Intelligent Automation And Soft Computing, 2019
Copyright © 2019, TSI® Press
Vol.25,no.1,105–115

Surgical Outcome Prediction in Total Knee Arthroplasty using Machine Learning

Belayat Hossain[a], Takatoshi Morooka[b], Makiko Okuno[b], Manabu Nii[a], Shinichi Yoshiya[b] and Syoji Kobashi[a]

[a]Graduate School of Engineering, University of Hyogo, Hyogo, Japan
[b]Department of Orthopaedics Surgery, Hyogo College of Medicine, Hyogo, Japan

ABSTRACT
This work aimed to predict postoperative knee functions of a new patient prior to total knee arthroplasty (TKA) surgery using machine learning, because such prediction is essential for surgical planning and for patients to better understand the TKA outcome. However, the main difficulty is to determine the relationships among individual varieties of preoperative and postoperative knee kinematics. The problem was solved by constructing predictive models from the knee kinematics data of 35 osteoarthritis patients, operated by posterior stabilized implant, based on generalized linear regression (GLR) analysis. Two prediction methods (without and with principal component analysis followed by GLR) along with their sub-classes were proposed, and they were finally evaluated by a leave-one-out cross-validation procedure. The best method can predict the postoperative outcome of a new patient with a Pearson's correlation coefficient (cc) of 0.84±0.15 (mean±SD) and a root-mean-squared-error (RMSE) of 3.27±1.42 mm for anterior-posterior vs. flexion/extension (*A-P* pattern), and a cc of 0.89±0.15 and RMSE of 4.25±1.92° for valgus-varus vs. flexion/extension (*i-e* pattern). Although these were validated for one type of prosthesis, they could be applicable to other implants, because the definition of knee kinematics, measured by a navigation system, is appropriate for other implants.

KEY WORDS: Total knee arthroplasty, knee implant, knee kinematics, machine learning, generalized linear regression, prediction.

1 INTRODUCTION

TOTAL knee arthroplasty (TKA) is a common and cost-effective surgical treatment for knee osteoarthritis (OA), rheumatoid arthritis (RA), and other conditions, which replaces a damaged knee joint by an artificial one (Murray et al., 2014). The fabricated knee joint, also known as a TKA prosthesis, consists mainly of the femoral component, the tibial component, and an insert. There are various types of knee prostheses, of which three types, cruciate retaining, posterior stabilized, and cruciate substituting, are well-known to surgeons. Because of the anatomical structure and functional variability of the knee joint from patient to patient, there should be an appropriate TKA implant for each individual for patient-specific treatment. Currently, the surgeon has to select a TKA operation method and an implant product model without predicting the outcome quantitatively. Nonetheless, the outcome of TKA depends strongly on the TKA prosthesis, surgical technique, expertise of the clinician, and postoperative physiotherapy (Berend et al., 2013; Victor et al., 2010). Thus, the prediction of postoperative knee function before surgery is essential for the best surgical planning and patients' satisfaction (Choi et al., 2016) as well.

A non-invasive method has recently become more commonly used for knee kinematic analysis than an invasive method (Hiroshi et al., 2012; Tei et al., 2012) because of patient safety. Using the first type, one can investigate implanted knee kinematics (i.e., postoperative knee functions) using 2-D/3-D image registration technique with 2-D X-ray digital radiograph movies and 3-D computer-aided design (Yamazaki et al., 2004; Kobashi et al., 2005) and 3-D magnetic position sensors (Tomaru et al., 2010) only after the surgery. Some studies have investigated the

CONTACT Belayat Hossain belayat@ieee.org

relationship between preoperative and postoperative knee kinematics after the surgery to evaluate surgical performance and/or other clinical issues (Seon et al., 2011; Onsem et al., 2016; Hasega et al., 2015). Although these methods are essential for such applications, they are inappropriate for predicting the postoperative knee kinematics of a new patient prior to surgery, which has been done in this work. Additionally, a predictive model has recently attracted substantial attention among researchers for predicting surgical outcomes (Reinbolt et al., 2009; Sridevi et al., 2017, Miao et al., 2017, Galarranga et al., 2017).

The present work proposes predictive models for forecasting postoperative implanted knee kinematics by using regression analysis to calculate a mapping function from preoperative to postoperative kinematics. The principal contribution of this work is to predict the postoperative knee functions of a new patient before surgery by measuring the preoperative knee functions for possible application in patient-specific knee surgical planning. Two prediction methods (without and with principal component analysis (PCA), followed by generalized linear regression, GLR, analysis) along with their sub-classes are proposed, and they were evaluated by a leave-one-out cross-validation (LOOCV) procedure. Furthermore, the methods were also optimized for best prediction precision with respect to different predictive variables. Another objective was to show the possibility of predicting postoperative kinematics, as well as investigating the effectiveness of PCA for dimension reduction and GLR analysis for predictive model construction. A part of this article has appeared previously in Hossain, et al. (2016).

2 SUBJECTS AND MATERIALS

2.1 Subjects

THIS study analyzed knee joint functions of 35 OA patients (11 males, 24 females; mean±standard deviation (SD) age 74.08±6.90, range 57-83 years). The TKA operations of all subjects were carried out individually by two experienced surgeons, using posterior stabilized (PS) type knee implants (Vega, Aesculap, B/Braun, Melsungen, Germany). This study was approved by the local Ethics Committee, and each subject provided informed consent.

2.2 Data acquisition

Basically, knee joint functions are described by three rotation angles and three translations. These angles, namely *flexion-extension* (*f-e*), *valgus-varus* (*v-v*), and *internal-external* (*i-e*), as well as the translations called *medial-lateral* (*M-L*), *anterior-posterior* (*A-P*), and *superior-inferior* (*S-I*) are defined along the X, Y and Z axes, respectively, following the Grood and Suntay coordinate system (Grood & Suntay, 1983), shown in Figure. 1.

