines of tibia. Although the 3-D knee models for the subject-specific geometry of bone and soft tissues are widely available from imaging data, such as CT and MRI [25]–[27], it is still challenging to obtain accurate location information of the ligament insertion sites and meniscal root attachments conveniently. To obtain an accurate anatomical structure, specific MRI sequences and configurations may be required for specific tissue structures [28]-[30], or one has to acquire data in vitro [28], [31]. Such procedures are generally expensive and often impractical for broad clinical applications. The supervised structure learning and prediction method developed in this study has a potential to provide accurate information of soft tissue insertion sites only using the tibia outlines that can be reliably and easily acquired from imaging data of the knee. The efficient quantitative analysis framework facilitates establishing a clinical applicable tool to assist surgeons with identifying soft tissue insertion sites, based on which a close replication of the native anatomy can be created and the risk of iatrogenic injury to adjacent tissue structures can be minimized. The quantitative modeling and analysis methodology in this study is among the first efforts to facilitate the highly demanded personalized surgical planning and achieve more precise and better-navigated surgeries in anatomical reconstruction of cruciate ligaments.

REFERENCES

- [1] S. Plaweski, D. Petek, and D. Saragaglia, "Morphometric analysis and functional correlation of tibial and femoral footprints in anatomical and single bundle reconstructions of the anterior cruciate ligament of the knee," *Orthopaedics Traumatol., Surgery Res.*, vol. 97, pp. S75–S79, 2011.
- [2] T. Zantop, N. Diermann, T. Schumacher, S. Schanz, F. Fu, and W. Petersen, "Anatomical and nonanatomical double-bundle anterior cruciate ligament reconstruction: Importance of femoral tunnel location on knee kinematics," *Am. J. Sports Med.*, vol. 36, pp. 678–685, 2008.
- [3] R. Simmons, S. Howell, and M. Hull, "Effect of the angle of the femoral and tibial tunnels in the coronal plane and incremental excision of the posterior cruciate ligament on tension of an anterior cruciate ligament graft: an in vitro study," *J. Bone Joint Surgery*, vol. 85-A, pp. 1018–1029, 2003.
- [4] T. Iriuchishima, G. Tajima, S. Ingham, K. Shirakura, and F. Fu, "PCL to graft impingement pressure after anatomical or non-anatomical singlebundle ACL reconstruction," *Knee Surgery, Sports Traumatol., Arthrosc.*, vol. 20, pp. 964–969, 2012.
- [5] M. Galloway, E. Grood, J. Mehalik, M. Levy, S. Saddler, and F. Noyes, "Posterior cruciate ligament reconstruction. an in vitro study of femoral and tibial graft placement," *Am. J. Sports Med.*, vol. 24, pp. 437–445, 1996.
- [6] D. Biau, C. Tournoux, S. Katsahian, P. Schranz, and R. Nizard, "Acl reconstruction: A metaanalysis of functional scores," *Clin. Orthop. Related Res.*, vol. 458, pp. 180–187, 2007.
