An Efficient $L_{2,1}$-Regularized and Structure-Supervised Multi-Response Prediction Model with Application to Tibial Soft Tissue Insertion Predictions

Kin Ming Puk, Shouyi Wang, *Member, IEEE,* Jay Rosenberger, Cao Xiao, *Member, IEEE,* Liying Zheng, Xudong Zhang, and Wanpracha Art Chaovalitwongse, *Senior Member, IEEE*

***Abstract*—The coexistence of high dimensionality in both responses and feature space with limited sample size poses unprecedented challenges in identifying important predictive relationships by current multi-response prediction models. This paper proposes an efficient high-dimensional multi-response learning and prediction framework which integrates intrinsic data-structure learning with an $L_{2,1}$-norm regularized sparse leaning method to construct multi-response prediction models under the high-dimension-low-sample-size setting. In particular, motivated by a challenging biomedical prediction problem, the proposed learning framework has been successfully applied to learn intrinsic knee structures and predict natural insertions of cartilages and soft tissues in knee reconstruction surgery from high dimensional 3D CT images. Such prediction problem is challenging because of the large number of features (508) and response variables (24), as opposed to the limited number of samples (20). The proposed multi-step structure-learning and prediction framework provides an efficient approach to precisely identify cruciate ligament natural positions in a 3D spatial space with a limited sample size of 20. The proposed method outperformed the state-of-the-art multi-response prediction models in the experiment. As a general learning and prediction framework, it can also be applied to other prediction problems that have both high dimensional feature and response variable space and low sample size.**

***Index Terms*—cruciate ligament, meniscal insertions, knee anatomic reconstruction, structured supervised learning, X-ray imaging, computed tomogrpahy, multi-response prediction, support vector regression, $L_{2,1}$ regularization**

I. Introduction

THE knee is a complex joint that provides basic human function and mobility, enabling ambulatory or sports activities such as walking, running, and jumping [4]. The structural components of a knee joint include cartilage, ligaments (connecting one bone to another), tendons (connecting muscle to bone) and bones (connected by ligaments, or by tendons to muscles). As a union of bones, the knee is where the thigh bone (femur) joins the shin bone (tibia) to form the tibiofemoral joint, and the kneecap (patella) joins the femur to form the patellofemoral joint. Articular cartilage covers the knee joint, helps absorb shock, and improves joint movement. Over-use and excessive stress or force can make the knee prone to a variety of injuries. Typical knee injuries include fractures, dislocations, anterior cruciate ligament (ACL) injuries, posterior cruciate ligament (PCL) injuries, collateral ligament injuries, meniscal tears and tendon tears [20], [42]. Injuries to the cruciate ligaments (ACL and PCL) are particularly devastating as they do not heal and would most often require surgical reconstruction. Among the two, the ACL is more commonly injured: an estimated 175,000 ACL reconstructions are performed annually, with a financial impact exceeding 2 billion dollars in the United States alone. On the other hand, PCL injuries occur less frequently and are thus believed to be under diagnosed; 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 bone tunnels and placing substituting grafts in the exact 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 [18], [24], [32], [36], [44]. However, current approaches to treating knee injuries may be inconsistent and ineffective in restoring knee function and preventing the development of osteoarthritis (OA). Analyses of long-term outcomes after ACL reconstruction have revealed that only 37% of patients were restored back to normal in terms of knee structure and function [3], and 90% of ACL reconstructed knees exhibited radiographic evidence of OA 3-10 years after injury [15]. A growing body of clinical evidence suggests that anatomic reconstruction can better restore joint function and deter the development of OA. However, a number of challenges are present in the practice of anatomic reconstruction of the cruciate ligaments.

Intraoperative identification of native cruciate ligament insertion sites, as a requisite for anatomical reconstruction,

K. Puk, S. Wang and J. Rosenberger are with the Department of Industrial and Manufacturing Systems Engineering, University of Texas, Arlington, TX 76019 USA (e-mail: pukkinming@gmail.com, shouyiw@uta.edu, jrosenbe@uta.edu).

C. Xiao is with Center for Computational Health, Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: cxiao@us.ibm.com).

L. Zheng is with Health Effects Lab Division, National Institute for Occupational Safety and Health, 1095 Willowdale Rd, Morgantown, WV 26505 USA (e-mail: lzheng2@cdc.gov).

X. Zhang is with the Departments of Industrial & Systems Engineering and Mechanical Engineering, Texas A&M University, Emerging Technologies Building, Suite 4077 College Station, TX 77843-3131 USA (e-mail: xudongzhang@tamu.edu).

W. A. Chaovalitwongse is with the Department of Industrial Engineering, University of Arkansas, Fayetteville, AR 72701 USA (e-mail: artchao@uw.edu).

Manuscript received XXX XX, 2017; revised XXX XX, 2017.

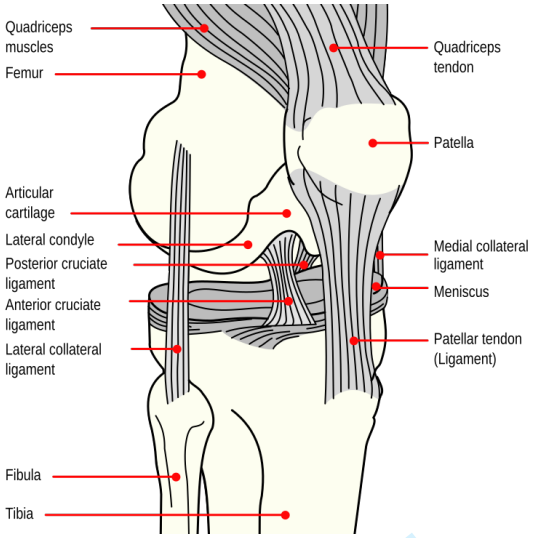

Fig. 1. Right knee, seen from an angle between anteriorly and laterally [11]. The three main parts of the knee are the thigh bone (femur), the shin bone (tibia) and the kneecap (patella). The femur joins the tibia to form the main joint, whereas the patella meets the femur to make another joint [4].

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 [19], 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) [14]. 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 such as meniscus roots. 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 [17], [28]. The key to anatomic surgery of cruciate ligaments with minimized risk of iatrogenic injury is accurate, quantitative knowledge 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 cruciate ligaments and other soft tissue components using statistical and quantitative approaches [22], [27], [35]. However, such quantitative analysis and measures cannot fully capture accurate spatial arrangement of soft tissues. The location and morphological measures cannot account for the interperson variability of cruciate ligaments and meniscus insertions, which are mostly characterized by qualitative measures [13], [41]. While in 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, identification of insertion sites of cruciate ligaments and meniscus roots in knee reconstruction surgery presents a great challenge intra-operatively or pre-operatively during surgical planning.

Predicting locations of multiple soft-tissue insertion sites simultaneously requires multi-target prediction models, also known as multi-response or multi-output learning. According to [5], [45], there are mainly two types of multi-response learning approaches:

- Problem transformation methods transform multi-target regression into other well-established, single-target learning models. One drawback of this type of methods is that the relationships among the targets are often ignored so that the prediction performance might not be satisfactory. Examples include multi-target regressor stacking [38], which uses the prediction of the first response as part of the input data in predicting the next response variable, and multi-output support vector regression (SVR) [46], which considers the correlations between the targets using a vector virtualization method.
- Algorithm adaption methods learn by considering all dependencies and internal relationships between response variables. This type of methods is easier to interpret and produces better learning performance when response variables are correlated. Examples include statistical methods [25], multi-output SVR [9], [33], kernel methods [1], regression trees [12] and more. Interested readers can refer to [5] for a list of the state-of-the-art algorithms in this category.

This study presents a new multi-response regression framework in order to achieve accurate prediction of the insertion sites for the six cruciate ligaments and two cartilage surfaces (collectively referred to as "cartilage and soft issues" in the following). The challenge of this study lies in the large number of features (508) and response variables (24) as opposed to the limited number of training samples (20). In particular, the proposed framework first digitalized outlines of the tibia from three-dimensional computed tomography (CT) images and aligned the outlines using generalized procrustes analysis (GPA) techniques. It then extracted patient-specific morphological features of the knee. Last but not least, the three-step structure-learning model identifies the insertion locations of all cartilages and soft tissues simultaneously.

