Deep Learning-Based Full-Process Automatic CPAK Classification System and Its Application in the Analysis of Alignment Outcomes Before and After Knee Arthroplasty

Kun Wu , Xiao Geng , Xinguang Wang , Jiazheng Chen , Hua Tian *

Posted Date: 26 March 2026

doi: 10.20944/preprints202603.2038.v1

Keywords: total knee arthroplasty; CPAK classification; deep learning; alignment transition; individualized alignment; HRNet-W32

 doi:10.20944/preprints202603.2038.v1

Article

Deep Learning-Based Full-Process Automatic CPAK Classification System and Its Application in the Analysis of Alignment Outcomes Before and After Knee Arthroplasty

Kun Wu [1], Xiao Geng [1], Xinguang Wang [1], Jiazheng Chen [2] and Hua Tian [1,*]

[1] Orthopaedic Department, Peking University Third Hospital, No. 49 North Garden Rd, Haidian District, 100191, Beijing, China

[2] Department of Orthopaedic, The First Affiliated Hospital, Zhejiang University School of Medicine, No. 79 Qingchun Road, Shangcheng District, 310003, Hangzhou, Zhejiang, China

* Correspondence: tianhua@bjmu.edu.cn

Abstract

Background: Coronal Plane Alignment of the Knee (CPAK) classification enables individualized alignment assessment in total knee arthroplasty (TKA), yet manual evaluation is time-consuming and lacks preoperative-to-postoperative transition analysis. **Methods:** This study aims to develop and validate a fully automated deep learning-based CPAK classification system and to investigate individual-level transition patterns and their association with short-term clinical outcomes using paired radiographic data from a large Chinese cohort. A retrospective analysis was conducted on 919 patients with knee osteoarthritis who underwent TKA. A keypoint detection model (HRNet-W32) was developed to automatically calculate medial proximal tibial angle, lateral distal femoral angle, arithmetic hip-knee-ankle angle, and joint line obliquity, from which CPAK types were derived. **Results:** On the validation set (92 cases), the model achieved a mean radius error of 1.22±0.43 mm for keypoint detection; mean absolute errors were ≤0.74° for angular measurements, with intraclass correlation coefficients ≥0.96 compared to manual annotations. Automatic CPAK classification accuracy was 80.98% (kappa = 0.767). Transition matrix analysis showed that only 8.30% of preoperative type I patients maintained their original type postoperatively, with most shifting to types IV, V, or VII. After propensity score matching, no significant differences in clinical outcomes were observed among transition groups (all $p > 0.05$). **Conclusions:** These results demonstrate that the proposed automated system enables efficient CPAK assessment, revealing substantial postoperative alignment transitions that were not associated with differential short-term outcomes, thereby supporting AI-assisted individualized alignment planning in TKA.

Keywords: total knee arthroplasty; CPAK classification; deep learning; alignment transition; individualized alignment; HRNet-W32

1. Introduction

Knee osteoarthritis (KOA) is a leading cause of functional impairment in the middle-aged and elderly population. Total knee arthroplasty (TKA), as an effective treatment for end-stage KOA, can significantly relieve pain and improve function[1,2]. Postoperative outcomes of TKA are closely related to the reconstruction of lower limb alignment, which directly affects prosthesis stress distribution, joint stability, and long-term patient prognosis[3,4]. Conventional TKA aims to restore mechanical alignment (MA) as its core objective. Although its 10-year implant survival rate can exceed 90%, approximately 20% of patients remain dissatisfied with postoperative outcomes[5–8].

 doi:10.20944/preprints202603.2038.v1

This is primarily because MA does not fully account for individual anatomical variations, potentially leading to soft tissue imbalance or abnormal joint line obliquity [9,10]. Consequently, alternative individualized strategies have been proposed, including anatomic alignment (AA), kinematic alignment (KA), restricted kinematic alignment (rKA), and functional alignment (FA)[11–14]. However, AA has limited applicability for complex deformities; KA may result in alignment deviations >3° with insufficient long-term evidence; the safe range for rKA remains undefined; and FA relies heavily on surgeon experience and navigation equipment, limiting its widespread adoption[15,16]

The core of individualized alignment strategies lies in identifying distinct anatomical phenotypes among patients[17,18]. In 2021, MacDessi et al. introduced the Coronal Plane Alignment of the Knee (CPAK) classification system. This system utilizes the medial proximal tibial angle (MPTA) and lateral distal femoral angle (LDFA) to calculate the arithmetic hip-knee-ankle angle (aHKA) and joint line orientation (JLO), classifying the knee into nine phenotypes and providing a systematic framework for personalized surgical planning [19–21]. A deeper understanding of the anatomical characteristics of CPAK types and their association with postoperative outcomes is crucial for optimizing surgical strategies [22,23].

However, the clinical implementation of this classification faces bottlenecks. Currently, CPAK classification relies on manual measurements, which are time-consuming and subjective, thus limiting its widespread clinical application[24]. In recent years, deep learning has made significant strides in medical image analysis, particularly excelling in automated skeletal keypoint detection and angle measurement [25,26]. Studies have shown that fully automated alignment measurement systems based on convolutional neural networks achieve excellent agreement with manual annotations, with intraclass correlation coefficients (ICC) > 0.90, laying the technical groundwork for automating CPAK classification [27].

Current research on the Chinese population has reported the distribution characteristics of CPAK types[28]; however, the individual-level transition pathways from preoperative to postoperative CPAK types remain unclear [9–11]. Key bottlenecks include: the acquisition of CPAK types still relies on manual effort, hindering large-scale application; and the individual-level transition patterns from preoperative to postoperative CPAK types have not been clarified. To address these gaps, this study aims to develop a fully automated deep learning-based CPAK classification system and, using paired pre- and postoperative data from a large cohort of Chinese TKA patients, systematically analyze the individual-level transition patterns of CPAK types and their association with clinical outcomes. The specific objectives are: 1. To develop and validate a fully automated CPAK classification model; 2. To construct a transition matrix to elucidate transition patterns; 3. To analyze the impact of different transition patterns on clinical outcomes.

2. Materials and Methods

This study was approved by the Medical Science Research Ethics Committee of Peking University Third Hospital (IRB00006761-M2023429) and was registered in the Chinese Clinical Trial Registry (NCT06010979).

2.1. Study Design and Population

A total of 1000 patients with knee osteoarthritis who underwent TKA at Peking University Third Hospital between January and December 2023 were initially identified. All patients had pre- and postoperative full-length weight-bearing anteroposterior lower limb radiographs. Patients were randomly divided into a model development set (n = 827, 1654 radiographs) and a model validation set (n = 92, 184 radiographs) at the patient level. Inclusion criteria were: Kellgren-Lawrence grade ≥ 3; underwent TKA; and had standard pre- and postoperative full-length weight-bearing lower limb radiographs. Exclusion criteria were: images with flexion or rotational deformity (rotation > 10°); severe bone defects on the femoral or tibial side; and poor image quality. A total of 919 patients were ultimately included for analysis. Patient demographic information, including age, sex, height,

 doi:10.20944/preprints202603.2038.v1

weight, body mass index (BMI), surgical technique, and preoperative Kellgren-Lawrence grade of the affected side, was collected from the medical records.

2.2. Clinical Outcome Assessment

Postoperative evaluation was conducted through outpatient follow-up or telephone interviews, with a follow-up period of 12–24 months. Assessment tools included the Knee Society Score (KSS), the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), the Forgotten Joint Score (FJS), and a 5-level Likert satisfaction scale (for statistical analysis, "satisfied" and "very satisfied" were combined into the "satisfied" group, while the remaining responses were combined into the "non-satisfied" group). All clinical outcome data were collected by independent researchers who were not involved in the surgical procedures or image annotation.

2.3. Image Acquisition, Manual Annotation, and Gold Standard

Image Acquisition: Standard full-length weight-bearing lower limb radiography was performed.

Keypoint Definitions[19]: Femoral head center, femoral intercondylar notch center, distal medial femoral condyle point, distal lateral femoral condyle point, tibial intercondylar eminence center, medial tibial plateau lowest point, lateral tibial plateau lowest point, talar dome center (ankle joint center).

Annotation Process: One orthopedic resident annotated the keypoints using Labelme software (as shown in Figure 1), and two senior chief physicians reviewed and corrected the annotations.

Gold Standard: Three senior chief physicians, blinded to the AI results, independently performed manual measurements (MPTA, LDFA, aHKA, JLO) and CPAK classification using professional imaging software. The final classification was based on the consensus of at least two experts. The specific angle values measured by the experts were recorded. Kendall's coefficient of concordance (W) was used to assess the consistency of the angle measurements among the three physicians, with $W > 0.8$ indicating extremely high consistency. The mean values were used as the reference for evaluating the accuracy of the AI measurements.