- [7] D. C. Fithian, E. W. Paxton, M. L. Stone, W. F. Luetzow, R. P. Csintalan, D. Phelan, and D. M. Daniel, "Prospective trial of a treatment algorithm for the management of the anterior cruciate ligament-injured knee," *Am. J. Sports Med.*, vol. 33, pp. 335–346, 2005.
- [8] W. E. Garrett, M. F. Swiontkowski, J. N. Weinstein, J. Callaghan, R. N. Rosier, D. J. Berry, J. Harrast, and G. P. Derosa, "American board of orthopaedic surgery practice of the orthopaedic surgeon: Part-II, certification examination case mix," *J. Bone Joint Surg. Am.*, vol. 88, pp. 660–667, 2006.
- [9] M. Ferretti, D. Doca, S. Ingham, M. Cohen, and F. Fu, "Bony and soft tissue landmarks of the ACL tibial insertion site: An anatomical study," *Knee Surgery Sports Traumatol. Arthrosc.*, vol. 20, pp. 62–68, 2012.
- [10] H. Gadikota, J. Sim, A. Hosseini, T. Gill, and G. Li, "The relationship between femoral tunnels created by the transtibial, anteromedial portal, and outside-in techniques and the anterior cruciate ligament footprint," *Am. J. Sports Med.*, vol. 40, pp. 882–888, 2012.
- [11] J. H. Lubowitz, "Anteromedial portal technique for the anterior cruciate ligament femoral socket: Pitfalls and solutions," *Arthroscopy*, vol. 25, pp. 95–101, 2009.
- [12] C. D. Harner, G. H. Baek, T. M. Vogrin, G. J. Carlin, S. Kashiwaguchi, and S. L.-Y. Woo, "Quantitative analysis of human cruciate ligament insertions," *Arthrosc. J. Arthrosc. Related Surgery*, vol. 15, no. 7, pp. 741–749, 1999.
- [13] R. Siebold, T. Ellert, S. Metz, and J. Metz, "Tibial insertions of the anteromedial and posterolateral bundles of the anterior cruciate ligament: morphometry, arthroscopic landmarks, and orientation model for bone tunnel placement," *Arthroscopy*, vol. 24, pp. 154–161, 2008.
- [14] K. Li, M. O'Farrell, D. Martin, S. Kopf, C. Harner, and X. Zhang, "Mapping ligament insertion sites onto bone surfaces in knee by co-registration of CT and digitization data," *J. Biomech.*, vol. 42, no. 15, pp. 2624–2626, 2009.
- [15] A. Edwards, A. Bull, and A. Amis, "The attachments of the anteromedial and posterolateral fibre bundles of the anterior cruciate ligament: Part 1: tibial attachment," *Knee Surgery, Sports Traumatol., Arthrosc.*, vol. 15, pp. 1414–1421, 2007.
- [16] G. Tajima, M. Nozaki, T. Iriuchishima, S. Ingham, W. Shen, P. Smolinski, and F. Fu, "Morphology of the tibial insertion of the posterior cruciate ligament," *J. Bone Joint Surgery*, vol. 91, pp. 859–866, 2009.
- [17] L. Zheng, C. D. Harner, and X. Zhang, "The morphometry of soft tissue insertions on the tibial plateau: Data acquisition and statistical shape analysis," *PLoS ONE*, vol. 9, no. 5, p. e96515, May 2014.
- [18] A. Argyriou, T. Evgeniou, and M. Pontil, "Multi-task feature learning," in Proc. Int. Conf. Adv. Neural Inf. Process. Syst., 2007, pp. 41–48.
- [19] F. YJ and W. Chaovalitwongse, "Optimizing feature selection to improve medical diagnosis," *Annal. Oper. Res.*, vol. 174, no. 1, pp. 169–183.
- [20] W. Chaovalitwongse, Y. Jeong, M. K. Jeong, S. Danish, and S. Wong, "Pattern recognition approaches for identifying subcortical targets during deep brain stimulation surgery," *IEEE Intelli. Syst.*, vol. 26, no. 5, pp. 54–63, Sep./Oct. 2011.