In the first learning step, an $L_{2,1}$ sparsity-inducing regularization was incorporated into the multi-response support vector regression (SVR) to cope with the challenge of having more features and response variables than observations. Secondly, sparse learning model was constructed to learn the intrinsic spatial relationship between the cartilages and soft tissues. Finally, a structure-supervised prediction model that considers both the feature-response relationship from step 1 and the intrinsic spatial relationship from step 2 was used to simultaneously predict the optimal insertion centroids of the cartilages and soft tissues.

In summary, the contribution of this study includes the following: 1) a new method of feature extraction was developed to capture morphological features of the knee from aligned

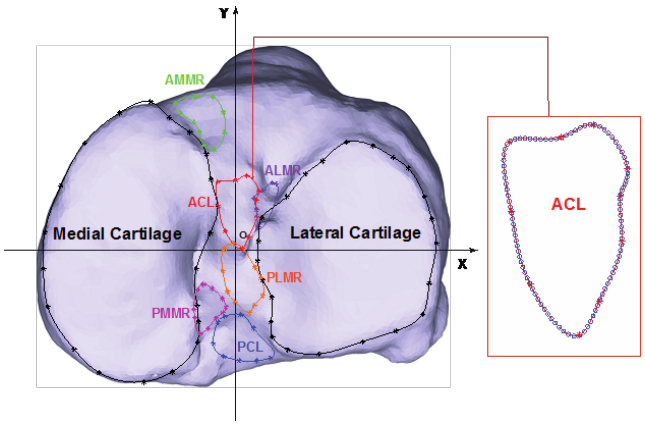

Fig. 2. Digitized cartilage and insertion site outlines mapped onto the CT-based 3-D tibia model. The digitized points (asterisks) were spline-fitted, generating 100 equidistant points (circles on the close-up view of ACL insertion outline) on the fitted outlines to facilitate the subsequent analyses [27], [47].

three-dimensional CT scan images; 2) a new $L_{2,1}$-regularized, $L_{2,2}$-loss multi-response SVR model was developed and solved via the Fast Iterative Shrinkage-Threshold Algorithm (FISTA); and 3) a novel three-step structure-learning and structure-supervised prediction framework was developed. The proposed framework outperforms the state-of-the-art multi-response regression models.

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 years) were used to acquire the morphometric data [47]. 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), the anterior and posterior medial meniscal roots (AMMR and PMMR), and the anterior-lateral and posterior-lateral meniscal roots (ALMR and PLMR). The digitization was performed by the same experimenter with 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% [27]. A closed spline was then fitted to each outline, resulting in 100 equidistant discrete points to represent the outline, as shown in figure 2.

A 3-D coordinate system was defined on each tibia based on its digitized and mapped cartilage outlines (see figure 2). First, the origin of the coordinate system was determined as the midpoint of the medial and lateral cartilage centroids. 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 through the origin and pointing laterally. The Y-axis was orthogonal to the X-axis, passing through 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 left-handed 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 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. 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. Figure 3 shows the outlines of tibial cartilage and six insertion sites from 20 subjects before and after cartilage-based GPA.

C. Feature Engineering Using Spherical Coordinates

The outline of the tibia can be easily and reliably measured, making it a feature candidate to predict the locations of intangible soft tissues. Features are selected via a spherical coordinate system, which consists of radial distance (r), polar angle (θ) and azimuthal angle (ϕ). See figure 4 for an illustration.

A 3-D point with Cartesian coordinates (x, y, z) can be transformed into polar coordinates (r, θ, ϕ) following the rules: $r = \sqrt{x^2 + y^2 + z^2}$, $\theta = \text{arc cos}(z/r)$, and $\phi = \text{arc tan}(y/x)$. Thus the direction of a tibia point from the origin can be described by the two angles (θ, ϕ). The entire range of each angle ($[-\pi, \pi]$) is divided into N equal intervals, resulting in a combination of $N \times N = N^2$ directions from the origin. Points along the outline tibia that are closest to these N^2 directions will be selected as features in the form of Cartesian coordinates. In this regard, N was chosen to be 13, and thus the number of selected features of each subject is $N^2 \times 3 = 507$.

In addition, raw CT image data consist of knees from left and right legs. Another useful feature is whether it is a left or right knee. Therefore the total number of features is $507 + 1 = 508$.

D. Response Variable

There is high variability in ligament and meniscus insertions. In a surgical procedure, the centroids of insertions are of particular importance to tunnel locations. Thus, this work focuses on the prediction of centroids of the ligament and

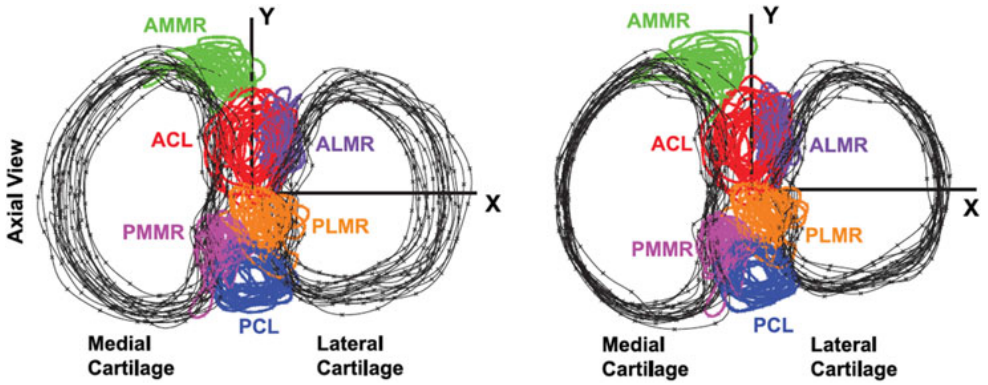

Fig. 3. Outlines of tibial cartilage and six insertion sites from 20 subjects before (left plot) and after (right plot) cartilage-based GPA [27], [47].

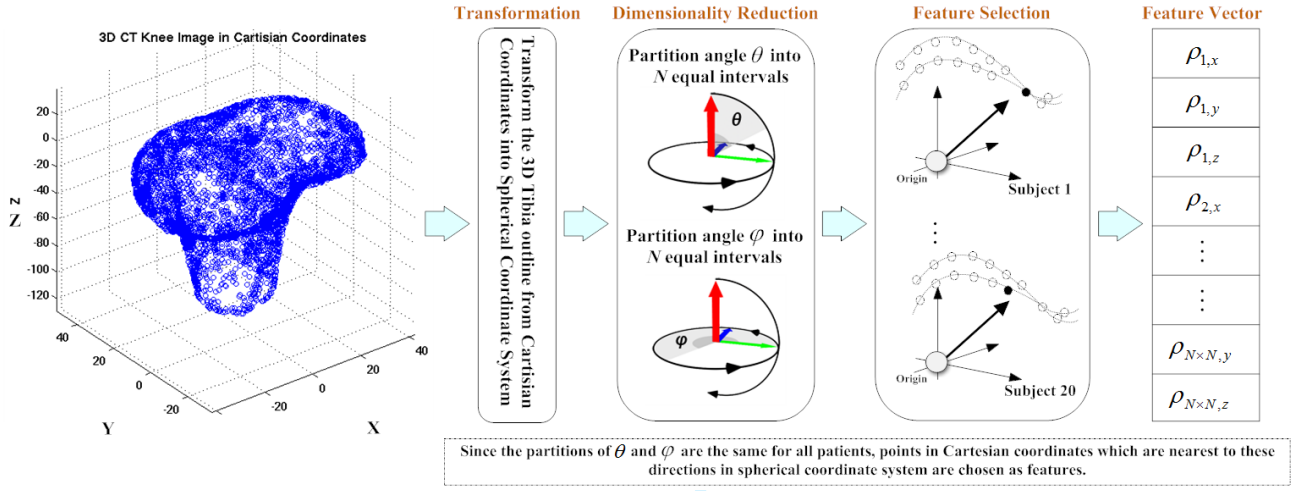

Fig. 4. A visual illustration of feature selection. Points in the tibia feature space are converted to a spherical coordinate system. The entire range of each angle in the spherical coordinate is divided into $N = 13$ equal intervals, rendering 169 combinations for θ and ϕ. Outline points that are nearest to these directions are chosen as features in the form of Cartesian coordinates. Together with the side of the knee (left or right), there are $169 \times 3 + 1 = 508$ features for each observation.

meniscus insertions instead of the complete morphology of insertions.