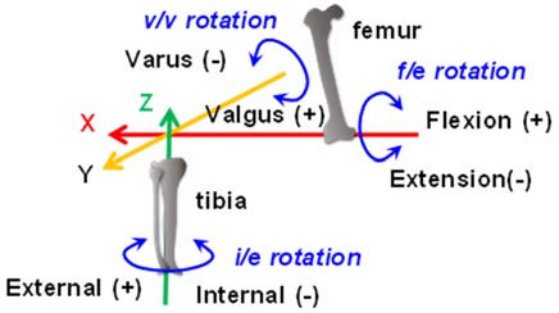


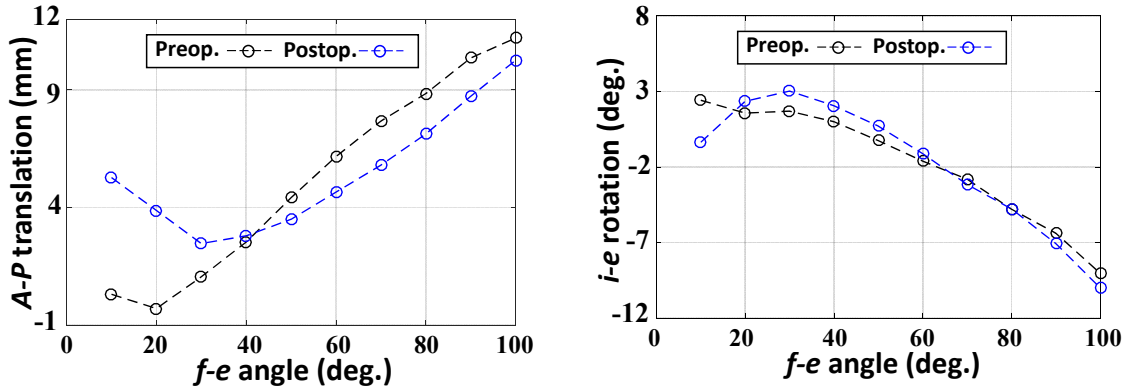
Figure 1. Knee joint function definition.

The *A-P* translations and the *i-e* rotations were measured for every 10° *f-e* angle (i.e., 10°, 20°, ...100°) by passively flexing the knee joint between flexion angles of 10° and 100° with the patient in a supine position under a non-load-bearing condition, hereafter known as the *A-P* and *i-e* patterns, respectively. The measurements were taken both before and after the TKA surgery of every patient in the operating room of Hyogo College of Medicine (Hyogo, Japan) between May 14, 2014 and November 27, 2015 using a CT-free navigation system (OrthoPilot, B/Braun, Aesculap, Melsungen, Germany) for implant positioning and limb alignment. The before and after measurements were confirmed as the preoperative and postoperative knee functions, respectively. Figure 2 shows examples of measured preoperative (dash dotted black color) and postoperative (dash dotted blue color) knee kinematics (both *A-P* and *i-e* patterns) of a patient. There were 35 pairs of samples for each type of pattern (total 70 pairs of patterns) in the dataset. Each pair comprised one preoperative and one postoperative kinematic data of a patient; one example is shown in Figure 2. Kinematic patterns of each sample of all subjects were investigated carefully, with non-linearity and discrete (*A-P*/*i-e* values available for every 10º f/e angle only) features found among preoperative and postoperative data in all pairs. It was also observed that the kinematic pattern varied from subject to subject, i.e. there was no direct correlation among them that could be represented by any generic algebraic equation. Therefore, to grasp the inter-individual variation, a statistical technique (i.e. PCA) combined with a machine learning method (i.e. GLR) was used to predict surgical outcomes of TKA patients, where, during training the models, the preoperative and postoperative data were assigned as predictive (input data) and response (output data) variables, respectively (details explained in Section 4).

3 PROPOSED METHODS

3.1 Data pre-processing

CONSIDER a pre-operative kinematic pattern of training data of size, $N_s \times N_m$, as shown in (1),



(a) *A-P* pattern (b) *i-e* pattern

Figure 2. Examples of knee kinematics data of a test patient before (dash dotted black color) and after (dash dotted blue color) the surgery.

$$\boldsymbol{f}^{pr} = \begin{pmatrix} f_1^{pr}(1) & f_1^{pr}(2) & \cdots & f_1^{pr}(N_m) \\ f_2^{pr}(1) & f_2^{pr}(2) & \ldots & f_2^{pr}(N_m) \\ \vdots & \vdots & \ddots & \vdots \\ f_{N_s}^{pr}(1) & f_{N_s}^{pr}(2) & \cdots & f_{N_s}^{pr}(N_m) \end{pmatrix} \quad (1)$$

where N_s is the number of subjects, N_m is the number of measurement points, and $f_i^{pr}(j)$ denotes reading (*i.e.*, *A-P* translation, *i-e* rotation value) of subject i at measurement point, j. Thus, an entire row corresponds to a preoperative kinematic pattern of one subject. Similarly, postoperative kinematic training data of the same subjects are also organized as shown in (2).

$$\boldsymbol{f}^{po} = \begin{pmatrix} f_1^{po}(1) & f_1^{po}(2) & \cdots & f_1^{po}(N_m) \\ f_2^{po}(1) & f_2^{po}(2) & \ldots & f_2^{po}(N_m) \\ \vdots & \vdots & \ddots & \vdots \\ f_{N_s}^{po}(1) & f_{N_s}^{po}(2) & \cdots & f_{N_s}^{po}(N_m) \end{pmatrix} \quad (2)$$

3.2 Model construction

Firstly, preoperative and postoperative kinematics training data (*Eq.* 1 & *Eq.* 2) were projected independently into a lower dimensional space using PCA (scaling is recommending if the range of values in the variables is larger) (Jollifee, 2002). In the following context, let principal components (PCs), principal axes (PAs), and mean kinematic pattern be $\boldsymbol{\lambda}^{pr}$ of size $N_s \times N_m$, $\boldsymbol{V}^{pr}$ of size $N_m \times N_m$, and $\boldsymbol{\mu}^{pr}$ of size $N_m \times 1$, respectively, for preoperative training data, and let $\boldsymbol{\lambda}^{po}$, $\boldsymbol{V}^{po}$, and $\boldsymbol{\mu}^{po}$ be the same parameters and size defined above but corresponding to the postoperative training data.

Let us propose a prediction method for dim_{pr} ($< N_m$) PCs at preoperative kinematics, and dim_{po} PCs at postoperative kinematics, but retaining at least a 95% cumulative contribution ratio (CCR), *i.e.*, dim_{po} predictive models are constructed from PCs of first dim_{pr} *PCs*. Thus, the first predictive model (say, glm_1) based on a general linear model (GLM) (McCullagh & Nelder, 1989) is constructed between the lower dimensional PCs of first dim_{pr} PAs of pre- and first PC of post-operative kinematic training data. The regression model is represented by (3),

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} \quad (3)$$

where predictor variables,

$$\boldsymbol{X} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,dim_{pr}} \\ x_{2,1} & x_{2,2} & \ldots & x_{2,dim_{pr}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N_s,1} & x_{N_s,2} & \cdots & x_{N_s,dim_{pr}} \end{pmatrix} \quad (4)$$

are constructed by first dim_{pr} PCs taken from $\boldsymbol{\lambda}^{pr}$. $x_{i,j}$ is the j^{th} PC of subject i; here $i = 1, 2, \ldots N_s$, $j = 1, 2, \ldots dim_{pr}$. $\boldsymbol{Y}$ is represented by (5), obtained from the 1st PC taken from $\boldsymbol{\lambda}^{po}$; where $i = 1,2, \ldots N_s$.