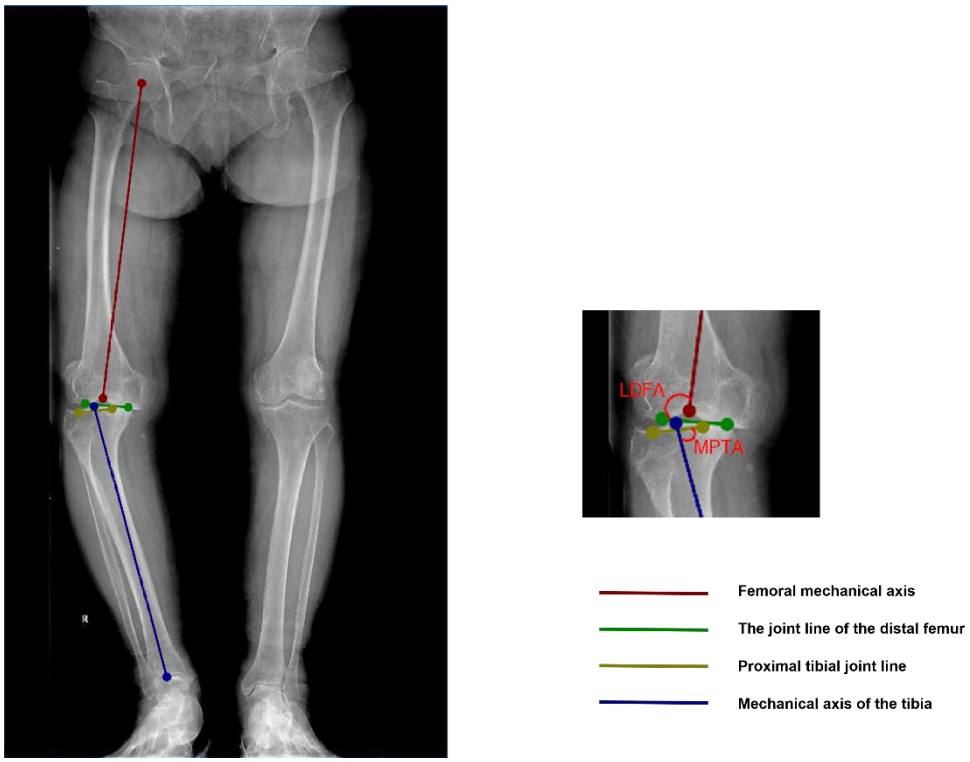


Figure 1. Labelme annotation diagram and schematic diagrams of LDFA and MPTA.

 doi:10.20944/preprints202603.2038.v1

2.4. CPAK Type Calculation

The formulas for CPAK type calculation were: aHKA = MPTA − LDFA, JLO = MPTA + LDFA. Classification thresholds were as follows: aHKA < −2° was classified as varus, −2° to +2° as neutral, and > +2° as valgus; JLO < 177° was classified as apex distal, 177°–183° as neutral, and > 183° as apex proximal. Preoperative and postoperative CPAK types (I–IX) were determined accordingly (Figure 2).

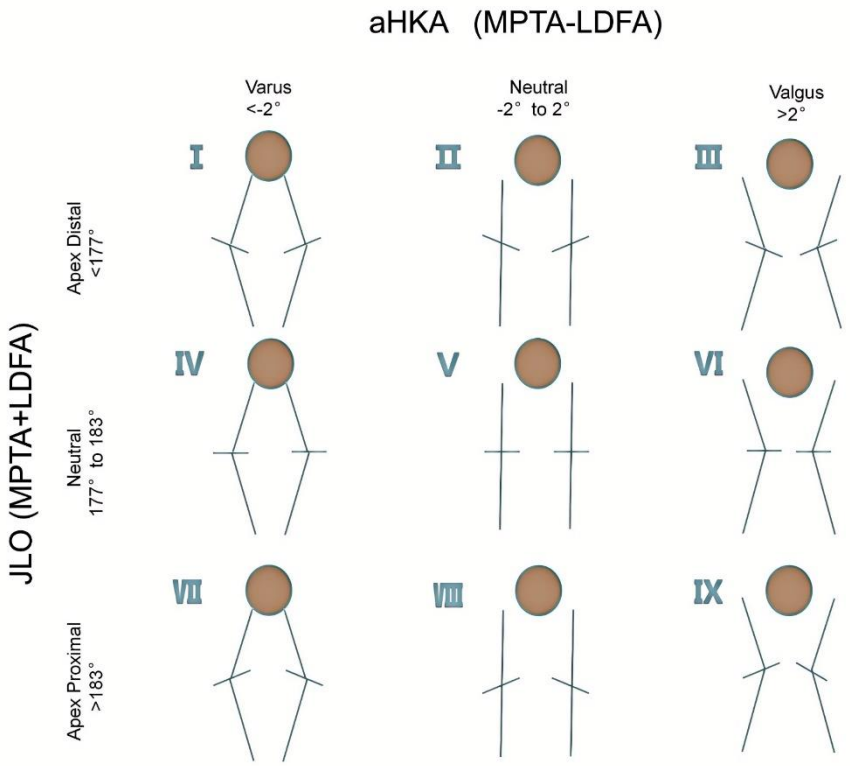


Figure 2. The nine phenotypes of knee coronal plane alignment (CPAK classification).

2.5. Deep Learning-Based Automatic CPAK Classification System

2.5.1. Overall Workflow

The deep learning-based automatic CPAK classification system developed in this study consisted of three main stages: image preprocessing, keypoint detection, and angle calculation with CPAK type output.

2.5.2. Image Preprocessing

Original full-length weight-bearing lower limb radiographs underwent side identification. For left knee images, horizontal flipping was applied to standardize them to the right knee orientation, eliminating the impact of anatomical differences between sides. Flipping parameters were recorded for subsequent coordinate mapping. All images were then uniformly resized to a fixed dimension for network input. During the model training phase, data augmentation strategies including random flipping, rotation, and brightness adjustment were further applied to enhance model generalization.

2.5.3. Keypoint Detection

This study employed the HRNet-W32 network for keypoint detection. HRNet-W32 is a deep neural network architecture based on the High-Resolution Network (HRNet)[29]. This network

 doi:10.20944/preprints202603.2038.v1

achieves an excellent balance between computational efficiency and accuracy and is widely used in the field of computer vision. The network consists of four parallel subnetworks across four stages, with each stage introducing a subsequent subnetwork. Within each subnetwork, the number of channels remains constant, while the number of channels in parallel subnetworks doubles at each stage, and the resolution decreases by 50%. This design allows the network to maintain high-resolution representations while capturing rich semantic features. Fusion modules connect the stages, exchanging and fusing feature information from different resolutions to obtain improved keypoint feature representations (Figure 3). For this specific task, based on the number of keypoints required for detection in the hip, knee, and ankle joints (eight keypoints, including the femoral head center, femoral intercondylar notch center, distal medial femoral condyle point, etc.), a convolutional layer with the corresponding number of channels was added as the output layer on top of the HRNet-W32 to form the final keypoint detection network. During model training, transfer learning was applied using HRNet-W32, with pre-trained weights from an image classification task as initial weights. After Gaussian transformation, the annotated keypoint coordinates were converted into heatmap labels for training. The trained model could detect heatmaps containing keypoint coordinate information from joint images, and a program was used to convert the heatmaps into keypoint coordinate values, thereby enabling automatic detection of the eight anatomical keypoints.