- [21] O. Seref, Y. Fan, and W. Chaovalitwongse, "Mathematical programming formulations and algorithms for discrete k-median clustering of timeseries data," *INFORMS J. Comput.*, vol. 26, no. 1, pp. 160–172, 2013.
- [22] R. Tibshirani, "Regression shrinkage and selection via the lasso," J. Royal Statist. Soc. B, vol. 58, pp. 267–288, 1996.
- [23] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.
- [24] B.-K. Yi, H. Jagadish, and C. Faloutsos, "Efficient retrieval of similar time sequences under time warping," in *Proc. 14th Int. Conf. Data Eng.*, Feb. 1998, pp. 201–208.
- [25] E. Pena, B. Calvo, M. Martinez, and M. Doblare, "A three-dimensional finite element analysis of the combined behavior of ligaments and menisci in the healthy human knee joint," *J. Biomech.*, vol. 39, pp. 1686–1701, 2006.
- [26] T. Guess, G. Thiagarajan, M. Kia, and M. Mishra, "A subject specific multibody model of the knee with menisci," *Med. Eng. Phys.*, vol. 32, pp. 505–515, 2010.
- [27] A. Sandholm, C. Schwartz, N. Pronost, M. de Zee, M. Voigt, and D. Thalmann, "Evaluation of a geometry-based knee joint compared to a planar knee joint," *Visual Comput.*, vol. 27, no. 2, pp. 161–171, 2011.
- [28] G. Li, J. Gi, A. Kanamori, and S. Woo, "A validated three-dimensional computational model of a human knee joint," *J. Biomech. Eng.*, vol. 121, pp. 657–662, 1999.
- [29] J. Brody, M. Hulstyn, B. Fleming, and G. Tung, "The meniscal roots: Gross anatomic correlation with 3-t MRI findings," *Am. J. Roentgenol.*, vol. 188, pp. W446–W450, 2007.
- [30] P. Araujo, C. van Eck, M. Torabi, and F. Fu, "How to optimize the use of MRI in anatomic ACL reconstruction," *Knee Surgery, Sports Traumatol.*, *Arthrosc.*, vol. 21, pp. 1495–1501, 2012.
- [31] T. Donahue, M. Hull, M. Rashid, and C. Jacobs, "A finite element model of the human knee joint for the study of tibio-femoral contact," *J. Biomech. Eng.*, vol. 124, pp. 273–280, 2002.



Cao Xiao (S'14) received the Bachelor's degree from Peking University, Beijing, China, the Master's degree in mathematics from Boston College, Boston, MA, USA, and the Ph.D. degree in industrial and systems engineering from the University of Washington, Seattle, WA, USA.

Her current research interests include data mining and optimization, machine learning, intelligent decision-making systems, and big data applications. Particularly, she enjoys developing efficient datamining methods to enable effective problem solving

for bioinformatics and social network applications.



Shouyi Wang (M'11) received the B.S. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2003, the M.S. degree in systems and control engineering from the Delft University of Technology, Delft, The Netherlands, in 2005, and the Ph.D. degree in industrial and systems engineering from Rutgers University, New Brunswick, NJ, USA, in 2012.

He is currently an Assistant Professor with the Department of Industrial, Manufacturing, and Systems Engineering and the Center on Stochastic Modeling,

Optimization, and Statistics, University of Texas at Arlington, Arlington, TX, USA.



Liying Zheng received the B.S. degree in mechanical engineering and automation from the East China University of Science and Technology, Shanghai, China, in 2002, and the Ph.D. degree in mechanical engineering from Washington State University, Pullman, WA, USA, in 2011.

She is currently a Postdoctoral Associate with the Departments of Mechanical Engineering and Materials Science and Orthopaedic Surgery, University of Pittsburgh, Pittsburgh, PA, USA. Her research interests include the areas of biomechanics, muscu-

loskeletal modeling and simulation, morphometrics, and advanced data analysis for improving the prevention, diagnosis, and treatment of musculoskeletal disorders.



Wanpracha Art Chaovalitwongse (M'05–SM'11) received the M.S. and Ph.D. degrees in industrial and systems engineering from the University of Florida, Gainesville, FL, USA, in 2000 and 2003, respectively.

He is currently a Professor with the Departments of Industrial and Systems Engineering, Radiology (joint), and Bioengineering (adjunct), University of Washington (UW), Seattle, WA, USA. He also serves as an Associate Director of the Integrated Brain Imaging Center, UW Medical Center. His research group conducts basic computational science, applied, and

translational research at the interface of engineering, medicine, and other emerging disciplines.



Xudong Zhang received the B.S. degree in mechanical engineering from Tsinghua University, Beijing, China, and the M.S. and Ph.D. degrees in industrial and operations engineering from the University of Michigan, Ann Arbor, MI, USA.

He is currently a Professor with Departments of Mechanical Engineering, Bioengineering, and Orthopedic Surgery, University of Pittsburgh, Pittsburgh, PA, USA. Previously, he was on the faculty of mechanical and industrial engineering with the University of Illinois at Urbana-Champaign.