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

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

E. $L_{2,1}$-Regularized, $L_{2,2}$-Loss Support Vector Regression - Version 1

The proposed regression model[1] is an algorithm adaption method, which considers all the response variables at the same time during learning. Recall the formulation of L_2-regularized, L_2-loss support vector regression (SVR) model, where N, M and P are the number of observations, features and response variables respectively, as follows:

$$\begin{aligned} \min_{\boldsymbol{w}_p, \boldsymbol{\xi}, \boldsymbol{\xi}^*} \quad & \sum_{p=1}^{P} \left[||\boldsymbol{w}_p||_2 + C \sum_{n=1}^{N} (\xi_{pn})^2 + (\xi_{pn}^*)^2 \right] \\ \text{s.t.} \quad & y_{pn} - \boldsymbol{w}_p^T \boldsymbol{x_n} \leq \epsilon + \xi_{pn}, \forall p = 1, \ldots, P, n = 1, \ldots, N \\ & \boldsymbol{w}_p^T \boldsymbol{x_n} - y_{pn} \leq \epsilon + \xi_{pn}^*, \forall p = 1, \ldots, P, n = 1, \ldots, N \\ & \boldsymbol{\xi}, \boldsymbol{\xi}^* \geq 0 \end{aligned} \quad (2)$$

$$f_p(\boldsymbol{x}) = \boldsymbol{w}_p^T \boldsymbol{x} + b, \ \ \forall p = 1, \ldots, P \quad (3)$$

The bias term b is eliminated for simplifying the presentation. It can be achieved by adding a column of one into the data matrix (or design matrix). Furthermore, the above can be generalized in matrix terms as follows, where $\mathbf{X} \in \mathbb{R}^{N \times M}$, $\mathbf{Y} \in \mathbb{R}^{N \times P}$, $\mathbf{W} \in \mathbb{R}^{M \times P}$, $\boldsymbol{\xi}, \boldsymbol{\xi}^*, \mathbf{1} \in \mathbb{R}^{N \times P}$ and $||\cdot||_{\mathrm{F}}$ is the

[1] In the model, a scalar is represented by a non-bold lower-case letter, a vector is represented by a bold lower-case letter and a matrix is represented by a bold upper-case letter.

Frobenius Norm:

$$\begin{aligned} \min_{\mathbf{W},\boldsymbol{\xi},\boldsymbol{\xi}^*} \quad & ||\mathbf{W}||_{\mathrm{F}} + C||\boldsymbol{\xi}||_{\mathrm{F}} + C||\boldsymbol{\xi}^*||_{\mathrm{F}} \\ \text{s.t.} \quad & \mathbf{Y} - \mathbf{XW} \leq \epsilon\mathbf{1} + \boldsymbol{\xi} \\ & \mathbf{XW} - \mathbf{Y} \leq \epsilon\mathbf{1} + \boldsymbol{\xi}^* \\ & \boldsymbol{\xi}, \boldsymbol{\xi}^* \geq 0 \end{aligned} \tag{4}$$

$$f(\mathbf{X}) = \mathbf{XW} \tag{5}$$

The unconstrained version of formulation (4) is:

$$\begin{aligned} \min_{\mathbf{W}} \; J(\mathbf{W}) = ||\mathbf{W}||_{\mathrm{F}} + C||(\mathbf{Y} - \mathbf{XW} - \epsilon\mathbf{1})_+||_{\mathrm{F}} \\ + C||(\mathbf{XW} - \mathbf{Y} - \epsilon\mathbf{1})_+||_{\mathrm{F}} \end{aligned} \tag{6}$$

where the hinge loss operator is $(\cdot)_+ = \max(\cdot, 0)$. To further modify the formulation, consider the $L_{2,1}$-norm, which is defined as:

$$||\mathbf{W}||_{2,1} = \sum_{m=1}^{M} \sqrt{\sum_{p=1}^{P} w_{mp}^2} = \sum_{m=1}^{M} ||\mathbf{w}_m||_2 \tag{7}$$

where $\mathbf{w}_m$ denotes the m-th row vector of matrix $\mathbf{W}$. The following formulation of $L_{2,1}$-regularized, $L_{2,2}$-loss multi-response support vector regression (L21-MSVR) - Version 1, is proposed:

$$\begin{aligned} \min_{\mathbf{W}} \; J_1(\mathbf{W}) = ||\mathbf{W}||_{2,1} + C||(\mathbf{Y} - \mathbf{XW} - \epsilon\mathbf{1})_+||_{\mathrm{F}} \\ + C||(\mathbf{XW} - \mathbf{Y} - \epsilon\mathbf{1})_+||_{\mathrm{F}} \end{aligned} \tag{8}$$

1) Motivation: As opposed to ridge regression in [43], support vector regression is chosen to be the base algorithm. It generally performs better than ridge regression or lasso regression with careful parameter tuning at the expense of one more parameter (ϵ) to tune. Instead of training the projection vector $\mathbf{w}_m$ one-by-one in formulation (2), L21-MSVR trains the projection matrix $\mathbf{W}$ considering all responses.

The $L_{2,1}$-norm is used instead of the $L_{2,2}$ norm in order to encourage row sparsity, where the number of rows corresponds to the number of features. As a result, the $L_{2,1}$-norm encourages parsimony of features, which is itself a method of feature selection for features in the matrix representation [10], [29]. It is especially sparse when the number of features is a lot larger than that of the observations.

2) Subgradient of Formulation: Specifically, the derivative of $||\mathbf{W}||_{2,1}$ with respect to (w.r.t.) $\mathbf{W}$ equals the derivative of the trace$(\mathbf{W}^T\mathbf{DW})$ w.r.t. $\mathbf{W}$ [10], where $\mathbf{D} \in \mathbb{R}^{M\times M}$ is a diagonal matrix in which each element on the diagonal is defined as $d_{mm} = \frac{1}{2||\mathbf{w}_m||_2}$, $\forall m \in 1, \ldots, M$. In view of this, the subgradient of formulation (8) for each observation i can be derived as follows [31], [34]:

$$\begin{aligned} \frac{\partial J_1(\mathbf{W})}{\partial \mathbf{W}} &= \frac{\partial}{\partial \mathbf{W}} \; ||\mathbf{W}||_{2,1} \\ &\quad + C\sum_{n=1}^{N} \frac{\partial}{\partial \mathbf{W}} ||(\mathbf{y}_n - \mathbf{W}^T\mathbf{x}_n - \epsilon\mathbf{1})_+||_{\mathrm{F}} \\ &\quad + C\sum_{n=1}^{N} \frac{\partial}{\partial \mathbf{W}} ||(\mathbf{W}^T\mathbf{x}_n - \mathbf{y}_n - \epsilon\mathbf{1})_+||_{\mathrm{F}} \\ &= \frac{\partial}{\partial \mathbf{W}} \; \mathrm{tr}(\mathbf{W}^T\mathbf{DW}) \\ &\quad + C\sum_{n=1}^{N} \frac{\partial}{\partial \mathbf{W}} ||\mathbf{y}_n - \mathbf{W}^T\mathbf{x}_n - \epsilon\mathbf{1}||_{\mathrm{F}} \\ &\qquad \circ \mathbb{I}[\mathbf{y}_n - \mathbf{W}^T\mathbf{x}_n > \epsilon\mathbf{1}] \\ &\quad + C\sum_{n=1}^{N} \frac{\partial}{\partial \mathbf{W}} ||\mathbf{W}^T\mathbf{x}_n - \mathbf{y}_n - \epsilon\mathbf{1}||_{\mathrm{F}} \\ &\qquad \circ \mathbb{I}[\mathbf{W}^T\mathbf{x}_n - \mathbf{y}_n > \epsilon\mathbf{1}] \end{aligned} \tag{9}$$

where

$$\begin{aligned} &\frac{\partial}{\partial \mathbf{W}} ||\mathbf{y}_n - \mathbf{W}^T\mathbf{x}_n - \epsilon\mathbf{1}||_{\mathrm{F}} \\ =& 2\mathbf{x}_n^T\mathbf{x}_n\mathbf{W} - 2\mathbf{x}_n\mathbf{y}_n^T + 2\epsilon\mathbf{x}_n\mathbf{1}^T \end{aligned} \tag{10}$$