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_s} \end{pmatrix} \quad (5)$$

$\boldsymbol{\beta} = \left(\beta_1, \beta_2, \cdots, \beta_{dim_{pr}}\right)^T$ is a vector of regression coefficients and must be estimated from the training data. Statistically significant predictor variables could be tested during training the model by Akaike's Information Criterion (AIC) (Akaike, 1974) to optimize the model.

So far, the first regression model (glm_1) was constructed by relating first dim_{pr} preoperative PCs as predictive variables with the first postoperative PCs as response variables. Similarly, dim_{po} regression models ($glm_2, \ldots glm_{dim_{po}}$) were constructed by considering subsequent 2nd, 3rd, .., $\text{dim}_{\text{po}}^{\text{th}}$ postoperative PCs as response variables without any change in predictor variables. Note that different preoperative PC dimensions, dim_{pr}=1, 2, ..., N_m-1, with postoperative PC, dim_{po}= 1st or 2nd or …, N_m^{th}-1, retaining at least 95% CCR were tested to optimize prediction precision

and finally chosen to train the model for kinematics prediction of a new patient.

3.3 Kinematic prediction

A new patient's same type preoperative kinematic pattern (f_{new}^{pr}) is projected to a lower-dimension of size, dim_{pr} by PCA as shown in (6),

$$\boldsymbol{\lambda}_{new}^{pr}(j) = \sum_{1}^{i=N_m} \left((\boldsymbol{f}_{new}^{pr} - \boldsymbol{\mu}^{pr}).\boldsymbol{V}^{pr}(i,j) \right) \quad (6)$$

where j = 1, 2, …, $\dim_{pr}$.

Then, postoperative PCs of the new patient are estimated using $\boldsymbol{\lambda}_{new}^{pr}$ from corresponding trained models, (glm_1, glm_2,… $glm_{dim_{po}}$), and are combined as shown in (7),

$$\boldsymbol{\lambda}_{new}^{po}(k) = \left(\lambda_{new}^{po}(1), \lambda_{new}^{po}(2), \cdots \lambda_{new}^{po}(dim_{po})\right) \quad (7)$$

Finally, the postoperative kinematic pattern of the patient is predicted using estimated $\boldsymbol{\lambda}_{new}^{po}$ as shown in (8).

$$\boldsymbol{f}_{new}^{pred}(i) = \sum_{k=1}^{dim_{po}} (\boldsymbol{\lambda}_{new}^{po}(k).\boldsymbol{V}^{po}(i,k)) + \boldsymbol{\mu}^{po} \quad (8)$$

where $i = 1, 2, \ldots, N_m$. Thus, the maximum dim_{pr} dimensions at preoperative and dim_{po} dimensions at postoperative are used for predicting the kinematic pattern.

4 EXPERIMENTAL METHODS AND EVALUATION

4.1 Training data organization

THIS study used two types of knee functions, *A-P* translation *vs. f-e* angle *(A-P* pattern*)* and *i-e* rotation *vs. f-e* angle *(i-e* pattern*)* from a dataset of 35 pairs of samples for each type of pattern (total 70 samples) to train the models to assess the proposed methods, and since the knee functions are supposed to be associated with each other, the training data matrices were organized according to the variables used for training the models, as follows,

(Individual mapping, IM) Predicting the *A-P or i-e* pattern by training the predictive models with preoperative and postoperative data of either *A-P* or *i-e* type only (9).

$$\begin{aligned} \boldsymbol{f}_{ap/ie}^{pr} &= \boldsymbol{f}^{pr}(i,j) \\ \boldsymbol{f}_{ap/ie}^{po} &= \boldsymbol{f}^{po}(i,j) \end{aligned} \quad (9)$$

where subject's number, *i=1, 2, … 35* and measurement points, *j= 1, 2, … 10*. $\boldsymbol{f}_{ap/ie}^{pr}$ and $\boldsymbol{f}_{ap/ie}^{po}$ were defined as the preoperative (predictor variables) and postoperative (response variables) *A-P* (or *i-e*) kinematic training data matrices, respectively.

(Combined to individual mapping, CIM) Predicting the *A-P* or *i-e* pattern by gathering data of both patterns (*A-P* and *i-e*) of preoperative kinematic data as predictor variables and corresponding postoperative *A-P or i-e* kinematic data as response variables (10).

$$\begin{aligned} \boldsymbol{h}_{ai}^{pr} &= [\boldsymbol{f}_{ap}^{pr} \ \boldsymbol{f}_{ie}^{pr}] \\ \boldsymbol{h}_{ap/ie}^{po} &= [\boldsymbol{f}_{ap/ie}^{po}] \end{aligned} \quad (10)$$

where $\boldsymbol{h}_{ai}^{pr}$ is preoperative training data (predictive variables) by gathering both *A-P* and *i-e* patterns, of size N_s-by-2×N_m (35×20); $\boldsymbol{h}_{ai}^{pr}$ and $\boldsymbol{h}_{ap/ie}^{po}$ are predictor and response variables of preoperative and postoperative data for predicting *A-P* or *i-e* patterns, respectively.

(Combined mapping, CM) Predicting *A-P* and *i-e* patterns by gathering preoperative and postoperative data of both patterns (*A-P* and *i-e*) as predictor variables and response variables, respectively (11).

$$\begin{aligned} \boldsymbol{h}_{ai}^{pr} &= [\boldsymbol{f}_{ap}^{pr} \ \ \boldsymbol{f}_{ie}^{pr}] \\ \boldsymbol{h}_{ai}^{po} &= [\boldsymbol{f}_{ap}^{po} \ \ \boldsymbol{f}_{ie}^{po} \ ] \end{aligned} \quad (11)$$

where the parameters were defined the same as before, with a difference; in this case, response variables ($\boldsymbol{h}_{ai}^{po}$) were defined by combining postoperative data of both patterns, $\boldsymbol{f}_{ap}^{po}$ and $\boldsymbol{f}_{ie}^{po}$.