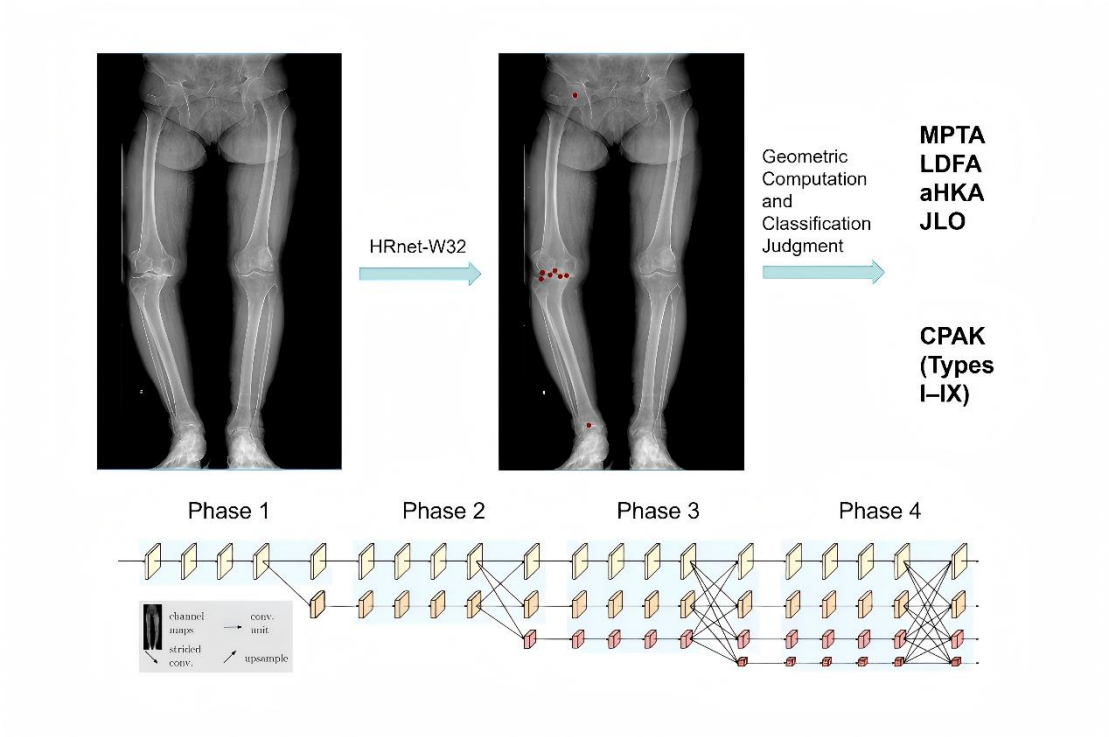


Figure 3. Schematic diagram of automatic detection of hip, knee, and ankle keypoints, angle calculation, and CPAK type generation based on the HRNet model.

2.5.4. Angle Calculation and CPAK Type Classification:

Based on the detected keypoint coordinates, MPTA, LDFA, aHKA, and JLO were geometrically calculated (see section 1.4 for formulas), and the CPAK type (I–IX) was automatically determined based on the aHKA and JLO thresholds.

2.5.5. Dataset Partitioning

All samples were divided at the patient level into a model development set (n = 827, 1654 radiographs) and a model validation set (n = 92, 184 radiographs). Subsequent clinical analyses (CPAK distribution, transition matrix, outcome association) were based on the model predictions for

 doi:10.20944/preprints202603.2038.v1

all 919 patients, i.e., combining the development and validation sets. The model development set was used for training and internal validation (using cross-validation) of the deep learning model. The model validation set was used to evaluate keypoint detection accuracy, angle measurement consistency, and CPAK classification accuracy. The model validation set was also included in the subsequent clinical distribution and transition analyses, as it was not used for any parameter updates or model selection during the development process, thus avoiding information leakage.

2.5.6. Model Evaluation Metrics

Keypoint detection accuracy was evaluated using the mean radius error (MRE). Angle measurement accuracy was expressed as mean absolute error (MAE) and standard deviation. Agreement between model measurements and manual annotations was assessed using intraclass correlation coefficients (ICC) and Bland-Altman plots. Evaluation metrics for automatic CPAK classification included overall accuracy, recall rate for each type, and Cohen's kappa coefficient.

2.6. Preoperative-to-Postoperative CPAK Type Transition Analysis

After model performance was validated, all 1838 radiographs from the 919 patients were input into the trained deep learning model. The model automatically output angle measurements and CPAK types for each image, which were used for the subsequent preoperative-to-postoperative transition analysis. It should be noted that this part of the analysis was entirely based on model predictions without manual verification.

2.6.1. Individual-Level Transition Matrix

A 9×9 contingency table was constructed, with rows representing preoperative types and columns representing postoperative types, and frequencies and row percentages were calculated.

2.6.2. Transition Pattern Grouping:

Based on the transition characteristics between preoperative and postoperative CPAK types, patients were divided into four groups: the Stable group (preoperative and postoperative types identical), the Alignment-changed group (only aHKA classification changed, JLO classification unchanged), the Joint line-changed group (only JLO classification changed, aHKA classification unchanged), and the Mixed-changed group (both aHKA and JLO classifications changed).

2.7. Association Analysis Between Transition Patterns and Clinical Outcomes

Due to baseline differences among the four groups, we performed 1:1 propensity score matching using the Stable group as the reference. Matching was performed separately with the other three groups. Covariates included age, sex, BMI, preoperative K-L grade, and surgical technique. After matching, the Stable group was combined with the three matched samples for exploratory multiple group comparisons (Kruskal-Wallis test), supplemented by pairwise comparisons between each group and the Stable group. Post-matching, the standardized mean difference (SMD) for all covariates was < 0.1, indicating good balance.

2.8. Statistical Analysis

SPSS 27.0 was used for data analysis, and Origin 24.0 was used for graph generation. Kendall's coefficient of concordance was used to evaluate consistency among expert measurements. Intraclass correlation coefficients (ICC) were used to assess consistency between model measurements and manual annotations. Continuous variables conforming to normal distribution were expressed as mean ± standard deviation ($\bar{x} \pm s$), while non-normally distributed continuous variables were expressed as median (interquartile range). Categorical variables were expressed as frequencies (percentages). Group comparisons were performed using independent samples t-test, Mann-Whitney U test, or chi-square (χ^2) test. For the dichotomized satisfaction variable, the χ^2 test was used

 doi:10.20944/preprints202603.2038.v1

for comparisons between groups. Multiple group comparisons were performed using one-way ANOVA or Kruskal-Wallis test. A two-sided P-value < 0.05 was considered statistically significant.

3. Results

3.1. Patient Baseline Characteristics

A total of 919 patients undergoing TKA for knee osteoarthritis were included in this study. Among them, there were 215 males (23.4%) and 704 females (76.6%). The mean age was 68.03 ± 6.23 years (range: 34–84 years), and the mean BMI was 26.85 ± 3.50 kg/m². Regarding K-L grades, 310 patients (33.73%) were grade 3, and 609 patients (66.27%) were grade 4. All patients completed clinical follow-up of 12–24 months postoperatively. Robotic-assisted surgery was performed in 217 cases (23.6%), and conventional surgery in 702 cases (76.4%).

3.2. Performance Validation of the Deep Learning Model

3.2.1. Keypoint Detection Accuracy

On the model validation set (92 patients, 184 radiographs), the HRNet-W32 model achieved a mean radius error (MRE) of 1.22 ± 0.43 mm for the eight anatomical keypoints. The detection accuracy for each keypoint is shown in Table 1. The femoral head center and the ankle joint center had the smallest detection errors, while the errors for the medial and lateral tibial plateau points were relatively larger.

Table 1. Keypoint detection accuracy (MRE).

Keypoint	MRE(mm)
Femoral head center	0.87±0.31
Femoral intercondylar notch center	1.02±0.38
Distal medial femoral condyle point	1.34±0.45
Distal lateral femoral condyle point	1.29±0.42
Tibial intercondylar eminence center	1.41±0.51
Medial tibial plateau lowest point	1.68±0.62
Lateral tibial plateau lowest point	1.72±0.65
Talar dome center	0.53±0.22

3.2.2. Angle Measurement Accuracy

Comparison between the model's automatically measured angles and the gold standard annotations is shown in Table 2:

The mean absolute error (MAE) for MPTA was 0.74 ± 0.42°, with 8.70% of errors ≥ 1.5°.

The MAE for LDFA was 0.72 ± 0.51°, with 15.22% of errors ≥ 1.5°.

The MAE for aHKA was 0.91 ± 0.70°, with 19.57% of errors ≥ 1.5°.

The MAE for JLO was 1.12 ± 0.66°, with 22.28% of errors ≥ 1.5°.

Table 2. Angle measurement error analysis.