$$\begin{aligned} &\frac{\partial}{\partial \mathbf{W}} ||\mathbf{W}^T\mathbf{x}_n - \mathbf{y}_n - \epsilon\mathbf{1}||_{\mathrm{F}} \\ =& 2\mathbf{x}_n^T\mathbf{x}_n\mathbf{W} - 2\mathbf{x}_n\mathbf{y}_n^T - 2\epsilon\mathbf{x}_n\mathbf{1}^T \end{aligned} \tag{11}$$

$$\begin{aligned} &\mathbb{I}[\mathbf{y}_n - \mathbf{W}^T\mathbf{x}_n > \epsilon\mathbf{1}] \\ =& \underbrace{[\mathbb{1}[\mathbf{y}_n - \mathbf{W}^T\mathbf{x}_n > \epsilon\mathbf{1}] \quad \ldots \quad \mathbb{1}[\mathbf{y}_n - \mathbf{W}^T\mathbf{x}_n > \epsilon\mathbf{1}]]^T}_{\text{replication of the column vector } \mathbb{1} \text{ for } M \text{ times}} \end{aligned} \tag{12}$$

$$\begin{aligned} &\mathbb{I}[\mathbf{W}^T\mathbf{x}_n - \mathbf{y}_n > \epsilon\mathbf{1}] \\ =& \underbrace{[\mathbb{1}[\mathbf{W}^T\mathbf{x}_n - \mathbf{y}_n > \epsilon\mathbf{1}] \quad \ldots \quad \mathbb{1}[\mathbf{W}^T\mathbf{x}_n - \mathbf{y}_n > \epsilon\mathbf{1}]]^T}_{\text{replication of the column vector } \mathbb{1} \text{ for } M \text{ times}} \end{aligned} \tag{13}$$

In the above, $\circ$ is the Hadamard product of matrices of equal size, and $\mathbb{1}$ and $\mathbb{I}$ are indicator functions that take the value of one if its argument is true and zero otherwise. The size of $\mathbb{1}$ and $\mathbb{I}$ are $\mathbb{R}^P$ and $\mathbb{R}^{M\times P}$ respectively. Therefore, the size of $\mathbf{W}$, $\frac{J_1(\mathbf{W})}{\partial \mathbf{W}}$ and $\mathbb{I}$ are all the same.

The design of having the Hadamard product on the projection matrix $\mathbf{W}$ in (9) is in accordance with [34] so that the subgradient of an observation n will become zero if $\mathbf{y}_n - \mathbf{W}^T\mathbf{x}_n - \epsilon\mathbf{1}$ (or $\mathbf{W}^T\mathbf{x}_n - \mathbf{y}_n - \epsilon\mathbf{1}$) is less than zero. For details, please refer to equations (10)-(13).

3) Optimization Algorithm - FISTA: Formulation (8) can be easily solved with a commercial convex programming package such as CVX [21]. However, doing so may take more than a week in order to tune the parameters and get an optimal solution. In order to enhance computational efficiency, the FISTA (Fast ISTA) algorithm [8], which combines Nesterov's accelerated gradient descent with ISTA (iterative shrinkage-threshold algorithm), is chosen to efficiently solve formulation (8) because i) the formulation is in the form of standard,

unconstrained quadratic programming; ii) a subgradient can be computed for the objective function consisting of $L_{2,2}$-norm and $L_{2,1}$-norm; iii) FISTA is easy to implement [2]; and iv) instead of having a sublinear convergence rate of $O(1/j)$ as in gradient descent, FISTA enjoys a favorable convergence rate of $O(1/j^2)$, where j is the number of iterations.

4) Implementation Specifics of the Algorithm: When $\mathbf{w}_m = 0$, then $d_{mm} = 0$ is a subgradient of $||\mathbf{W}||_{2,1}$, w.r.t. $\mathbf{w}_m$ [10]. However, if d_{mm} is set to 0 when $\mathbf{w}_m = 0$, the derived algorithm may not converge (i.e. $d_{mm} = \frac{1}{0} = \text{inf}$). Therefore, a small number ρ is used for regularization as in $d_{mm} = \frac{1}{2\sqrt{(\mathbf{w}_m)^T\mathbf{w}_m+\rho}}$.

There are a lot of choices for termination criteria such as the difference of objective values between successive iterations. For better computational efficiency, the ratio of the Frobenius Norms $\left(\frac{||\mathbf{W}^{(j+1)}-\mathbf{W}^{(j)}||_{\text{F}}}{||\mathbf{W}^{(j)}||_{\text{F}}}\right)$ is used at [10] in order to avoid the expensive computation of the objective value.

Throughout the computation, terms in equations (10) and (11) such as $2C\mathbf{x}_n^T\mathbf{x}_n$, $2C\mathbf{x}_n\mathbf{y}_n^T$ and $2\epsilon C\mathbf{x}_n\mathbf{1}^T$ will remain unchanged in different iterations. These terms can be precomputed (or cached in some literature) first and passed to the appropriate function/procedure for direct reuse. As the number of observations and features are relatively small, memory occupied by these cache variables would be acceptable, but surprisingly one-third of computational time can be reduced in practice.

In practice, the starting step size γ and the parameter σ in backtracking ($\gamma = \sigma\gamma$) are set to 1 and 0.8 respectively [6].

Finally, a summary of FISTA can be found in Algorithm (1). Its convergence is well analyzed and can be found in [2].

F. $L_{2,1}$-Regularized, $L_{2,2}$-Loss Support Vector Regression - Version 2 (Relaxing Non-Negativity Constraints of Slack Variables)

As a heuristics to further enhance computational efficiency, the formulation (6) is reformulated with equality loss constraints as follows [40]:

$$\begin{aligned} \min_{\mathbf{W},\boldsymbol{\xi},\boldsymbol{\xi}^*} \quad & ||\mathbf{W}||_{2,1} + C||\boldsymbol{\xi}||_{\text{F}} + C||\boldsymbol{\xi}^*||_{\text{F}} \\ \text{s.t.} \quad & \mathbf{Y} - \mathbf{XW} = \epsilon\mathbf{1} + \boldsymbol{\xi} \\ & \mathbf{XW} - \mathbf{Y} = \epsilon\mathbf{1} + \boldsymbol{\xi}^* \end{aligned} \tag{14}$$

$$\begin{aligned} \min_{\mathbf{W}} \; J_2(\mathbf{W}) = {} & ||\mathbf{W}||_{2,1} + C||\mathbf{Y} - \mathbf{XW} - \epsilon\mathbf{1}||_{\text{F}} \\ & + C||\mathbf{XW} - \mathbf{Y} - \epsilon\mathbf{1}||_{\text{F}} \end{aligned} \tag{15}$$

The reformulation of the slack variables $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^*$ removes the hinge-loss operator so that the gradient of the Frobenius loss term can be computed directly instead of computing its subgradient. The derivative of formulation (15) is consequently as follows [31]:

$$\begin{aligned} \frac{\partial J_2(\mathbf{W})}{\partial \mathbf{W}} &= \frac{\partial}{\partial \mathbf{W}} \; ||\mathbf{W}||_{2,1} \\ &\quad + C\frac{\partial}{\partial \mathbf{W}}||\mathbf{Y} - \mathbf{XW} - \epsilon\mathbf{1}||_{\text{F}} \\ &\quad + C\frac{\partial}{\partial \mathbf{W}}||\mathbf{XW} - \mathbf{Y} - \epsilon\mathbf{1}||_{\text{F}} \\ &= \frac{\partial}{\partial \mathbf{W}} \; \text{tr}(\mathbf{W}^T\mathbf{DW}) + C(4\mathbf{X}^T\mathbf{XW} - 4\mathbf{X}^T\mathbf{Y}) \\ &= 2\mathbf{DW} + 4C\mathbf{X}^T\mathbf{XW} - 4C\mathbf{X}^T\mathbf{Y} \end{aligned} \tag{16}$$

One may be tempted to develop analytical solution as in [40]. From derivative (16), it is found that $\mathbf{W} = (2\mathbf{D} + 4C\mathbf{X}^T\mathbf{X})^{-1}4C\mathbf{X}^T\mathbf{Y}$. However, as pointed out in [10], the matrix $\mathbf{D}$ is dependent on $\mathbf{W}$ and is unknown. Therefore, an iterative optimization algorithm will be more appropriate for computation.