4.2 Performance assessment

For each kinematic pattern, the prediction performance of the methods was tested by the RMSE (12) and the cc between the original (f_i^{test}) and the predicted kinematics pattern (f_i^{pred}) to observe prediction error and their linear relationship, respectively. Here, N_{pts} is the total number of data points both in the original and predicted patterns.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_{pts}} \left(f_i^{test} - f_i^{pred}\right)^2}{N_{pts}}} \quad (12)$$

The proposed method was evaluated by a cross-validation test, where part of the data is usually used for training the model, and the rest of the data is used for prediction. In this work, the LOOCV procedure (Duda et al., 2000; Miller, 1974) was used in our datasets of 35 patterns from each type: when a patient *s* was tested, both preoperative and postoperative kinematics data of patient *s* were removed from the training set. Then, generalized linear regression was performed with data of the other remaining patients, and, subsequently, the kinematics data of patient *s* were tested on the trained model. This process was repeated for all patients in the database. Finally, to compare the methods, the mean RMSE and mean cc were calculated for each type of kinematic pattern.

5 EXPERIMENTAL RESULTS

THE methods were implemented in the open source software library R Ver. 3.2.2 (R core team, 2016). Different PCA dimensions (1 to 10 and/or 20) were tested corresponding to the best prediction precision (high mean correlation coefficient (cc) and low mean root mean squared error (RMSE)) for

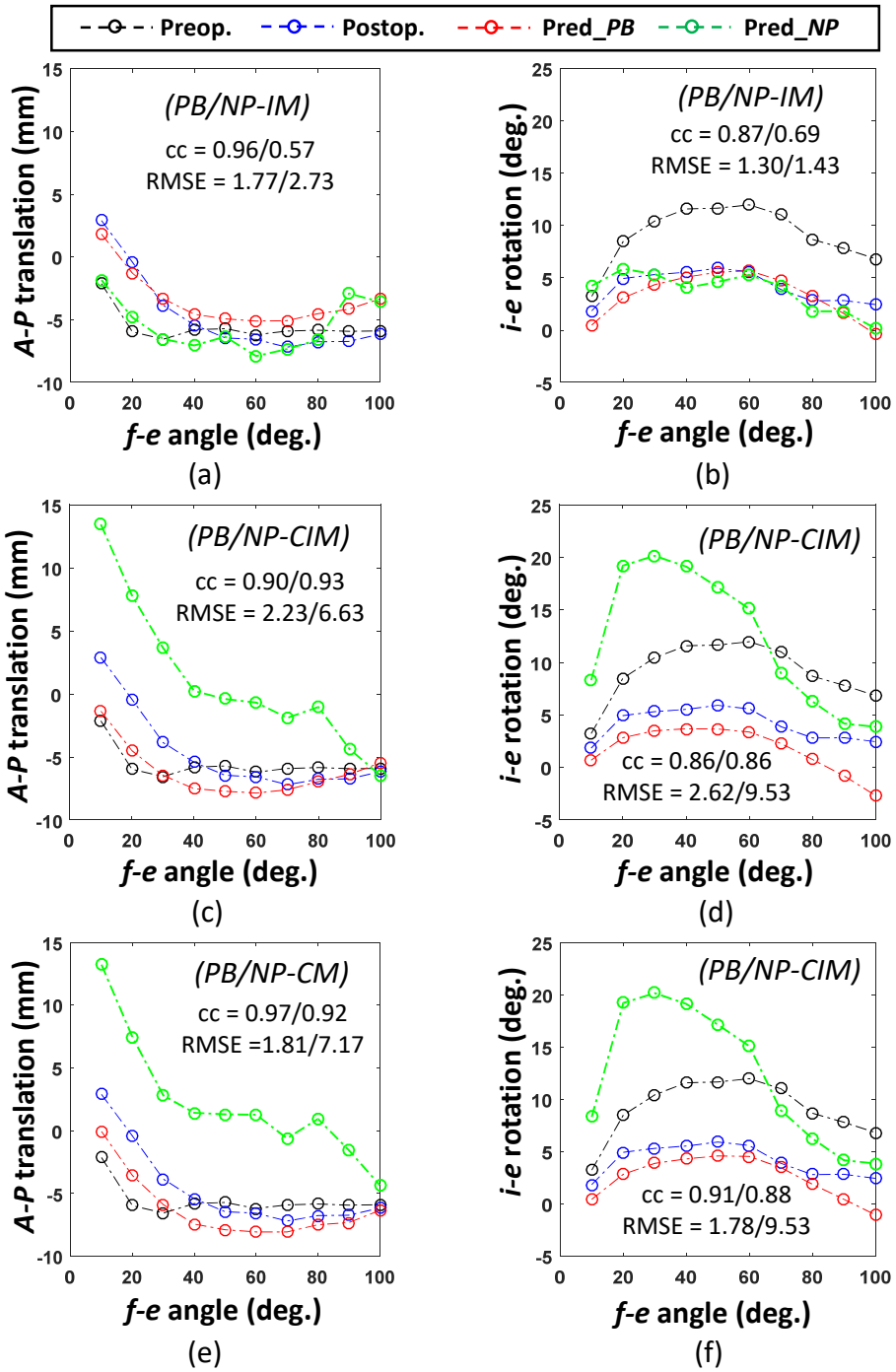


Figure 3. **Output of the methods for a test patient obtained by the proposed methods, PB and NP. Dash-dotted black, blue, red (PB), and green (NP) lines correspond to the preoperative, postoperative, and predicted knee functions, respectively. In the first column, the first to last rows (a, c, & e) represent A-P patterns obtained from (PB/NP IM), (PB/NP CIM), and (PB/NP CM) methods, respectively. In the second column, the first to last rows (b, d & f) represent *i-e* patterns obtained from (PB/NP IM), (PB/NP CIM), and (PB/NP CM) methods, respectively. In-set prediction accuracy is shown, i.e. refer to (a), cc of PB-IM and NP-IM is 0.96 and 0.57, respectively.**

regression analysis, retaining more than 95% CCR. Basically, in the case of kinematic data acquisition with a navigation system, the range of data values in the *A-P* or *i-e* pattern was not larger. Therefore, PCA was applied without any scaling, although the units of the *A-P* and *i-e* patterns were different. Thus, the percentage of total variance (also known as CCR) was explained by the first five PCA dimensions for both cases (10 or 20 variables). Finally, the models were trained with a PCA dimension of 5 for preoperative data and 3 for postoperative data, both of which contained at least 98% of the total variance in all methods. This method, PCA followed by generalized linear regression, was identified as the PCA-based (PB) method, and sub-categorized as *(PB-IM)*, *(PB-CIM)*, and *(PB-CM)*, according to Eq. (9), (10), and (11), respectively, to optimize prediction precision with high cc and low RMSE among these methods of both *A-P* and *i-e* patterns, since they are supposed to be associated with each other.