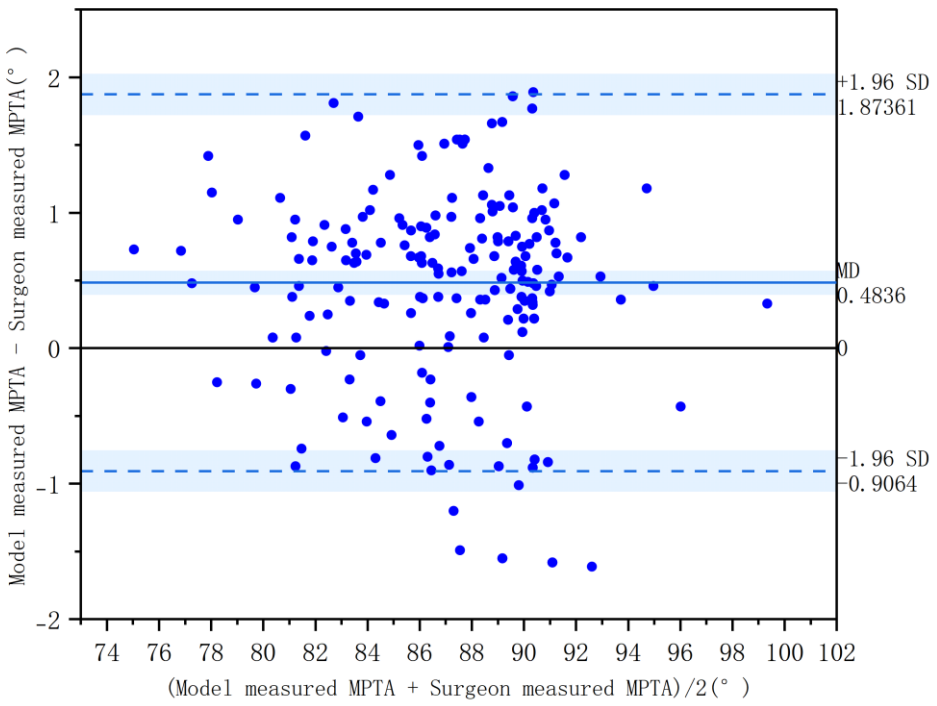
 doi:10.20944/preprints202603.2038.v1

Angle	MAE(°)	Proportion of errors ≥ 1.5° (%)
MPTA	0.74±0.42	8.70
LDFA	0.72±0.51	15.22
aHKA	0.91±0.70	19.57
JLO	1.12±0.66	22.28

3.2.3. Agreement with Manual Annotations

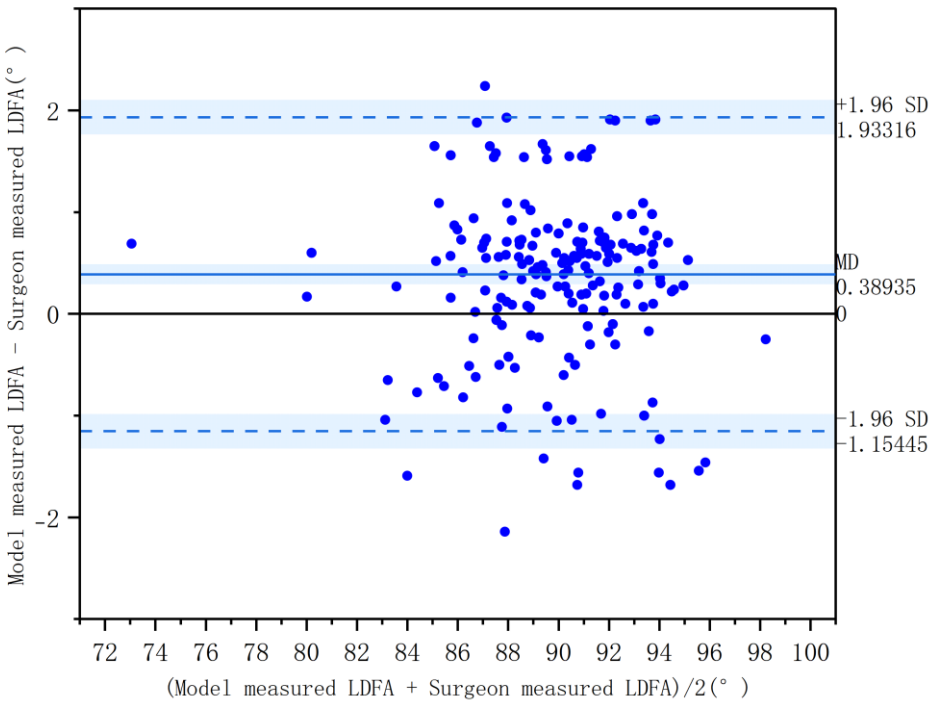
The three orthopedic surgeons demonstrated extremely high consistency in measuring MPTA, LDFA, aHKA, and JLO classifications (all $W > 0.95$, all $P < 0.05$). Their mean values were: MPTA: 87.19 ± 3.90°, LDFA: 90.01 ± 3.22°, aHKA: -2.82 ± 4.78°, JLO: 177.20 ± 5.33°. The automatic model measurements were: MPTA: 86.72 ± 3.93°, LDFA: 89.67 ± 3.25°, aHKA: -2.96 ± 4.88°, JLO: 176.39 ± 5.30°.

ICC analysis showed excellent agreement between model measurements and manual annotations: MPTA ICC = 0.977 (95% CI: 0.930–0.990); LDFA ICC = 0.963 (95% CI: 0.940–0.980); aHKA ICC = 0.972 (95% CI: 0.960–0.980); JLO ICC = 0.970 (95% CI: 0.880–0.990) (all $P < 0.001$). As shown in the Bland-Altman plots (Figure 4A–D), the measurement differences for each angle fell within the 95% limits of agreement, indicating no systematic bias.

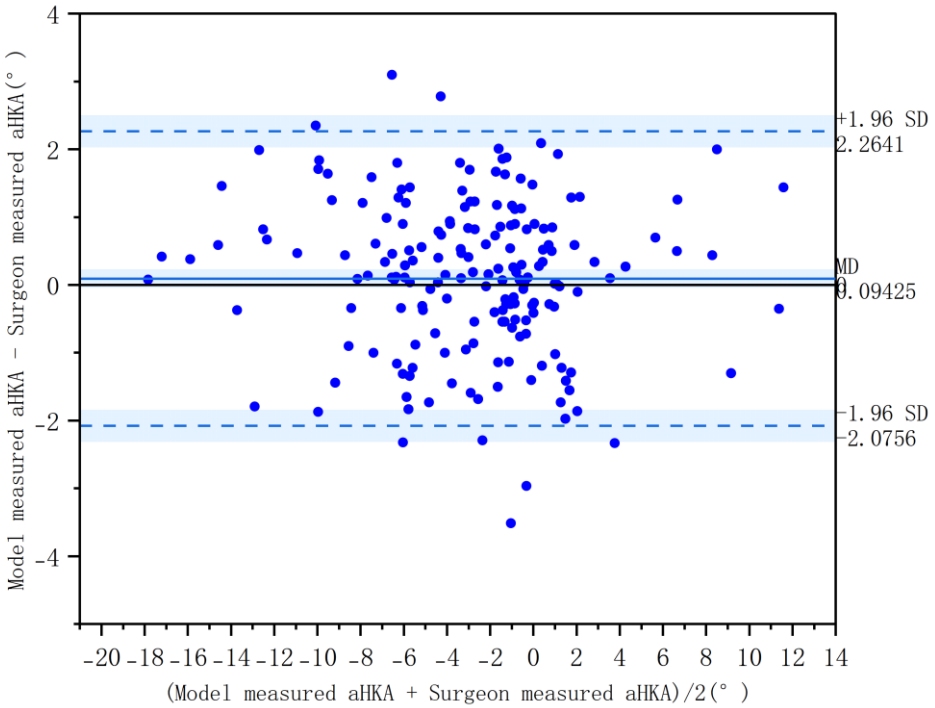


(**a**)

 doi:10.20944/preprints202603.2038.v1

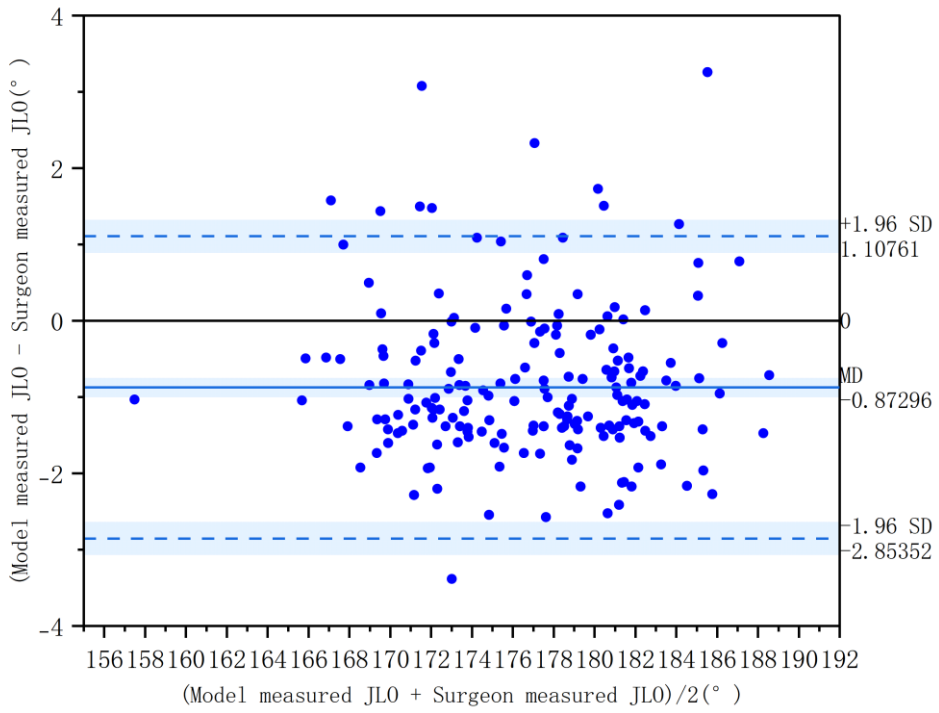


(**b**)



(**c**)

 doi:10.20944/preprints202603.2038.v1



(**d**)

Figure 4. a-d). Bland-Altman plots for assessing agreement of model-predicted angles on the validation set.(a) Medial proximal tibial angle (MPTA); (b) Lateral distal femoral angle (LDFA); (c) Arithmetic hip-knee-ankle angle (aHKA); (d) Joint line obliquity (JLO). The solid line represents the mean difference, and the dashed lines indicate the 95% limits of agreement.