A summary of the revised optimization algorithm can be found in Algorithm (1). The only difference is at line 6, which has a different subgradient. The objective history for both algorithms can be found in figure 5. Both of them converge in under 300 iterations.

Algorithm 1 Solving multi-response support vector regression with $L_{2,1}$-norm regularization [2], [10]

Require: $\mathbf{X} \in \mathbb{R}^{N\times M}$ (data matrix), $\mathbf{Y} \in \mathbb{R}^{N\times P}$ (response matrix), $C, \epsilon, \alpha \in \mathbb{R}$ (learning parameters), $\rho \in \mathbb{R}$ (regularization parameter), $\sigma \in \mathbb{R}$ (backtracking parameter), tol $\in \mathbb{R}$ (termination parameter).

1: Initialize projection matrix $\mathbf{W}^{(0)} = \{w_{mp} = 0\}, \forall m, p$, $\mathbf{Z}^{(0)} = \mathbf{W}^{(0)}$, $t^{(0)} = 1$.
2: Set $j = 0$.
3: **while** termination criterion is not met (tol $> \frac{||\mathbf{W}^{(j+1)}-\mathbf{W}^{(j)}||_{\text{F}}}{||\mathbf{W}^{(j)}||_{\text{F}}}$) **do**
4: Initialize diagonal matrix $\mathbf{D}^{(j+1)}$, where the m-th diagonal element is $d_{mm}^{(j+1)} = \frac{1}{2||\mathbf{w}_m^{(j)}||_2}$ (or $d_{mm}^{(j+1)} = \frac{1}{2\sqrt{(\mathbf{w}_m^{(j)})^T\mathbf{w}_m^{(j)}+\rho}}$ if $||\mathbf{w}_m^{(j)}||_2 = 0$).
5: $t^{(j+1)} = \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4(t^{(j)})^2}$.
6: Version 1: $\mathbf{W}^{(j+1)} = \mathbf{Z}^{(j)} - \gamma\frac{\partial J_1(\mathbf{Z}^{(j)})}{\partial \mathbf{Z}}$ according to (9)-(13).
Version 2: $\mathbf{W}^{(j+1)} = \mathbf{Z}^{(j)} - \gamma\frac{\partial J_2(\mathbf{Z}^{(j)})}{\partial \mathbf{Z}}$ according to (16).
7: $\mathbf{Z}^{(j+1)} = \mathbf{W}^{(j+1)} + \frac{t^{(j)}-1}{t^{(j+1)}}(\mathbf{W}^{(j+1)} - \mathbf{W}^{(j)})$.
8: Use backtracking to find the appropriate step size γ [2], [6].
9: Set $j = j + 1$.
10: **end while**

For runtime comparison, Version 1 needs around 121 seconds to complete a leave-one-out cross validation, whereas Version 2 needs only 33 seconds.

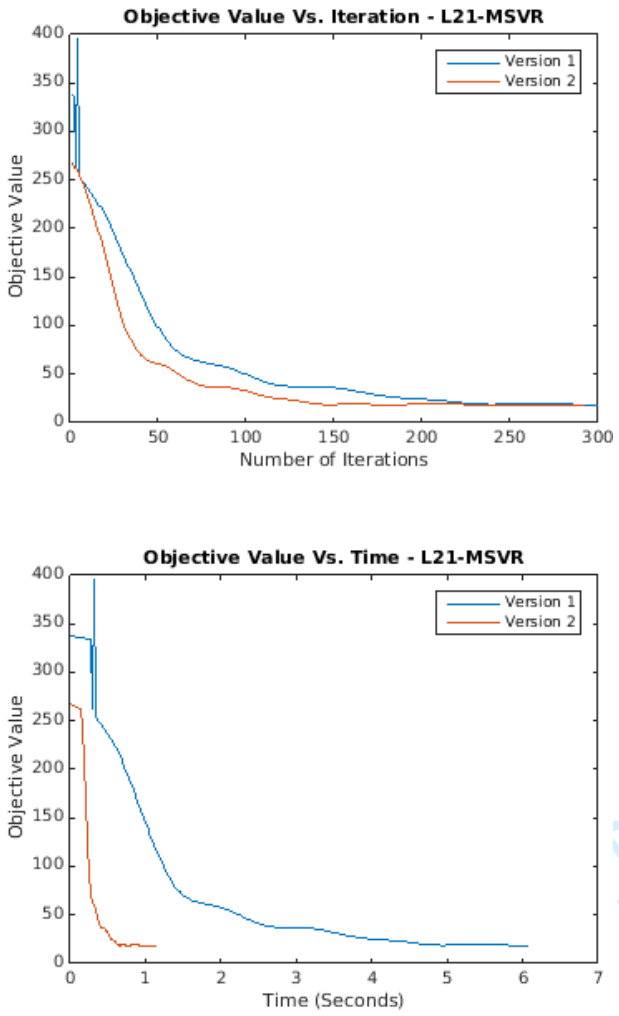

Fig. 5. Objective history for Versions 1 and 2. For Version 1, there are some oscillations for the first few iterations. The objective value decreases steadily afterwards. Both versions converge in under 300 iterations. This figure is generated with Matlab.

G. Proposed Three-Step Spatial-Structure Supervised Learning and Prediction Model (L21-SSSL)

Extended from the $L_{2,1}$-Regularized, $L_{2,2}$-Loss Support Vector Regression Model, the following L21-SSSL model is proposed, where $\mathbf{A} \in \mathbb{R}^{P\times P}$ is a matrix describing the linear relationship between each component of $\mathbf{Y}$:

Learning Step 1: Solve $L_{2,1}$-Regularized, $L_{2,2}$-Loss Support Vector Regression Model with either formulation (8) (L21-SSSL Version 1) or formulation (15) (L21-SSSL Version 2) to obtain the projection matrix $\mathbf{W}$.

Learning Step 2:

$$\begin{aligned} \min_{\mathbf{A}} \ & \alpha||\mathbf{YA}||_1 + ||\mathbf{A}||_1 \\ \text{s.t.} \ & \sum_{p=1}^{P} y_{np}a_{pk} = 0, \quad a_{kk} \neq 0, \\ & n \in \{1,\ldots,N\}, k \in \{1,\ldots,P\} \end{aligned} \tag{17}$$

Prediction Step:

$$\min_{\hat{\mathbf{Y}}} \quad ||\hat{\mathbf{Y}} - \mathbf{XW}||_1 + \alpha||\hat{\mathbf{YA}}||_1 \tag{18}$$

As suggested in [43], the $\mathbf{A}$ matrix in (17) learns the spatial arrangement of the centroid of the soft tissues, by formulating each coordinate of the soft tissues as the linear representation of the remaining response variables. By forcing one of the elements along each column (or each row) of the $\mathbf{A}$ matrix to be non-zero, a linear representation is enforced.

In other words, the objective function of formulation (18) considers both the prediction error and the error of the spatial arrangement in the three-dimensional feature space, thus avoiding too much deviation of the centroid prediction from the original centroid position. In [43], formulation (8) (or formulation (15)) and formulation (17) are combined. However, these two formulations can be separated.

For the prediction step (18), the L_1-norm is used to minimize the prediction error of the objective function.

Both formulations (17) and (18) can be solved efficiently by any convex programming package such as CVX [21].

III. Evaluations

Experiments were executed on a workstation with an Intel i7-5960X CPU and 64 GB RAM with Ubuntu 16.04 operating system. Programming code was developed in Matlab R2016b and can be found at the author's github[2].