Table 1a. Performance comparison of the proposed methods (mean±SD)- *A-P* translation prediction.

Methods	cc		RMSE [mm]	
	With PCA (*PB*)	Without PCA (*NP*)	With PCA (*PB*)	Without PCA (*NP*)
IM	0.84±0.15	0.71±0.26	3.55±1.83	3.99±1.91
CIM	0.81±0.16	0.54±0.39	3.27±1.42	5.59±2.95
CM	0.80±0.20	0.56±0.34	3.30±1.38	5.19±2.82

Table 1b. Performance comparison of the proposed methods (mean±SD)- *i-e* rotation prediction.

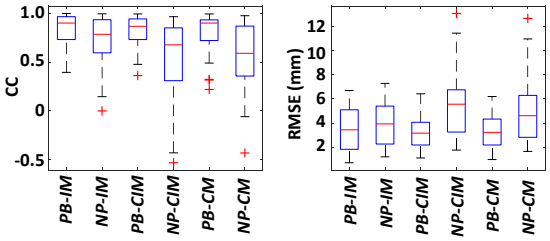
Methods	cc		RMSE [deg.]	
	With PCA (*PB*)	Without PCA (*NP*)	With PCA (*PB*)	Without PCA (*NP*)
IM	0.88±0.12	0.86±0.17	4.43±1.98	4.55±2.32
CIM	0.88±0.17	0.69±0.39	4.25±1.92	5.59±2.68
CM	0.89±0.15	0.69±0.39	4.25±1.95	5.99±2.68

Furthermore, regression models were also created without PCA pre-processing, which predicted *A-P* and *i-e* patterns directly by constructing GLMs (10 models for (9) & (10); 20 models for (11)) whose variables, used for prediction, were either of *A-P* or *i-e*, or both as described in equations (9), (10), and (11). The method without PCA was identified as the non-PCA (*NP*)-based method and sub-classified as *(NP-IM)*, *(NP-CIM)*, and *(NP-CM)*, respectively.

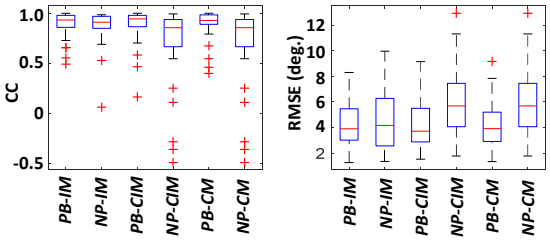
Finally, the outcome of the method (*NP*) was compared to the PCA-based method (*PB*) with corresponding sub-classes. The predicted kinematic curves of a test patient obtained by the proposed methods are summarized in Figure 3, which depicts differences in predicted kinematics in terms of both correlation and prediction error between the methods, *(PB)* and *(NP)*, which were also observed for other test patients.

Table 1 summarizes the results of all subjects for the different methods. Statistically significant differences between two methods (*PB* and *NP*) for both *A-P* (Table 1a) and *i-e* patterns (Table 1b) ($p < 0.05$) were evaluated by the paired-samples *t*-test, and in every case, the null hypothesis (H_0) was rejected. Method *PB* outperformed *NP* in terms of cc and RMSE for both patterns. However, method (*IM*) provided similar performance in terms of cc between methods (*PB* and *NP*) for *i-e* rotation prediction only, whereas the performance of methods (*CIM*) and (*CM*) was very different performance compared to method (*IM*).

For further quantification in addition to mean cc and mean RMSE, Figure 4 shows the distribution of the performance of the methods, which also shows the superiority of the PCA-based method (*PB*) for kinematics prediction over the non-PCA method (*NP*). Higher mean cc and lower mean RMSE were observed for all subgroups of the *PB* method compared to those of the *NP* method. Both cc and RMSE values of most subjects were more closely distributed around the respective mean values for all *PB* methods than that for respective *NP* methods. However, both methods have some outliers, the possible reasons for which are explained in the Discussion section. These findings are valid in both types of pattern.