3.2.4. Automatic CPAK Type Classification Evaluation

Using the manually annotated CPAK type as the gold standard, the overall accuracy of the model's automatic classification was 80.98%. Cohen's kappa coefficient was 0.767. The recall rates for each type are shown in Table 3 (type VIII had a low recall due to the small sample size), indicating good overall performance of the automatic CPAK classification and high agreement with expert judgment.

Table 3. Recall rates for each CPAK type on the model validation set.

CPAK Type	True Number	Correctly Predicted	Recall Rate (%)
1	51	49	96.08
2	27	23	85.19
3	8	6	75.00
4	32	26	81.25
5	38	29	76.32
6	6	3	50.00

 doi:10.20944/preprints202603.2038.v1

CPAK Type	True Number	Correctly Predicted	Recall Rate (%)
7	13	10	76.92
8	7	1	14.29
9	2	2	100.00
Total	**184**	**149**	**80.98**

3.3. Preoperative and Postoperative CPAK Type Distribution

The trained deep learning model was applied to automatically measure the images of all 919 patients, yielding the corresponding angle parameters and CPAK types. The distribution of preoperative and postoperative CPAK types among the 919 patients is shown in Table 4. Preoperatively, the most common types were Type I (535 cases, 58.22%), followed by Type II (139 cases, 15.13%), Type IV (97 cases, 10.55%), and Type III (75 cases, 8.16%). Types V–IX were relatively rare, collectively accounting for 7.94%. Postoperatively, the most common types were Type V (283 cases, 30.79%), Type IV (263 cases, 28.62%), and Type VII (137 cases, 14.91%), collectively accounting for 74.32%.

Table 4. Preoperative and postoperative CPAK type distribution.

CPAK Type	Preoperative n	Preoperative %	Postoperative n	Postoperative %
I	535	58.22	33	3.59
II	139	15.13	45	4.90
III	75	8.16	22	2.39
IV	97	10.55	263	28.62
V	38	4.13	283	30.79
VI	15	1.63	52	5.66
VII	9	0.98	137	14.91
VIII	2	0.22	75	8.16
IX	9	0.98	9	0.98
Total	919	100.00	919	100.00

3.4. Individual-Level Preoperative-to-Postoperative CPAK Type Transition Matrix

3.4.1. Transition Matrix

An individual-level transition matrix from preoperative to postoperative CPAK types was constructed (Table 5, row percentages). There was a significant difference in the overall distribution between preoperative and postoperative CPAK types (Stuart-Maxwell test, $\chi^2 = 647.2$, df = 8, $P < 0.001$).

Table 5. Preoperative-to-postoperative CPAK type transition matrix (row percentages, %).

 doi:10.20944/preprints202603.2038.v1

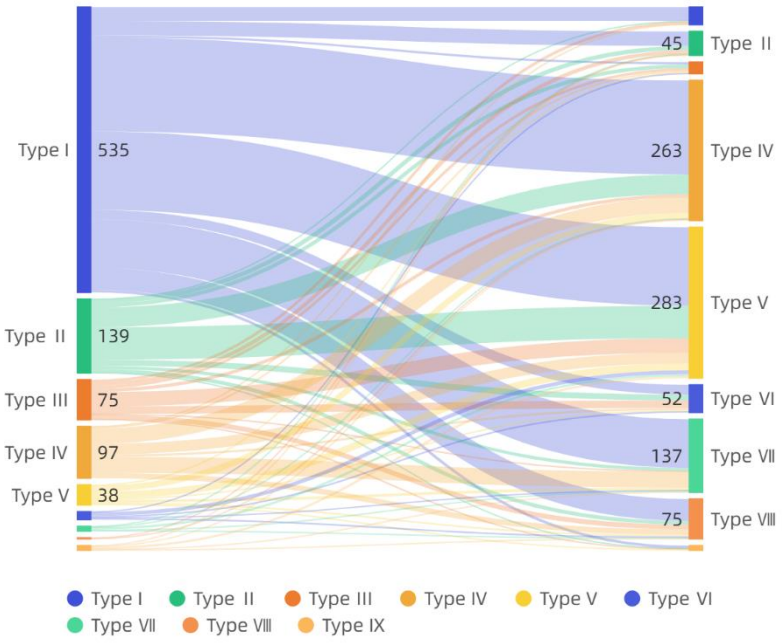
Preop\Postop	I(%)	II(%)	III(%)	IV(%)	V(%)	VI(%)	VII(%)	VIII(%)	IX(%)
I	4.86	5.05	0.75	32.90	27.48	3.36	17.01	7.66	0.93
II	0.72	5.76	5.04	26.62	43.88	7.91	4.32	5.04	0.72
III	6.67	9.33	6.67	8.00	36.00	18.67	1.33	12.00	1.33
IV	1.03	0.00	3.09	29.90	20.62	4.12	30.93	10.31	0.00
V	0.00	0.00	5.26	28.95	34.21	5.26	13.16	7.89	5.26
VI	0.00	0.00	6.67	0.00	46.67	20.00	6.67	20.00	0.00
VII	0.00	11.11	0.00	11.11	44.44	0.00	22.22	11.11	0.00
VIII	0.00	50.00	0.00	50.00	0.00	0.00	0.00	0.00	0.00
IX	0.00	11.11	0.00	22.22	44.44	0.00	11.11	11.11	0.00

3.4.2. Transition Pattern Grouping

Based on the transition patterns, patients were divided into four groups: Stable group (preoperative and postoperative types identical): 76 cases (8.30%); Alignment-changed group (only aHKA classification changed): 114 cases (12.40%); Joint line-changed group (only JLO classification changed): 386 cases (42.00%); Mixed-changed group (both aHKA and JLO classifications changed): 343 cases (37.32%).

3.4.3. Visualization of Transition Pathways

Figure 5 presents a Sankey diagram of the preoperative-to-postoperative CPAK type transitions, visually illustrating the main flow directions and patient numbers for each preoperative type.



 doi:10.20944/preprints202603.2038.v1

Figure 5. Sankey diagram of preoperative-to-postoperative CPAK type transitions.

3.5. Association Between Transition Patterns and Clinical Outcomes

3.5.1. Baseline Balance After Propensity Score Matching

The "Alignment-changed," "Mixed-changed," and "Joint line-changed" groups were each matched 1:1 with the "Stable" group using nearest-neighbor propensity score matching. Covariates included age, sex, BMI, preoperative K-L grade, and surgical technique. Before matching, some imbalances existed across the four groups in sex, preoperative K-L grade, and surgical technique (SMD > 0.1 for some comparisons), while age and BMI were generally balanced (SMD < 0.1). After matching, all covariates had absolute SMD values < 0.1, indicating good balance between the groups.

Table 6. Baseline characteristics and balance comparison before matching (full sample, n=919).

Variable	Stable (n=76)	Alignment-changed (n=114)	Mixed-changed (n=343)	Joint line-changed (n=386)
Age (years)	68.08 ± 5.73	67.63 ± 4.72 (SMD = 0.09)	68.61 ± 6.56 (SMD = -0.08)	67.64 ± 6.35 (SMD = 0.07)
BMI (kg/m²)	27.13 ± 3.59	26.81 ± 3.68 (SMD = 0.09)	26.72 ± 3.40 (SMD = 0.12)	26.92 ± 3.54 (SMD = 0.06)
Sex, n (%)				
- Male	20 (26.32%)	42 (36.84%) (SMD = 0.23)	72 (20.99%) (SMD = -0.12)	90 (23.32%) (SMD = -0.07)
- Female	56 (73.68%)	72 (63.16%)	271 (79.01%)	296 (76.68%)
Preop K-L grade, n (%)				
- Grade 3	18 (23.68%)	40 (35.09%) (SMD = 0.25)	110 (32.07%) (SMD = 0.18)	127 (32.90%) (SMD = 0.20)
- Grade 4	58 (76.32%)	74 (64.91%)	233 (67.93%)	259 (67.10%)
Surgical technique, n (%)				
- Robotic-assisted	12 (15.79%)	34 (29.82%) (SMD = 0.34)	77 (22.45%) (SMD = 0.17)	90 (23.32%) (SMD = 0.18)
- Conventional	64 (84.21%)	80 (70.18%)	266 (77.55%)	296 (76.68%)

 doi:10.20944/preprints202603.2038.v1

Table 7. Baseline characteristics and balance comparison after matching.