A. Performance Measures

As presented in [5], evaluation measures for multi-response regression include average correlation coefficient, mean square error, average relative error, average root mean squared error (aRMSE) and average relative root mean squared error (aRRMSE). Among these measures, aRRMSE is considered as a more robust measure over others for multi-response prediction problems. It can be viewed as a normalized root mean squared error (RMSE) for each response variable as follows:

$$\text{aRRMSE} = \frac{1}{P}\sum_{p=1}^{P}\sqrt{\frac{\sum_{n=1}^{N_{test}}(y_{pn}-\hat{y}_{pn})^2}{\sum_{n=1}^{N_{test}}(y_{pn}-\bar{y}_p)^2}}, \tag{19}$$

where $\bar{y}_p$ is the average of responses for the p-th response variable, y_{pn} and $\hat{y}_{pn}$ are the actual and predicted responses of the p-th response variable and n-th observation respectively. During parameter tuning of the model, the performance using aRRMSE is considered so that learning will not be skewed towards variables of large error values and will not be affected by different scaling of the response variables. Finally, a leave-one-out cross validation was used for predicting the out-of-sample error. For each fold, one of the 20 observations is used as testing, and the remaining 19 observations are used for training.

B. Comparison with Baseline Methods

To illustrate the advantages of the proposed methods, four other multi-target prediction models are compared. All implementations are Matlab-based except for multi-target random forest, which is implemented via the scikit-learn package in

[2] https://github.com/pukkinming/IEEE_paper_l21_msvr

Python. Parameters in each model, if any, are aRRMSE-tuned via a two-level grid search to ensure the learning performance is fair. The two-level grid search [23] consists of two parts - a coarse grid is applied to find a good region first, and then a finer grid search is applied on the identified area. Instead of running a full grid search, which is time consuming and computationally expensive, a two-level grid search is robust and at the same time reduces computational time.

The following methods in the literature were used to compare with the proposed method. Notations in the following models have been changed for better consistency with the above.

- Spatial Structure Supervised Learning (SSSL): SSSL was the first of its kind to learn an optimal insertion position of tibia soft tissue of knees in a two-dimensional setting and is the basis of this work [43]. The $\mathbf{A}$ matrix, which stores the spatial structure, is learnt during the first step and is carried over to the prediction step. SSSL is an algorithm adaption method. Without the structural learning term, such method is reduced to the regular lasso model.
  Learning step:

$$\min_{\mathbf{A},\mathbf{W}} \quad \sum_{n=1}^{N} ||\mathbf{y}_n - \mathbf{W}^T\mathbf{x}_n||_2 + \alpha_1 \sum_{n=1}^{N} ||\mathbf{A}\mathbf{y}_n||_2 + \alpha_2 \sum_{i,j=1}^{P} |A_{ij}| + \alpha_3 \sum_{j=1}^{M}\sum_{p=1}^{P} |\mathrm{W}_{mp}|$$
$$\text{s.t.} \quad [\mathbf{y}_1, \ldots, \mathbf{y}_P]\,\mathbf{a}_p = \mathbf{0}, \;\; a_{pp} \neq 0, p = 1, \ldots, P \tag{20}$$

  Prediction step:

$$\min_{\mathbf{y}} \quad ||\mathbf{y} - \mathbf{W}^T\mathbf{x}||_1 + \alpha_1||\mathbf{A}\mathbf{y}||_1 \tag{21}$$

- Generalized linear model via penalized maximum likelihood (GLMNET, [16], [37]): GLMNET is an extension of the single-target case where it is designed to model correlated responses with a multi-response Gaussian distribution. The absolute penalty on each single coefficient is replaced with a group-lasso penalty.

$$\min_{(\mathbf{W}_0,\mathbf{W})\in\mathbb{R}^{(M+1)\times P}} \frac{1}{2N}\sum_{n=1}^{N} ||\mathbf{y}_n - \mathbf{W}_0^T - \mathbf{W}^T\mathbf{x}_n||_F^2 + \lambda\left[\frac{1}{2}(1-\alpha)||\mathbf{W}||_F^2 + \alpha\sum_{m=1}^{M}||\mathbf{w}_m||_2\right] \tag{22}$$

- Multi-target regressor stacking using decision tree (MTRS-DT, [39]): MTRS-DT is inspired from "stacked generalization" which was designed for multi-label classification. There are two stages in the learning process. Firstly, P single-target models are learned as usual. MTRS-DT then predicts the target one-by-one with stacking: after predicting the first response variable, its output will be attached to the original testing dataset to build a new meta model, which is in turned used to predict the second target variable. The process continues until all response variables are predicted in this manner [5], [39]. The stacking idea is powerful and can be applied to any single-target regression model. However, the sequence of stacking affects the regression performance. In this experiment, the prediction of the first response variable will be used in the prediction of the next.
- Multi-target random forest (MTRF, [7], [26], [30]): A random forest is a meta-learner that trains a number of decision trees, each of which uses a subset of the original dataset and outputs the final prediction by averaging the prediction from each decision tree. Generally speaking, MTRF differs from its single-target counterpart in that during the splitting of a tree node, the performance measure considers all the output variables instead of considering a single output variable at a time.

TABLE I
THE FOLLOWING TABLE SHOWS THE PREDICTION PERFORMANCE IN AVERAGE RELATIVE ROOT MEAN SQUARED ERROR (ARRMSE) OF THE CENTROIDS OF EIGHT INSERTION SITES USING LEFT-SIDE, RIGHT-SIDED AND BOTH KNEES USING LEAVE-ONE-OUT CROSS VALIDATION. A LOWER ARRMSE INDICATES A BETTER PREDICTION PERFORMANCE. BOLD ENTRIES CORRESPOND TO THE BEST PREDICTION FOR THE INSERTION AREA. GENERALLY, LEARNING PERFORMANCE IS ENHANCED IF THE KNEE DATA OF BOTH SIDES IS USED AND IF MULTIPLE INSERTION SITES ARE LEARNED SIMULTANEOUSLY.

	Left					
	SSSL	GLMNET	MTRS -DT	MTRF	**L21-SSSL (Ver. 1)**	**L21-SSSL (Ver. 2)**
ACL	1.0057	1.0206	1.1018	1.3590	**0.8977**	**0.8977**
ALMR	1.0687	1.0655	1.1018	1.4793	0.8973	**0.8905**
AMMR	0.8764	1.1000	1.0914	1.4421	0.8694	**0.8638**
LCART	1.0336	1.0247	1.1203	1.7627	**0.8353**	0.8391
MCART	1.2163	1.2023	1.0839	1.4627	1.0253	**1.0215**
PCL	1.4237	1.3492	**1.0921**	1.6554	1.2321	1.2344
PLMR	1.1319	**1.0168**	1.0947	1.5869	1.0324	1.0350
PMMR	1.6138	1.3762	**1.1382**	1.5326	1.3790	1.3821
All	1.1713	1.1444	1.1212	1.5351	1.0211	**1.0205**

	Right					
	SSSL	GLMNET	MTRS -DT	MTRF	**L21-SSSL (Ver. 1)**	**L21-SSSL (Ver. 2)**
ACL	1.3455	1.3259	**1.1067**	1.3336	1.1736	1.1679
ALMR	1.2636	1.2609	1.0909	1.1339	1.0748	**1.0676**
AMMR	1.0767	1.0926	1.1340	1.3009	0.9991	**0.9975**
LCART	1.1871	1.0845	1.1063	1.1748	0.9703	**0.9617**
MCART	0.9374	0.7977	1.1026	1.1124	**0.7364**	0.7434
PCL	1.3829	1.0841	**1.0734**	1.4416	1.1354	1.1303
PLMR	1.2519	1.1376	1.0955	1.1747	1.0515	**1.0474**
PMMR	1.2564	1.1213	1.0992	1.2166	**1.0283**	1.0343
All	1.2127	1.1131	1.0891	1.2361	1.0212	**1.0188**

	Both					
	SSSL	GLMNET	MTRS -DT	MTRF	**L21-SSSL (Ver. 1)**	**L21-SSSL (Ver. 2)**
ACL	1.0380	1.0055	1.0320	1.4772	0.9534	**0.9458**
ALMR	0.9673	0.8844	1.2293	1.4026	0.8707	**0.8574**
AMMR	0.7801	0.7255	1.2682	1.3257	0.7203	**0.7157**
LCART	0.7200	0.7087	1.1775	1.2702	**0.7086**	0.7090
MCART	0.6961	0.6839	1.2931	1.2407	0.6823	**0.6783**
PCL	1.1923	1.2238	1.4040	1.3068	**1.0862**	1.0918
PLMR	0.9900	0.9115	1.2621	1.4934	0.8961	**0.8868**
PMMR	0.9073	**0.9038**	1.3180	1.3834	0.9210	0.9341
All	0.9114	0.8809	1.2480	1.3625	0.8548	**0.8524**

C. Discussion of Computational Result

Table I summarizes the performance of predicting each of eight insertion centroids as well as predicting all eight locations at the same time with the proposed models and baseline models using data of the left knee only, the right knee only and both knees. For each soft tissue centroid, the average relative root mean squared error (aRRMSE) over the 20 subjects is reported.