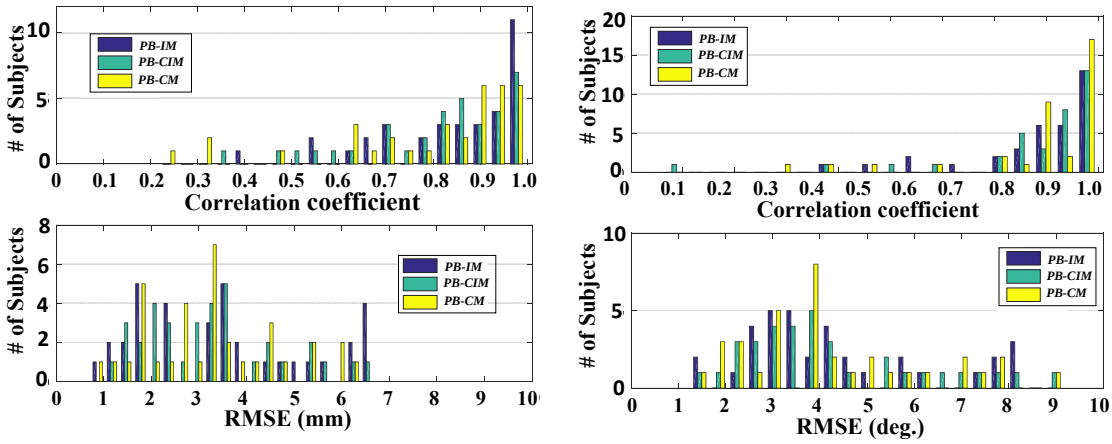


(a) *A-P* pattern



(b) *i-e* pattern

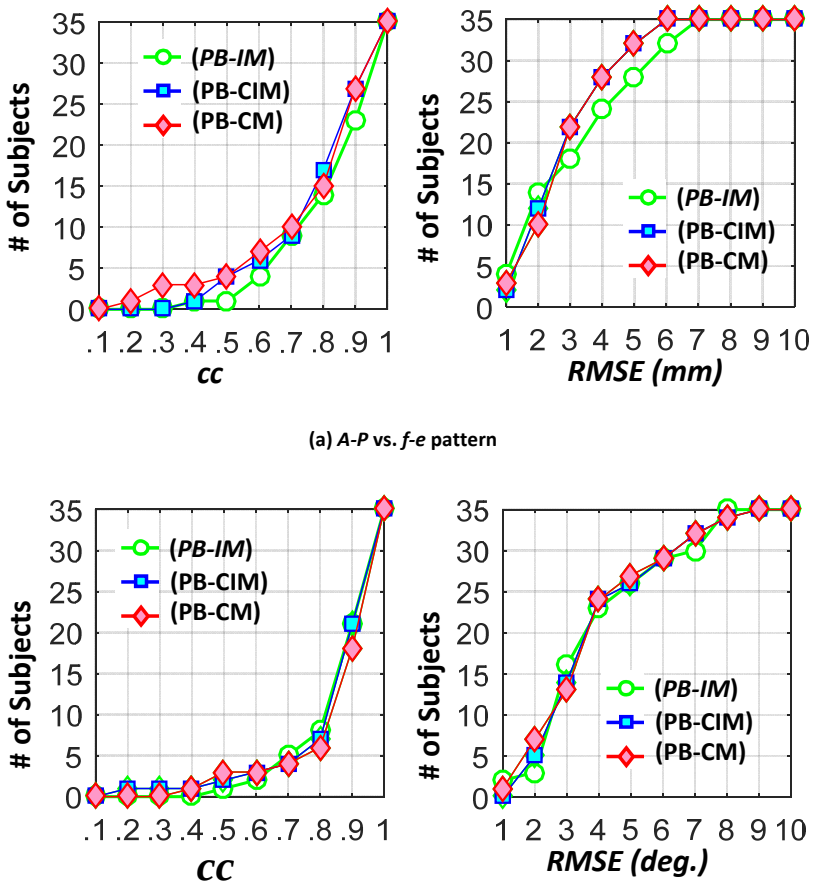
Figure 4. Boxplots show the performance of the methods. They show the distributions of cc and RMSE of the predicted patterns obtained from the methods and comparisons among them.



(a) A-P translation prediction

(b) *i-e* rotation prediction

Figure 5. Histograms of cc and RMSE obtained from the (*PB*) methods for both the *A-P* (a) and *i-e* (b) patterns. It shows the number of subjects that corresponds to cc and RMSE.



(a) *A-P* vs. *f-e* pattern

(b) *i-e* vs. *f-e* pattern

Figure 6. Cumulative histogram of cc and RMSE derived from different methods. Green-circle, blue-square, and red-diamond lines correspond to methods (*PB-IM*), (*PB-CIM*), and (*PB-CM*), respectively.

Since the *A-P* and *i-e* patterns are seemed to be associated with each other, to evaluate prediction precision with high cc and low RMSE within (*PB*), histograms for cc and RMSE of these methods (*PB-IM*, *PB-CIM, & PB-CM*) were plotted by counting the number of subjects for a specific value of cc and RMSE, respectively (Figure 5). In such a plot, ideally, all subject cc values are expected to be close to 1, which means a single bar at the cc value of 1. Pragmatically, the outcomes depicted in the figure (location of the maximum bars at the right side of the plot) show that a relatively large number of subjects has higher cc values are accumulated around the value of 1. In contrast, theoretically, all subject RMSE values are expected to be zero. In the present findings (location of the maximum bars at the left side of the plot), although none of the subjects has zero RMSE, most of them are accumulated close to zero RMSE. Thus, it can be concluded that the predicted pattern obtained by the proposed methods has a good correlation for pattern shape, but it is not free of error.

Since there was no statistically significant difference ($p<0.05$) within (*PB*) methods, the area ratio under the cumulative histogram curve (AUCHC, mathematically known as the definite integral) was also calculated to test the performance within (*PB*) methods. First, the cumulative histogram of each method was plotted, and then the area under the histogram curve of the corresponding method was calculated, following the principle of the definite integral. Then, the AUCHC was defined as the ratio of the area under the curve to the whole area of the plot. It takes a value between 0 and 1; a smaller AUCHC value of *cc* and a larger AUCHC value of RMSE indicate better performance, because this means fewer worse cases. The horizontal range was set to [0.1, 1.0] for cc and [1, 10] for RMSE to include all cases. Figure 6 visualizes the cumulative histogram curves of the three subgroups of *PB* methods for both patterns. The method corresponds to the smallest area in the cc plot and the greatest area in the RMSE plot under the curve to the horizontal axis has high prediction precision. Thus, for comparison, the AUCHC derived from the cumulative histograms of cc and RMSE (Table 2) suggests that *PB-CM* and *PB-CIM* have the best performance with respect to cc and RMSE, respectively, for predicting the *A-P* pattern, whereas *PB-CM* and *PB-IM* are the best methods with respect to cc and RMSE, respectively, for predicting the *i-e* pattern.

Table 2. Comparison of the area ratio under the cumulative histogram curve (AUCHC) among the proposed methods.

Methods	*A-P* translation		*i-e* rotation	
	cc	RMSE [mm]	cc	RMSE [deg.]
PB-IM	0.199	0.687	0.233	0.721
PB-CIM	0.199	0.717	0.156	0.616
PB-CM	0.156	0.616	0.150	0.626

6 DISCUSSION

THE present approach is the first to enable us to predict the postoperative knee functions of a new TKA patient by measuring the preoperative knee functions. Other methods are usually effective for postoperative knee kinematic analysis to evaluate the surgical outcome after the surgery (Seon et al., 2011 & Hasegawa et al., 2015) and/or postoperative complications (Onsem et al., 2016). Basically, the proposed methods derived a mapping function from the clinical training data by using generalized regression analysis.

In the case of method (*NP),* since every 10° *f-e* angle was used as a predictive variable (10 for *Eq*. 9 and 10; 20 for *Eq*. 12) for both preoperative and postoperative kinematics to train the predictive models, the prediction accuracy decreased due to irregularity in pattern shape, and thus it was incompatible with clinical application. One possible reason for the reduced precision of this method could be construction of individual predictive models for every 10° *f-e* angle between the preoperative and postoperative kinematics, and hence data redundancy affected prediction performance.

In contrast, method (*PB)* overcame this limitation by projecting the training data into a lower dimensional space by PCA, retaining at least 95% CCR (Jolliffe, 2002), and after that, training the predictive models by the PCs. The role of PCA was essentially to optimize features. Hence, higher prediction precision was achieved with dimensionality reduction than without it. Thus, the outcome of this work suggests PCA-based predictive methods for forecasting postoperative knee kinematics.

Additionally, for PCA-based intra-methods comparison, with respect to cc, the best method was *PB-IM* for *A-P* pattern prediction (0.84±0.15), but it provided a higher RMSE value than other methods, and *PB-CM* was the best method for *i-e* pattern prediction (0.89±0.15). However, the AUCHC area ratio comparison showed that *PB-CM* is the best for both patterns (0.156 (*A-P*) and 0.150 (*i-e*)). Thus, *PB-CM* is expected to give the best prediction and fewer inconsistent cases. This suggests that we should evaluate both preoperative *A-P* and *i-e* patterns to achieve better prediction, and it also means that there is a relationship between the *A-P* and *i-e* patterns.