Variable	Stable (n=76)	Alignment-changed (n=76)	Mixed-changed (n=76)	Joint line-changed (n=76)
Age (years)	68.08 ± 5.73	67.92 ± 5.51 (SMD = 0.03)	68.21 ± 5.89 (SMD = 0.02)	67.85 ± 5.64 (SMD = 0.04)
BMI (kg/m²)	27.13 ± 3.59	27.01 ± 3.48 (SMD = 0.03)	27.21 ± 3.52 (SMD = 0.02)	26.98 ± 3.61 (SMD = 0.04)
Sex, n (%)				
- Male	20 (26.32%)	21 (27.63%) (SMD = 0.03)	19 (25.00%) (SMD = 0.03)	20 (26.32%) (SMD = 0.00)
- Female	56 (73.68%)	55 (72.37%)	57 (75.00%)	56 (73.68%)
Preop K-L grade, n (%)				
- Grade 3	18 (23.68%)	19 (25.00%) (SMD = 0.03)	17 (22.37%) (SMD = 0.03)	18 (23.68%) (SMD = 0.00)
- Grade 4	58 (76.32%)	57 (75.00%)	59 (77.63%)	58 (76.32%)
Surgical technique, n (%)				
- Robotic-assisted	12 (15.79%)	13 (17.11%) (SMD = 0.04)	11 (14.47%) (SMD = 0.04)	12 (15.79%) (SMD = 0.00)
- Conventional	64 (84.21%)	63 (82.89%)	65 (85.53%)	64 (84.21%)

3.5.2. Comparison of Clinical Outcomes Between Groups

The comparison of clinical outcomes among the four matched groups is presented in Table 8:

Satisfaction:Overall satisfaction rates were above 90% in all four groups, with no significant difference ($P > 0.05$).

KSS score:No significant difference among the four groups ($P > 0.05$).

WOMAC score:No significant difference among the four groups ($P > 0.05$).

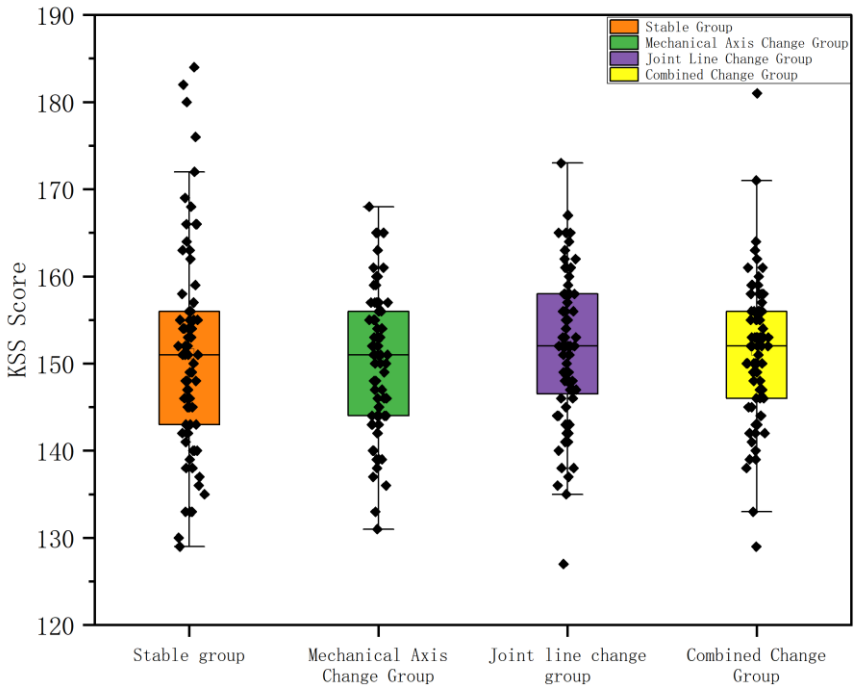
FJS score:No significant difference among the four groups ($P > 0.05$).

Figure 6 shows boxplots of the distribution of KSS scores (A), WOMAC scores (B), and FJS scores (C) for the Stable, Alignment-changed, Joint line-changed, and Mixed-changed groups.

 doi:10.20944/preprints202603.2038.v1

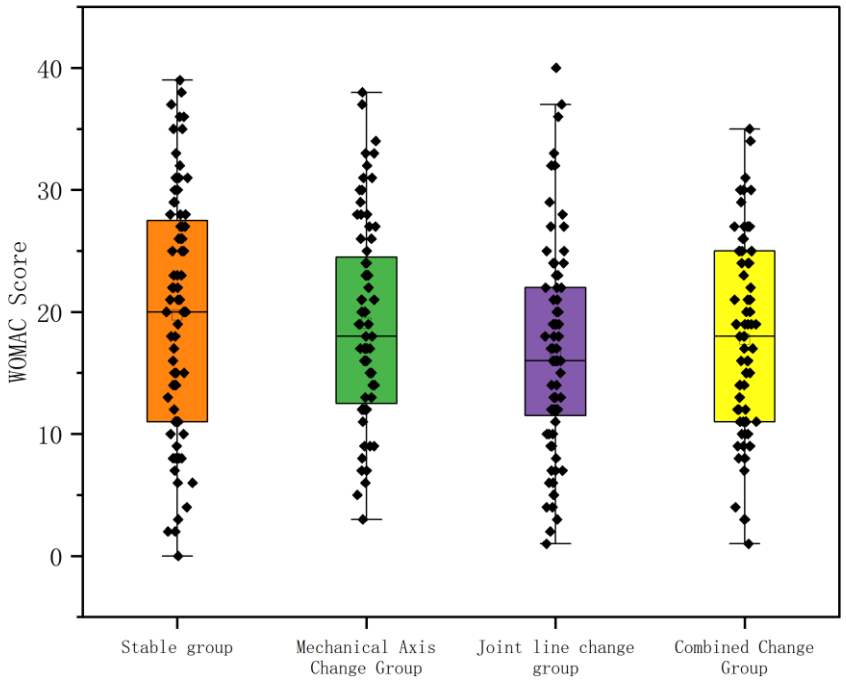
Table 8. Comparison of clinical outcomes after matching across different transition pattern groups.

Outcome	Stable (n=76)	Alignment-changed (n=76)	Joint line-changed (n=76)	Mixed-changed (n=76)	Statistic	P-value
Satisfaction, n (%)	70(92.11%)	70(92.11%)	69(90.79%)	71(93.42%)	χ^2=0.36	0.948
KSS score	151.00 (143.00,156.00)	151.00 (144.00,156.00)	152.00 (146.75,158.00)	152.00 (146.00,156.00)	H=1.68	0.642
WOMAC score	20.00 (11.00,27.25)	18.00 (12.75,24.25)	16.00 (11.75,22.00)	18.00 (11.00,25.00)	H=5.12	0.163
FJS score	61.78 ± 19.28	68.42 ± 18.82	63.32 ± 21.51	66.80 ± 18.40	F=1.87	0.135

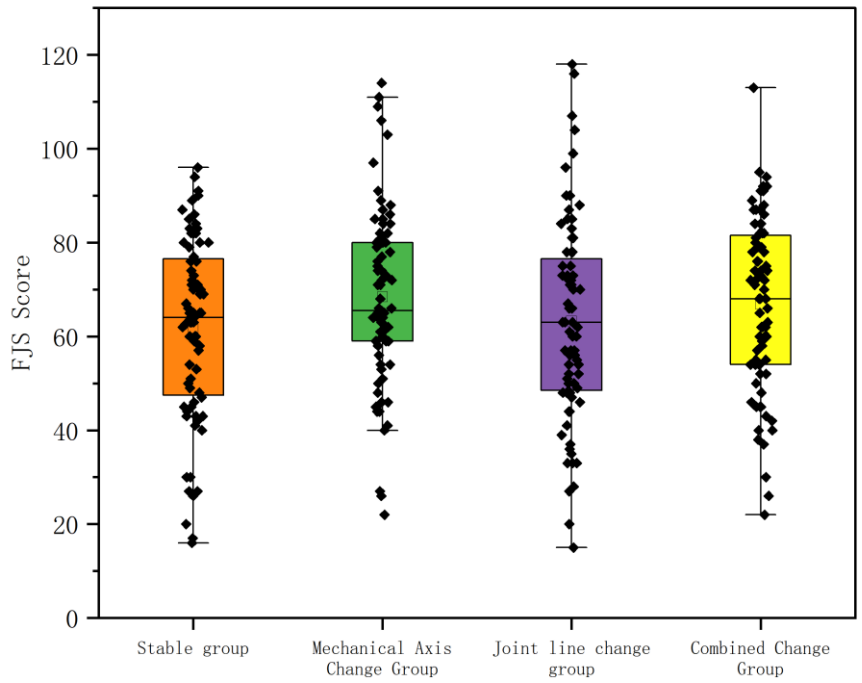


(a)

 doi:10.20944/preprints202603.2038.v1



(**b**)



(**c**)

 doi:10.20944/preprints202603.2038.v1

Figure 6. a-c). Boxplots showing clinical outcome scores across transition pattern groups. (a)Knee Society Score(KSS);(b)Western Ontario and McMaster Universities Osteoarthritis Index(WOMAC)score;(c)Forgotten Joint Score(FJS).Groups:Stable,Alignment-changed,Joint line-changed,and Mixed-changed.No statistically significant differences were observed among groups(all p>0.05p>0.05).