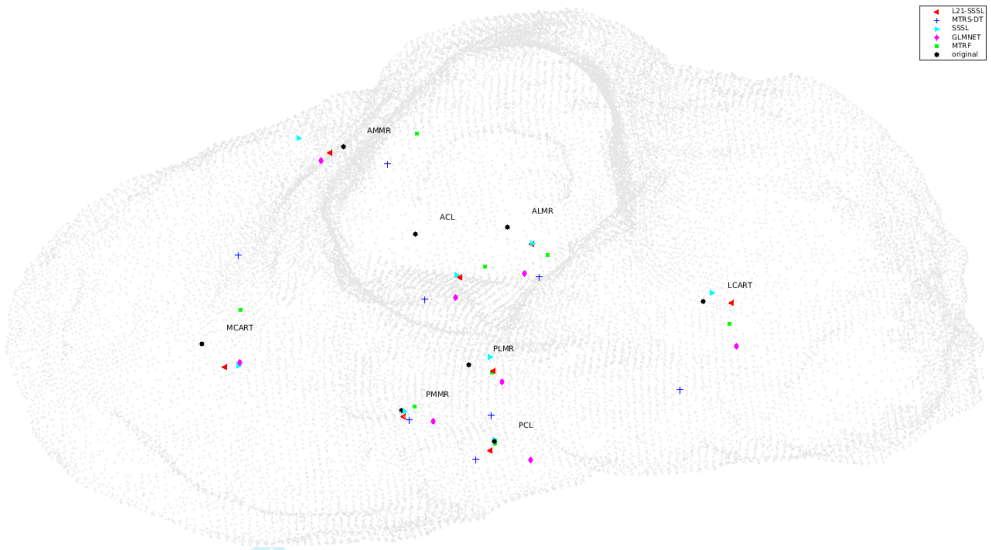

Fig. 6. Outline of tibia and the centroids of soft tissues of a subject viewed at $(\theta, \phi) = (0, 90)$. Original centroid position of a soft issue is highlighted as black. The predicted centroids are as follows: red triangle for L21-SSSL Version 2, blue cross for MTRS-DT, cyan triangle for SSSL, pink diamond for GLMNET and green square for MTRF. The L21-SSSL Version 2 predicted centroids are closer to the original centroids. Predictions from L21-SSSL Version 1 are not shown because their result is very similar to those of Version 2 as in Table I. Plots at other angles can be found at the author's github. This figure is generated with Matlab.

In general, the proposed L21-SSSL Version 2 performs better than other baseline models in terms of aRRMSE in five insertion areas (see ACL, ALMR, AMMR, MCART, and PLMR using data of both knees at Table I). The proposed model is also the best method for predicting all centroids simultaneously (0.8524 for L21-SSSL Version 2).

The advantage of the proposed model, as well as GLMNET and SSSL, is that the more responses available in the dataset, the better the learning performance, provided that responses are related to each other, such as having spatial relationship in this case. Because of the limited number of observations, a non-linear regression model, such as kernelized learning, may not be applicable because of possible over-fitting.

MTRS-DT and MTRF perform worse than the other three methods in this application. This is likely due to a lack of available observations. Decision trees and random forests are particularly suitable for datasets with many observations and high non-linearity in the decision boundary. Computational speed is also their strength.

Figure 6 shows the insertion centroids of the cartilages and soft tissues (in black) as well as the predicted positions using different algorithms - red triangle for L21-MSVR Version 2, blue cross for MTRS-DT, cyan triangle for SSSL, pink diamond for GLMNET and green square for MTRF. L21-MSVR is noticeably closer to the original position, followed by GLMNET and SSSL. Such experimental results confirm that the proposed supervised learning can improve prediction of the eight soft tissues simultaneously.

IV. Concluding Remarks

The motivation of this study is to develop a quantitative, automated framework in order to localize the optimal position of soft tissue insertion using patient-specific features 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, imaging alignment using GPA, patient-specific morphological feature extraction from tibia outlines and multi-response support vector regression. The proposed methodology yielded the best prediction performance as compared with other baseline models for eight soft tissue structure locations.

A few directions are recommended for future exploration. A non-linear learning model might be considered, in addition to increasing the number of observations and other innovative ideas of feature extraction. Predicting the shape and size of the insertion sites, and predicting the performing angle of the surgical tools can be a promising direction to continue this work. This quantitative analysis framework could be the basis of a clinical application to assist surgeons to better identify soft tissue insertion sites. Such an application, preferably on a computer-aided surgery platform, would render a close replication of the native anatomy so that the risk of non-anatomic tunnel placement or iatrogenic injury to adjacent tissue structures can be minimized.

V. Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

References

[1] Luca Baldassarre, Lorenzo Rosasco, Annalisa Barla, and Alessandro Verri. Multi-output learning via spectral filtering. *Machine learning*, 87(3):259–301, 2012.

[2] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[3] David J Biau, Caroline Tournoux, Sandrine Katsahian, Peter Schranz, and Rémy Nizard. Acl reconstruction: a meta-analysis of functional scores. *Clinical orthopaedics and related research*, 458:180–187, 2007.

[4] Turner A Blackburn and Emily Craig. Knee anatomy: a brief review. *Physical therapy*, 60(12):1556–1560, 1980.

[5] Hanen Borchani, Gherardo Varando, Concha Bielza, and Pedro Larrañaga. A survey on multi-output regression. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(5):216–233, 2015.

[6] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[7] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[8] Sébastien Bubeck et al. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

[9] Feng Cai and Vladimir Cherkassky. Svm+ regression and multi-task learning. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pages 418–424. IEEE, 2009.

[10] Xiao Cai, Feiping Nie, Heng Huang, and Chris Ding. Multi-class l2, 1-norm support vector machine. In *Data Mining (ICDM), 2011 IEEE 11th International Conference on*, pages 91–100. IEEE, 2011.

[11] Wikimedia Commons. File:knee diagram.svg — wikimedia commons, the free media repository, 2016. [Online; accessed 31-July-2017].

[12] Glenn De'Ath. Multivariate regression trees: a new technique for modeling species–environment relationships. *Ecology*, 83(4):1105–1117, 2002.

[13] Andrew Edwards, Anthony MJ Bull, and Andrew A Amis. The attachments of the anteromedial and posterolateral fibre bundles of the anterior cruciate ligament. *Knee Surgery, Sports Traumatology, Arthroscopy*, 16(1):29–36, 2008.

[14] Mario Ferretti, Daniel Doca, Sheila M Ingham, Moises Cohen, and Freddie H Fu. Bony and soft tissue landmarks of the acl tibial insertion site: an anatomical study. *Knee Surgery, Sports Traumatology, Arthroscopy*, 20(1):62–68, 2012.

[15] Donald C Fithian, Elizabeth W Paxton, Mary Lou Stone, William F Luetzow, Rick P Csintalan, Daniel Phelan, and Dale M Daniel. Prospective trial of a treatment algorithm for the management of the anterior cruciate ligament-injured knee. *The American journal of sports medicine*, 33(3):335–346, 2005.

[16] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

[17] Hemanth R Gadikota, Jae Ang Sim, Ali Hosseini, Thomas J Gill, and Guoan Li. The relationship between femoral tunnels created by the transtibial, anteromedial portal, and outside-in techniques and the anterior cruciate ligament footprint. *The American journal of sports medicine*, 40(4):882–888, 2012.

[18] Marc T Galloway, Edward S Grood, John N Mehalik, Martin Levy, Stephen C Saddler, and Frank R Noyes. Posterior cruciate ligament reconstruction: an in vitro study of femoral and tibial graft placement. *The American journal of sports medicine*, 24(4):437–445, 1996.

[19] William E Garrett Jr, Marc F Swiontkowski, James N Weinstein, John Callaghan, Randy N Rosier, Daniel J Berry, John Harrast, G Paul Derosa, et al. American board of orthopaedic surgery practice of the orthopaedic surgeon: part-ii, certification examination case mix. *JBJS*, 88(3):660–667, 2006.