With respect to RMSE, there were also no significant difference within any combination of the PCA-based method ($p<0.05$). However, the AUCHC area ratio comparison showed that the best method is *PM-CIM* for *A-P* translation (0.717) and *PB-IM* for *i-e* rotation (0.721). Thus, *PB-CIM* is the best *A-P* pattern method to give the best prediction and fewer inconsistent cases, and *PB-IM* is the best *i-e* pattern prediction method. Basically, the *PB-CM* method produced the best prediction accuracy in terms of the cc value, as described in the previous paragraph, which should also be true in terms of the RMSE value.

In other words, the *PB-CM* method was also expected to produce the best prediction outcome in terms of the RMSE value. However, this discrepancy could be due to calibration error of the navigation system, as described in the following context.

In comparing cc and RMSE, the results suggested that RMSE performance was relatively lower than that of cc performance, and there were some outliers, as shown in Figure 4. In our measurement procedure, preoperative and postoperative kinematics were measured using a CT-free navigation system in an operating room. It calibrated the patient's knee anatomical coordinate system used for measuring kinematics before the surgery, and it also used the same coordinate system through preoperative and postoperative measurements because of the system's limitation and to measure the kinematic change in the same coordinate system. However, in some clinical cases (referring to outliers in Figure 4), the surgeon implanted TKA intentionally with some rotation angle to revise the patient's knee joint alignment, and it could affect the zero-position and orientation of the coordinate system. This could be one reason for deterioration of the method's performance, specifically prediction error. Thus, prediction precision was degraded for some cases because the proposed predictive methods were trained by kinematic data that could include calibration error of the navigation system. However, it could be improved by calibrating the knee anatomical coordinate system after implantation or by measuring the kinematics a few months after the TKA operation by using, for example, 2-D/3-D image registration technique with a 2-D X-ray digital radiograph movie and 3-D computer-aided design (Yamazaki et al., 2004 & Kobashi et al., 2005). Actually, it is very effective for predict the outcome of TKA a few months after the operation because there would be a slight change with recovery.

Although the proposed methods were validated to one type of prosthesis, they could be applicable to other implants as well, because the definition of knee kinematics, measured by the same navigation system, is appropriate for other implants. Furthermore, the proposed method could also be applicable to the load-bearing condition, because Yoshiya, et. al. (2005) showed that the basic kinematics patterns were similar in both conditions (non-load bearing and load-bearing). In addition, it could also be applicable for suggesting the appropriate TKA implant in patient-specific knee surgical treatment before the surgery for accurate TKA planning, if it is trained by knee function data operated with other implants and by surgery code. This result shows its possible application in predicting postoperative knee function by measuring preoperative function, which could help the surgeon investigate the postoperative outcome of a patient before surgery. The outcome of the system could also serve as motivation for patients to proceed with TKA.

As a limitation, for some cases, prediction accuracy could be less because the operation outcome depends on the surgeon's skill and could deviate slightly for other surgeons (Victor et al., 2010), which was not considered during model implementation.

7 CONCLUSIONS

THIS study introduced PCA-based generalized linear regression analysis and optimized predictive variables during training the models to predict the most likely surgery outcome, specifically for TKA, before the surgery, by measuring only the preoperative knee functions. The methods were validated for two types of knee functions, and acceptable performance was achieved. The methods can predict the postoperative outcome of a new patient with a mean *cc* of at most 0.84±0.15 (mean ±SD) (*PB-IM*) and mean RMSE of at least 3.27±1.42 mm (*PB-CIM*) for the *A-P* pattern, and a mean *cc* of at most 0.89±0.15 (*PB-CM*) and a mean RMSE of at least 4.25±1.92 deg. (*PB-CIM*) for the *i-e* pattern. The best optimized prediction accuracy was obtained by method *PB-CM* for both patterns, although it provides a slightly higher RMSE value than others, which could be due to navigation calibration error. Thus, this study suggests method *PB-CM* for postoperative kinematics prediction in the future, considering the relationship between both the *A-P* and *i-e* patterns.

Although this work dealt with two types of patterns of knee functions, and one type of prosthesis, it could be applicable to other patterns, i.e. *v/v* rotation, because the data type of kinematics is the same as for other rotations, i.e. *i-e* rotation, and prostheses as well, since the knee kinematic definition using the same navigation system is similar for other prostheses. Finally, this work represents encouraging progress towards predicting postoperative kinematics before surgery using preoperative kinematics of TKA patients, in order to help clinicians to choose the optimal treatment and to help patients better understand the operation outcome. By incorporating more training data and/or investigating preoperative knee function, the model's performance could be enhanced in the future.

8 ACKNOWLEDGMENTS

A MEXT Scholarship was awarded to BH. The authors would also like to thank the anonymous reviewers for their insightful comments that supported us in preparing the final manuscript.

9 DISCLOSURE STATEMENT

NO potential conflict of interest was reported by the authors.

10 REFERENCES

H. Akaike. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19(6)*, 716–723.

K. R. Berend, A. V. Lombardi Jr, & J. B. Adams. (2013). Which total knee replacement implant should I pick? Correcting the pathology: the role of knee bearing designs. *Bone & Joint Journal, 95-B(11)*, 119-132.

Y. J. Choi & H. J. Ra. (2016). Patient satisfaction after total knee arthroplasty. *Knee surgery Related. Research, 28(1),* 1-15.

R. O. Duda, P. E. Hart, & D. G. Stork. (2000). *Pattern Classification* (2nd ed.). 22-24, New York, NY: Wiley-Blackwell.

O. A. Galarranga, C. V. Vigneron, B. Dorizzi, N. Khouri, & E. Desailly. (2017). Predicting postoperative gait in cerebral palsy. *Gait & posture, 52,* 45-51.

E. S. Grood & W. J. Suntay. (1983). A joint coordinate system for the clinical description of three-dimensional motions: application to the knee. *Journal of Biomechanical Engineering, 105(2),* 136-144.

M. Hasegawa, H. Takagita, & A. Sudo. (2015). Prediction of postoperative range of motion using intraoperative soft tissue balance in total knee arthroplasty with navigation. *Computer Aided Surgery, 20(1),* 47-51.