4. Discussion

In this study, we developed and validated a fully automated CPAK classification system based on HRNet-W32 and applied it to analyze individual-level preoperative-to-postoperative transition patterns. We systematically investigated the distribution characteristics of CPAK types, their transition patterns, and their association with clinical outcomes. The main findings and their clinical implications are discussed below.

4.1. Performance and Value of the Deep Learning-Based Automatic CPAK Classification System

Using HRNet-W32 as the backbone network for keypoint detection, our model achieved a mean radius error (MRE) of 1.22 ± 0.43 mm on the validation set. The femoral head center and talar center exhibited the highest detection accuracy, while errors for the medial and lateral tibial plateau points were relatively larger, consistent with previous reports indicating higher difficulty in detecting distal keypoints[30]. In terms of angle measurement, the MAE for MPTA and LDFA were 0.74° and 0.72°, respectively, while those for aHKA and JLO were 0.91° and 1.12°, respectively. The ICC with manual annotations were all above 0.96, and Bland-Altman plots showed no systematic bias, indicating excellent clinical consistency.

At the CPAK type level, the overall accuracy of automatic classification was 80.98%, with a kappa coefficient of 0.767, suggesting high agreement with expert judgment. Notably, the recall rate for some rare types (e.g., type VIII) was lower, likely due to their sparse distribution in the training set. This suggests that future improvements could focus on data augmentation or class weighting strategies to enhance model generalizability for rare phenotypes. Overall, the measurement accuracy and classification consistency of this system reached an acceptable level, reducing analysis time per full-length radiograph from 3–5 minutes manually to seconds, providing an efficient and objective tool for large-scale clinical studies and personalized surgical planning.

4.2. Distribution Characteristics of Preoperative and Postoperative CPAK Types in the Chinese TKA Population

The automatic classification results revealed that Type I was the predominant preoperative CPAK type in Chinese KOA patients (58.22%), followed by Types II and IV, with Type V accounting for only 4.13%. This distribution significantly differs from the predominantly Type II distribution reported by MacDessi et al. in an Australian population[19], highlighting the heterogeneity of knee anatomical phenotypes across different ethnic groups. Compared with reports from Japanese and Indian populations, our results align more closely with the actual distribution in mainland Chinese populations, further confirming the significant racial and regional specificity of the CPAK classification[9,31,32]. These findings are consistent with our team's previous research[28].

The postoperative CPAK type distribution changed significantly ($P < 0.001$), with Types V, IV, and VII predominating, collectively accounting for 74.32%. This reflects the general postoperative "correction" of alignment toward a neutral alignment or mild varus with altered joint line obliquity, consistent with current surgical strategies guided by mechanical or functional alignment principles.

4.3. Individual-Level Preoperative-to-Postoperative CPAK Type Transition Patterns

Based on individual paired data, we constructed a preoperative-to-postoperative CPAK type transition matrix, revealing detailed pathways of alignment transition. The results showed that patients with the most common preoperative Type I primarily transitioned to Types IV, V, and VII postoperatively, indicating that most patients with varus knees were corrected to a neutral or mild

 doi:10.20944/preprints202603.2038.v1

varus alignment after TKA, though some retained a varus phenotype. Preoperative Type II patients mainly transitioned to Types V and IV, while preoperative Type III patients predominantly transitioned to Types V and VI. According to the transition pattern grouping, only 8.30% of patients maintained their original type postoperatively, while the Joint line-changed and Mixed-changed groups together accounted for 79.32%. This demonstrates that TKA significantly impacts the natural coronal alignment characteristics of patients. This finding provides a comprehensive view of postoperative alignment transition and offers quantitative evidence for developing individualized alignment strategies.

4.4. Association Between Transition Patterns and Clinical Outcomes

This study found no statistically significant differences in postoperative clinical outcomes among different transition patterns. After propensity score matching, the Stable, Alignment-changed, Joint line-changed, and Mixed-changed groups showed no significant differences in KSS, WOMAC, FJS scores, or satisfaction ($P > 0.05$). This result partially aligns with the study by Sappey-Marinier et al., which also found that postoperative changes in joint line obliquity did not significantly affect clinical scores[17]. Notably, although the FJS score differences did not reach statistical significance, the mean values in the Alignment-changed and Mixed-changed groups were higher than those in the Stable group, suggesting a potential clinically meaningful trend that warrants further investigation with larger sample sizes or longer follow-up periods. The lack of a significant impact of transition patterns on clinical outcomes might be related to the relatively short follow-up period (12–24 months) and the high individual variability in patient-reported outcomes. It may also indicate that in the short term, patient satisfaction is more influenced by overall surgical outcomes such as pain relief and functional recovery, rather than by the specific pattern of alignment or joint line change.

4.5. Innovations and Clinical Significance

The innovations of this study are primarily reflected in the following aspects: it combined deep learning-based automatic CPAK classification with individual-level paired pre- and postoperative analyses, achieving fully automated assessment of alignment transition; it constructed a CPAK type transition matrix based on a large sample (919 cases) of paired pre- and postoperative data from the Chinese population, revealing individual-level transition patterns; and it systematically analyzed the association between different transition patterns and clinical outcomes, providing real-world data support for selecting individualized alignment strategies.

4.6. Limitations

This study has several limitations. It is a single-center retrospective design; all data came from one medical center, which may introduce selection bias, and the generalizability of the conclusions needs to be verified. The follow-up period was relatively short (12–24 months); the association between long-term implant survival and alignment transition remains to be investigated. Patient-reported outcomes are subjective; FJS, KSS scores, etc., may be influenced by subjective patient factors and postoperative rehabilitation differences, potentially introducing information bias. The generalizability of the deep learning model may be limited; the model was trained on images from a single center, and its adaptability to different equipment and imaging conditions requires external validation with multi-center data. In this study, the model validation set was included in the clinical analysis to maintain the completeness and representativeness of the clinical sample. Although this set was not used for model training or validation, and theoretically does not introduce overfitting bias, strictly adhering to the principle of an independent validation set would require its exclusion from subsequent analyses. This set was retained primarily to maximize the clinical sample size, and as the randomly divided validation set and clinical analysis set showed no significant baseline differences (SMD < 0.1), the impact on the main conclusions is likely limited. Furthermore, the analysis of the association between transition patterns and clinical outcomes employed a strategy of

 doi:10.20944/preprints202603.2038.v1

separate matching followed by exploratory combined comparisons. Due to the limited sample size of the Stable group, we matched subsets from the other three groups to achieve baseline balance with the Stable group and then performed multiple group comparisons. In this method, the Stable group was used repeatedly, and the three matched samples originated from different original populations, compromising the statistical independence required for multiple group comparisons, potentially affecting the accuracy of P-values. Therefore, the results of this part of the analysis should be considered exploratory, and future studies should use methods such as multinomial propensity score matching or inverse probability weighting for validation. The CPAK classification itself has limitations: it focuses only on the coronal plane and does not incorporate factors such as sagittal or rotational alignment or soft tissue balance, which also significantly impact postoperative function. Future research directions include conducting multicenter prospective studies with extended follow-up and incorporating more imaging and functional indicators; exploring the combined application value of CPAK classification with other alignment strategies (e.g., kinematic alignment, functional alignment); and further optimizing deep learning models to improve the identification of rare phenotypes.

5. Conclusions

This study, utilizing a deep learning-based automatic CPAK classification system, revealed that the predominant preoperative CPAK type in Chinese TKA patients is Type I, while postoperative types IV, V, and VII are most common. An individual-level transition matrix was constructed, demonstrating that most patients undergo significant changes in alignment and joint line post-TKA. However, no statistically significant differences in short-term clinical outcomes were observed among different transition patterns. These findings provide data-driven support for individualized TKA alignment strategies and offer a feasible tool for artificial intelligence-assisted preoperative planning and postoperative evaluation.