[20] Thomas J Gill, Samuel K Van de Velde, David W Wing, Luke S Oh, Ali Hosseini, and Guoan Li. Tibiofemoral and patellofemoral kinematics after reconstruction of an isolated posterior cruciate ligament injury: in vivo analysis during lunge. *The American journal of sports medicine*, 37(12):2377–2385, 2009.

[21] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, March 2014.

[22] Christopher D Harner, Goo Hyun Baek, Tracy M Vogrin, Gregory J Carlin, Shinji Kashiwaguchi, and Savio LY Woo. Quantitative analysis of human cruciate ligament insertions. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, 15(7):741–749, 1999.

[23] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification. 2003.

[24] Takanori Iriuchishima, Goro Tajima, Sheila JM Ingham, Kenji Shirakura, and Freddie H Fu. Pcl to graft impingement pressure after anatomical or non-anatomical single-bundle acl reconstruction. *Knee Surgery, Sports Traumatology, Arthroscopy*, 20(5):964–969, 2012.

[25] Alan Julian Izenman. Reduced-rank regression for the multivariate linear model. *Journal of multivariate analysis*, 5(2):248–264, 1975.

[26] H Tim Kam. Random decision forest. In *Proc. of the 3rd Int'l Conf. on Document Analysis and Recognition, Montreal, Canada, August*, pages 14–18, 1995.

[27] Kang Li, Madelyn O'Farrell, Daniel Martin, Sebastian Kopf, Christopher Harner, and Xudong Zhang. Mapping ligament insertion sites onto bone surfaces in knee by co-registration of ct and digitization data. *Journal of biomechanics*, 42(15):2624–2626, 2009.

[28] James H Lubowitz. Anteromedial portal technique for the anterior cruciate ligament femoral socket: pitfalls and solutions. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, 25(1):95–101, 2009.

[29] Feiping Nie, Heng Huang, Xiao Cai, and Chris H Ding. Efficient and robust feature selection via joint l2, 1-norms minimization. In *Advances in neural information processing systems*, pages 1813–1821, 2010.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[31] K. B. Petersen and M. S. Pedersen. The matrix cookbook, nov 2012. Version 20121115.

[32] 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 & Traumatology: Surgery & Research*, 97(6):S75–S79, 2011.

[33] Matilde Sánchez-Fernández, Mario de Prado-Cumplido, Jerónimo Arenas-García, and Fernando Pérez-Cruz. Svm multiregression for nonlinear channel estimation in multiple-input multiple-output systems. *IEEE transactions on signal processing*, 52(8):2298–2307, 2004.

[34] Shai Shalev-Shwartz, Yoram Singer, Nathan Srebro, and Andrew Cotter. Pegasos: Primal estimated sub-gradient solver for svm. *Mathematical programming*, 127(1):3–30, 2011.

[35] Rainer Siebold, Thomas Ellert, Stefan Metz, and Juergen 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: The Journal of Arthroscopic & Related Surgery*, 24(2):154–161, 2008.

[36] Richard Simmons, Stephen M Howell, and ML 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. *JBJS*, 85(6):1018–1029, 2003.

[37] Noah Simon, Jerome Friedman, and Trevor Hastie. A blockwise descent algorithm for group-penalized multiresponse and multinomial regression. *arXiv preprint arXiv:1311.6529*, 2013.

[38] Eleftherios Spyromitros-Xioufis, Grigorios Tsoumakas, William Groves, and Ioannis Vlahavas. Multi-label classification methods for multi-target regression. *arXiv preprint arXiv*, 1211, 2012.

[39] Eleftherios Spyromitros-Xioufis, Grigorios Tsoumakas, William Groves, and Ioannis Vlahavas. Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, 104(1):55–98, 2016.

[40] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[41] Goro Tajima, Masahiro Nozaki, Takanori Iriuchishima, Sheila JM Ingham, Wei Shen, Patrick Smolinski, and Freddie H Fu. Morphology of the tibial insertion of the posterior cruciate ligament. *JBJS*, 91(4):859–866, 2009.

[42] HOWARD B Tandeter and PESACH Shvartzman. Acute knee injuries: use of decision rules for selective radiograph ordering. *American family physician*, 60(9):2599–2608, 1999.

[43] Cao Xiao, Shouyi Wang, Liying Zheng, Xudong Zhang, and Wanpracha Art Chaovalitwongse. A patient-specific model for predicting tibia soft tissue insertions from bony outlines using a spatial structure supervised learning framework. *IEEE Transactions on Human-Machine Systems*, 46(5):638–646, 2016.

[44] Thore Zantop, Nadine Diermann, Tobias Schumacher, Steffen Schanz, Freddie H Fu, and Wolf Petersen. Anatomical and nonanatomical double-bundle anterior cruciate ligament reconstruction: importance of femoral tunnel location on knee kinematics. *The American journal of sports medicine*, 36(4):678–685, 2008.

[45] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, 26(8):1819–1837, 2014.

[46] Wei Zhang, Xianhui Liu, Yi Ding, and Deming Shi. Multi-output ls-svr machine in extended feature space. In *Computational Intelligence for Measurement Systems and Applications (CIMSA), 2012 IEEE International Conference on*, pages 130–134. IEEE, 2012.

[47] Liying Zheng, Christopher D Harner, and Xudong Zhang. The morphometry of soft tissue insertions on the tibial plateau: data acquisition and statistical shape analysis. *PloS one*, 9(5):e96515, 2014.

Kin Ming (Frank) Puk received the Bachelor's degree in Information Engineering from the Chinese University of Hong Kong and the Master's degree in Operations Research from the University of Southern California. He is currently a Ph.D. student at the University of Texas at Arlington. His research interests include medical imaging, data mining and optimization.

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.

Jay Rosenberger has a B.S. in Mathematics from Harvey Mudd College, an M.S. in Industrial Engineering from the University of California at Berkeley, and a Ph.D. in Industrial Engineering from the Georgia Institute of Technology.

His research interests include mathematical optimization and simulation in transportation, defense, and health care. He is currently a Professor with the Department of Industrial, Manufacturing, and Systems Engineering, the Director of the Center on Stochastic Modeling, Optimization, & Statistics (COSMOS), and an associate editor of Omega: The International Journal of Management Science.

Cao (Danica) Xiao (S'14) is a Research Staff Member in the AI for Healthcare Team of IBM Research. Before joining IBM, she got her Ph.D. degree from University of Washington, Seattle in 2016. Her research interests focus on developing novel machine learning and data mining models to solve real world healthcare challenges. Particularly, she is interested in deep computational phenotyping and risk prediction, deep representation learning for graphs, causal inference from observational data, tensor decomposition for integrating multiple medical data sources, well as translational informatics research.

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 did her postdoctoral training in department of orthopaedic surgery at University of Pittsburgh, Pittsburgh, PA, USA.

She is currently a biomechanical research engineer with Health Effects Lab Division at the National Institute for Occupational Safety and Health (NIOSH), Morgantown, WV, USA. Her research interests include the areas of biomechanics, musculoskeletal modeling and simulation, morphometrics, and advanced data analysis for improving the prevention, diagnosis, and treatment of musculoskeletal disorders.

Xudong Zhang received his B.S. degree in mechanical engineering from Tsinghua University, Beijing, China, and 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 in the Department of Industrial and Systems Engineering, with joint appointments in the Departments of Mechanical Engineering and Biomedical Engineering, at Texas A&M University, College Station, Texas. Previously, he had been on faculty of University of Pittsburgh and University of Illinois at Urbana-Champaign. His research centers on modeling, analysis, and simulation of the human musculoskeletal system and its multi-scale structures and components during functional, particularly dynamic activities, and identifies a wide range of applications in healthcare and human-machine systems engineering.

Wanpracha Art Chaovalitwongse (M'05–SM'11) is currently the 21st Century Leadership Chair in Engineering, a Professor of Industrial Engineering, and the Co-Director with the Institute of Advanced Data Analytics, University of Arkansas, Fayetteville. He previously worked as a Professor with the Departments of Industrial and Systems Engineering, Radiology (Joint), and Bioengineering (Adjunct), University of Washington (UW), Seattle. He also served as an Associate Director with 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.