H. Hiroshi, A. Shaw, T. Tetsuya, K. Sugamoto, T. Yamazaki, & N. Shimizu. (2012). In vivo kinematic analysis of cruciate-retaining total knee arthroplasty during weight-bearing and non-weight-bearing deep knee bending. *The Journal of Arthroplasty, 27(6),* 1196-1202.

B. M. Hossain, M. Nii, T. Morooka, M. Okuno, S. Yoshiya, & S. Kobashi. (2016). Post-operative implanted knee kinematics prediction in total knee arthroscopy using clinical big data. In *Lecture notes in Computer Science, Springer, 9835(2),* 405-412.

I. T. Jolliffe. (2002). *Principal Component Analysis*, 10-27, New York, NY: Springer-Verlag.

S. Kobashi, S. T. Tomosada, N. Shibanuma, M. Yamaguchi, M. Muratsu, K. Kondo, S. Yoshiya, and M. Kurosaka. (2005). Fuzzy image matching for pose recognition of occluded knee implants using fluoroscopy images. *Journal of Advanced Computational Intelligence and Intelligent Informatics, 9(2),* 181-195.

Y. Z. Miao, X. P. Ma, & S. P. Bu. (2017). Research on the Learning Method Based on PCA-ELM. *Intelligent Automation & Soft Computing. 23(4),* 637-642.

P. McCullagh & J. A. Nelder. (1989). *Generalized Linear Models* (2nd ed.). 25-32, London, Chapman and Hall.

R. G. Miller, (1974). The jackknife-a review. *Biometrika, 61(1),* 1-15.

D. W. Murray, G. S. MacLennan, S. Breeman, H. A. Dakin, L. Johnston, M. K. Campbell, and A. M. Gray, KAT group. (2014). A randomized controlled trial of the clinical effectiveness and cost-effectiveness of different knee prostheses: the knee arthroplasty trial (KAT). *Health Technology 352 Assessment., 18(19),* 1-235, vii-viii.

S. V. Onsem, V. D. Straeten, N. Arnout, P. Deprez, G. Damme, & J. Victoe. (2016). A new prediction model for patient satisfaction after Total Knee Arthroplasty. *Journal of Arthroplasty, 31(12),* 2660-2667.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

J.A. Reinbolt, M. D. Fox, M. H. Schwartz, & S. L. Delp. (2009). Predicting outcomes of rectus femoris transfer surgery. *Gait & posture, 30(1),* 100-105.

J. K. Seon, J. K. Park, M. S. Jeong, W. B. Jung, K. S. Park, T. R. Yoon, & E. K. Song. (2011). Correlation between preoperative and postoperative knee kinematics in total knee arthroplasty using cruciate retaining designs. *International Orthopaedics, 35,* 515-520.

M. Sridevi, P. Prakasam, S. Kumaravel, & P. Madhavsarma. (2017). Tibia Fracture Healing Prediction Using Adaptive Neuro Fuzzy Inference System. *Intelligent Automation & Soft Computing, 23(2),* 359-363.

K. Tei, N. Shibanuma, S. Kubo, T. Matsumoto, A. Matsumoto, H. Tateishi, M. Kurosaka, and R. Kuroda. (2012). Kinematic analysis of mobile-bearing total knee arthroplasty using image matching technique. *Journal of Bone & Joint Surgery Br, 94,* 242.

A. Tomaru, S. Kobashi, Y. Tsumori, S. Yoshiya, K. Kuramoto, S. Imawaki, & Y. Hata. (2010, October). A 3-DOF knee joint angle measurement system with inertial and magnetic sensors. *Proceedings of 2010 IEEE International conference on system, man and cybernetics,* (pp. 1261-1266), Istanbul, Turkey, IEEE System, Man & Cybernetics Society.

J. Victor, J. K. Mueller, R. D. Komistek, A. Sharma, M.C. Nadaud, & J. Bellemans. (2010). In vivo kinematics after a cruciate-substituting TKA. *Clinical Orthopaedics Related Research, 468(3),* 807–814.

T. Yamazaki, T. Watanabe, Y. Nakajima K. Sugamoto, T. Tomita, H. Yoshikawa, & S. Tamura. (2004). Improvement of depth position in 2-D/3-D Registration of knee implants using single-plane fluoroscopy. *IEEE Transactions Medical Imaging, 23(5),* 602-612.

S. Yoshiya, N. Matsui, R.D. Komistek, D. A. Dennis, M. Mahfouz, & M. Kurosaka. (2005). In vivo kinematic comparison of posterior cruciate-

retaining and posterior stabilized total knee arthroplasties under passive and weight-bearing conditions. *Journal of Arthroplasty, 20(6),* 777-83.

NOTES ON CONTRIBUTORS

B. Hossain is a doctoral student at the graduate school of engineering, University of Hyogo. He has also been serving as a research assistant at the Advanced medical engineering center, Himeji since 2016. His research interests include Biomedical signal analysis and Artificial intelligent in medical engineering. Currently, he is a research scholar of MEXT fellowship and student member of IEEE.

T. Morooka graduated Hyogo College of Medicine and received M.D. in 2005. He is currently an assistant professor of Department of Orthopaedic Surgery, Hyogo College of Medicine since 2015.

M. Okuno is an Orthopaedic Surgeon in Hyogo college of Medicine. She was received PhD degree from Hyogo college of Medicine in 2015. She is currently an associate professor at the department of Orthopaedic Surgery Hyogo College of Medicine. She is specialized in total knee arthroplasty. Her current study is intraoperative analysis of flexion kinematics of osteoarthritis knee.

M. Nii received Ph.D. from Osaka Prefecture University in 2001. Presently, he is working as Assistant Professor, Graduate School of Engineering, University of Hyogo. His research interests are Computational Intelligence, Machine Learning, Soft Computing, and Intelligent Systems for Nursing. He is members of the Institute of Systems, Control and Information Engineers, the Institute of Electronics, Information and Communication Engineers, Japan Society for Fuzzy Theory and Intelligent Informatics, the Virtual Reality Society of Japan, and the IEEE.

S. Yoshiya graduated Kobe University School of Medicine and received M. D. in 1979. He went through orthopaedic residency of Kobe University from 1979 to1984, and special fellowship in orthopaedic research in the Cleveland Clinic Foundation from 1984 to 1986. He has been serving as Chairman of Department of Orthopaedic Surgery, Hyogo College of Medicine since 2005.

S. Kobashi received Dr. of Engineering from Himeji Institute of Technology. He is currently a professor and director of AMEC at University of Hyogo since 2017. His research interests include artificial intelligence in medical engineering. He is the editor at large of Autosoft journal, and a senior member of IEEE.