Author Contributions: Hua Tian,Kun Wu and Xiao Geng designed this study.Xinguang Wang,Jiazheng Chen collected data.Kun Wu analysed data and wrote this paper.

Funding: This work was supported by the Beijing Natural Science. Foundation(L234012)and Clinical Cohort Construction Program of Peking University Third Hospital(BYSYDL2023007).

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki, and ethical review and approval were waived due to the retrospective design of the study (Institutional Review Board of the authors' affiliated institution, IRB00006761-M2023429). The study was registered with the Chinese Clinical Trial Registry (NCT06010979).

Informed Consent Statement: Patient consent was waived due to the retrospective design of the study; however, written informed consent for publication was obtained from all participants.

Data Availability Statement: **T**he data presented in this study are only available on request from the corresponding author, as they are part of an ongoing project.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. DeFrance MJ, Scuderi GR. Are 20% of Patients Actually Dissatisfied Following Total Knee Arthroplasty A Systematic Review of the Literature. J Arthroplasty. 2023. 38(3): 594-599.
2. Singh M, Harary J, Schilling PL, Moschetti WE. Patient Satisfaction Is Nearly 90% After Total Knee Arthroplasty; We Are Better Than We Were. J Arthroplasty. 2025. 40(6): 1521-1525.e1.

 doi:10.20944/preprints202603.2038.v1

3. Rodriguez-Merchan EC. Patient Satisfaction Following Primary Total Knee Arthroplasty: Contributing Factors. Arch Bone Jt Surg. 2021. 9(4): 379-386.
4. Muertizha M, Cai X, Ji B, Aimaiti A, Cao L. Factors contributing to 1-year dissatisfaction after total knee arthroplasty: a nomogram prediction model. J Orthop Surg Res. 2022. 17(1): 367.
5. Roussot MA, Vles GF, Oussedik S. Clinical outcomes of kinematic alignment versus mechanical alignment in total knee arthroplasty: a systematic review. EFORT Open Rev. 2020. 5(8): 486-497.
6. Sr SJ, Sr MC, Sr SJC, Sr SJC. Methods of alignment in total knee arthroplasty, systematic review. Orthop Rev (Pavia). 2024. 16: 117769.
7. Migliorini F, Maffulli N, Pilone M, et al. Mechanical versus kinematic alignment for total knee arthroplasty: a meta-analysis. Arch Orthop Trauma Surg. 2025. 145(1): 212.
8. Alhifzi Z. Systematic review and meta-analysis of long term outcomes and innovations in Total Knee Arthroplasty: KINEMATIC, PERSONALIZED KNEE vs. CONVENTIONAL. Orthop Rev (Pavia). 2024. 16: 122318.
9. Gao YH, Qi YM, Huang PH, Zhao XY, Qi X. Distribution of coronal plane alignment of the knee classification in Chinese osteoarthritic and healthy population: a retrospective cross-sectional observational study. Int J Surg. 2024. 110(5): 2583-2592.
10. Hsu CE, Chen CP, Wang SP, Huang JT, Tong KM, Huang KC. Validation and modification of the Coronal Plane Alignment of the Knee classification in the Asian population. Bone Jt Open. 2022. 3(3): 211-217.
11. Pagan CA, Karasavvidis T, Lebrun DG, Jang SJ, MacDessi SJ, Vigdorchik JM. Geographic Variation in Knee Phenotypes Based on the Coronal Plane Alignment of the Knee Classification: A Systematic Review. J Arthroplasty. 2023. 38(9): 1892-1899.e1.
12. Rahman A, Lee M, Tan L, et al. Coronal Plane Alignment of the Knee (CPAK) distribution in a diverse Asian population: Influence of ethnicity, sex and bilaterality. J Exp Orthop. 2025. 12(2): e70192.
13. Nam HS, Park SH, Ho J, Park SY, Cho JH, Lee YS. Key-Point Detection Algorithm of Deep Learning Can Predict Lower Limb Alignment with Simple Knee Radiographs. J Clin Med. 2023. 12(4): 1455.
14. Lee HS, Hwang S, Kim SH, et al. Automated analysis of knee joint alignment using detailed angular values in long leg radiographs based on deep learning. Sci Rep. 2024. 14(1): 7226.
15. Hu Z, Cullen D, Thompson P, et al. Deep Learning-based Alignment Measurement in Knee Radiographs. Med Image Comput Comput Assist Interv. 2025 : 121-130.
16. Liu Z, Wu J, Gao X, Qin Z, Tian R, Wang C. Deep learning-based automatic measurement system for patellar height: a multicenter retrospective study. J Orthop Surg Res. 2024. 19(1): 324.
17. Sappey-Marinier E, Batailler C, Swan J, et al. Mechanical alignment for primary TKA may change both knee phenotype and joint line obliquity without influencing clinical outcomes: a study comparing restored and unrestored joint line obliquity. Knee Surg Sports Traumatol Arthrosc. 2022. 30(8): 2806-2814.
18. Ahmed AH, Din SU, Chahal K, Rowe C, Shakshak M. Kinematic Alignment Versus Mechanical Alignment in Total Knee Arthroplasty: A Systematic Review of Mid-term Functional Outcomes. Cureus. 2026. 18(1): e100874.
19. MacDessi SJ, Griffiths-Jones W, Harris IA, Bellemans J, Chen DB. Coronal Plane Alignment of the Knee (CPAK) classification. Bone Joint J. 2021. 103-B(2): 329-337.
20. M Akagawa LV, C Harman CR. Kinematic alignment in total knee arthroplasty: principles, outcomes, and evolving perspectives. Journal of Joint Surgery and Research. 2026 .
21. Karasavvidis T, Pagan CA, Debbi EM, Mayman DJ, Jerabek SA, Vigdorchik JM. No Difference in Limb Alignment Between Kinematic and Mechanical Alignment Robotic-Assisted Total Knee Arthroplasty. J Arthroplasty. 2024. 39(8S1): S200-S205.
22. Verhaar J. Patient satisfaction after total knee replacement-still a challenge. Acta Orthop. 2020. 91(3): 241-242.
23. Davey MS, Hughes AJ, Sculco PK, Star AM. Alignment Philosophies in Total Knee Arthroplasty: A Comprehensive Narrative Review. Arthroplast Today. 2026. 37: 101940.
24. Konishi T, Hamai S, Tsushima H, et al. Pre- and postoperative Coronal Plane Alignment of the Knee classification and its impact on clinical outcomes in total knee arthroplasty. Bone Joint J. 2024. 106-B(10): 1059-1066.

 doi:10.20944/preprints202603.2038.v1

25. Young SW, Tay ML, Kawaguchi K, et al. The John N. Insall Award: Functional Versus Mechanical Alignment in Total Knee Arthroplasty: A Randomized Controlled Trial. J Arthroplasty. 2025. 40(7S1): S20-S30.e2.
26. Howell SM, Zabiba A, Nedopil AJ, Hull ML. The Forgotten Joint Score after total knee arthroplasty with a kinematic alignment-optimized femoral component matches total hip arthroplasty. Knee Surg Sports Traumatol Arthrosc. 2025. 33(10): 3646-3653.
27. X Zhong ZZ, Fang H. Automatic landmark detection and angle measurement in radiographs based on deep learning in application to coronal plane alignment of the knee classification. Biomedical Signal Processing and Control. 2025 .
28. Chen J, Geng X, Wang C, et al. Coronal plane alignment of the knee classification in patients with osteoarthritis and the clinical outcomes of its alteration in total knee arthroplasty: a cross-sectional analysis of a Chinese cohort. Int Orthop. 2025. 49(5): 1081-1090.
29. K Sun BX, D Liu JW. Deep high-resolution representation learning for human pose estimation. Proceedings of the IEEE/CVF conference on computer vision and Pattern Recognition (pp. 5693-5703). 2019 .
30. Moon KR, Lee BD, Lee MS. A deep learning approach for fully automated measurements of lower extremity alignment in radiographic images. Sci Rep. 2023. 13(1): 14692.
31. Toyooka S, Osaki Y, Masuda H, et al. Distribution of Coronal Plane Alignment of the Knee Classification in Patients with Knee Osteoarthritis in Japan. J Knee Surg. 2023. 36(7): 738-743.
32. Mulpur P, Desai KB, Mahajan A, Masilamani A, Hippalgaonkar K, Reddy A. Radiological Evaluation of the Phenotype of Indian Osteoarthritic Knees based on the Coronal Plane Alignment of the Knee Classification (CPAK). Indian J Orthop. 2022. 56(12): 2066-2076.