



Predicting joint space changes in knee osteoarthritis over 6 years: a combined model of TransUNet and XGBoost

Jiangrong Guo¹, Pengfei Yan², Hao Luo², Yingkai Ma¹, Yuchen Jiang², Chaojie Ju³, Wang Chen¹, Meina Liu⁴, Songcen Lv¹, Yong Qin¹

¹Department of Orthopedics and Sports Medicine, The Second Affiliated Hospital of Harbin Medical University, Harbin, China; ²Department of Control Science and Engineering, Harbin Institute of Technology, Harbin, China; ³Ninth Department of Orthopedics, Fifth Hospital of Harbin, Harbin, China; ⁴Department of Biostatistics, School of Public Health, Harbin Medical University, Harbin, China

Contributions: (I) Conception and design: J Guo, P Yan; (II) Administrative support: S Lv, Y Qin; (III) Provision of study materials or patients: Y Ma, C Ju, W Chen; (IV) Collection and assembly of data: J Guo, Y Jiang; (V) Data analysis and interpretation: M Liu, H Luo; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Meina Liu, PhD. Department of Biostatistics, School of Public Health, Harbin Medical University, 157 Baojian Road, Harbin 150076, China. Email: liumeina369@163.com; Songcen Lv, MD, PhD; Yong Qin, MD, PhD. Department of Orthopedics and Sports Medicine, The Second Affiliated Hospital of Harbin Medical University, 246 Xuefu Road, Harbin 150086, China. Email: lsc2022@yeah.net; qinyong0125@126.com.

Background: The progression of knee osteoarthritis is mainly characterized by the reduction in joint space width (JSW). The goal of this study was to build a knee joint space segmentation model through deep learning (DL) methods and develop a model for automatically measuring JSW. Furthermore, we predicted JSW changes in the sixth year based on regression models.

Methods: The data for this study was sourced from the Osteoarthritis Initiative database. We filtered knee X-ray images from 1,947 participants and tested six neural networks for segmentation to build an automatic JSW measurement model. Subsequently, we combined the clinical data with the JSW measurement results to predict the sixth-year knee JSW using six different regression models.

Results: The segmentation results showed that TransUNet performed the best, with an overall Dice coefficient of 0.889. The intraclass correlation coefficient (ICC) between manually measured and TransUNet's automatically measured JSW reached 0.927 ($P < 0.01$). Among the regression models, eXtreme Gradient Boosting (XGBoost) demonstrated the best predictive performance, with a mean absolute error (MAE) of 0.48 and an ICC of 0.887 ($P < 0.01$). To better align with clinical practice, we reduced the prediction model to utilize only 2 years of JSW images. The results showed that using the 0- and 12-month X-ray images still achieved high accuracy, with an MAE of 0.585 ($P < 0.05$) and an ICC of 0.805 ($P < 0.01$).

Conclusions: We developed a novel JSW measurement model that significantly improves accuracy compared to previous methods and identified the best prediction model by combining TransUNet and XGBoost. Additionally, in our built model, predicting the 72-month JSW using only 2 years of knee X-ray images and several clinical features achieved high accuracy.

Keywords: Osteoarthritis; deep learning (DL); joint space width (JSW); machine learning; prediction model

Submitted Jul 09, 2024. Accepted for publication Nov 29, 2024. Published online Jan 08, 2025.

doi: 10.21037/qims-24-1397

View this article at: <https://dx.doi.org/10.21037/qims-24-1397>

Introduction

Knee osteoarthritis (KOA) is a severe degenerative disease of the knee joint that leads to a decline in quality of life (1). KOA affects the majority of adults aged 65 years and above, with a prevalence of 33.6% in the United States, and can cause significant inconvenience in daily life. In severe cases, individuals may even lose their ability to move freely (2).

The diagnosis of KOA primarily relies on clinical symptoms combined with X-ray imaging. Clinical symptoms include joint pain, stiffness, and a reduced range of motion (3). Key imaging indicators used to determine the severity of KOA include the presence of osteophytes and changes in joint space width (JSW) (4). JSW measurement is an indirect assessment of cartilage thickness (4), which due to its high reliability and responsiveness, has been recommended by the United States Food and Drug Administration (FDA) as a biological imaging biomarker for KOA (5). A reduction in JSW is believed to reflect a decrease in cartilage thickness and meniscus integrity, indicating the need for clinical intervention (6).

Although magnetic resonance imaging (MRI) provides a more precise evaluation of cartilage morphology, the low cost, high availability, and simplicity of X-ray imaging make JSW measurement the gold standard for assessing KOA progression (7). Compared to minimum JSW (mJSW), fixed JSW (fJSW) measurement is more sensitive for analyzing JSW changes (8). This allows for a comprehensive observation of joint space loss, aiding in the preventive assessment or therapeutic intervention of KOA (9). However, manual measurement of fJSW is time-consuming and requires experienced radiologists or orthopedic surgeons, prompting researchers to develop machine learning (ML)-based automatic fJSW measurement models (10).

With the advancement of medical artificial intelligence, deep learning (DL) has demonstrated a superior ability to extract complex features from various types of data (11). Convolutional neural network (CNN), a cutting-edge type of network in this field, can automatically identify and learn features in images by simulating human visual perception mechanisms, showing high accuracy in medical image analysis (12). Especially in KOA research, CNNs have been widely applied in various models (13). In recent years, multiple studies have shown that DL-based automatic fJSW measurement technology can help detect KOA progression, reduce manual measurement errors, and enhance support for clinical doctors (14,15).

The onset and progression of KOA are complex and

multifactorial. Known risk factors include age, body mass index (BMI), gender (female), repetitive knee trauma, and kneeling (16,17), among others. Predicting the progression of KOA has always been a research focus. Researchers aim to achieve precise predictions of KOA progression or pain development through various methods (18-20). Some have even predicted the likelihood of future knee replacement surgery (21). However, most existing prediction methods and models result in binary or multiclass classification predictions; they have not been able to quantify the progression of KOA or predict specific numerical outcomes.

Therefore, the objectives of this study were as follows: (I) to build a joint space segmentation model and an automatic measurement model for multiple fJSWs using X-ray; (II) to build a regression prediction model that predicts the fJSW at the sixth year using the patient's clinical data combined with 5 years of X-ray data; and (III) to simplify the model that enhance the practical clinical value by reducing data while ensuring accuracy in predicting fJSW. We expect that the automatic measurement model for fJSW will help clinicians to more accurately and quickly determine the loss of JSW, and the prediction model will accurately predict the future state of patients' fJSW, indirectly assessing the progression of KOA. We present this article in accordance with the TRIPOD+AI reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-1397/rc>).

Methods

Data collection

This retrospective study utilized data entirely sourced from the public Osteoarthritis Initiative (OAI) database, a multi-center, longitudinal, prospective observational study on KOA. All data can be accessed at the OAI database: <https://nda.nih.gov/oai> (22). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). OAI obtained written informed consent from all participants.

The database recruited 4,796 participants for a 108-month longitudinal follow-up on KOA. We collected relevant data from the 0- (baseline), 12-, 24-, 36-, 48-, and 72-month intervals. Participants included men and women of all ethnicities, aged 45–79 years, who had KOA or were at risk for KOA. We excluded patients who dropped out of follow-up, had missing clinical data or images, or had undergone knee replacement surgery during the 6-year

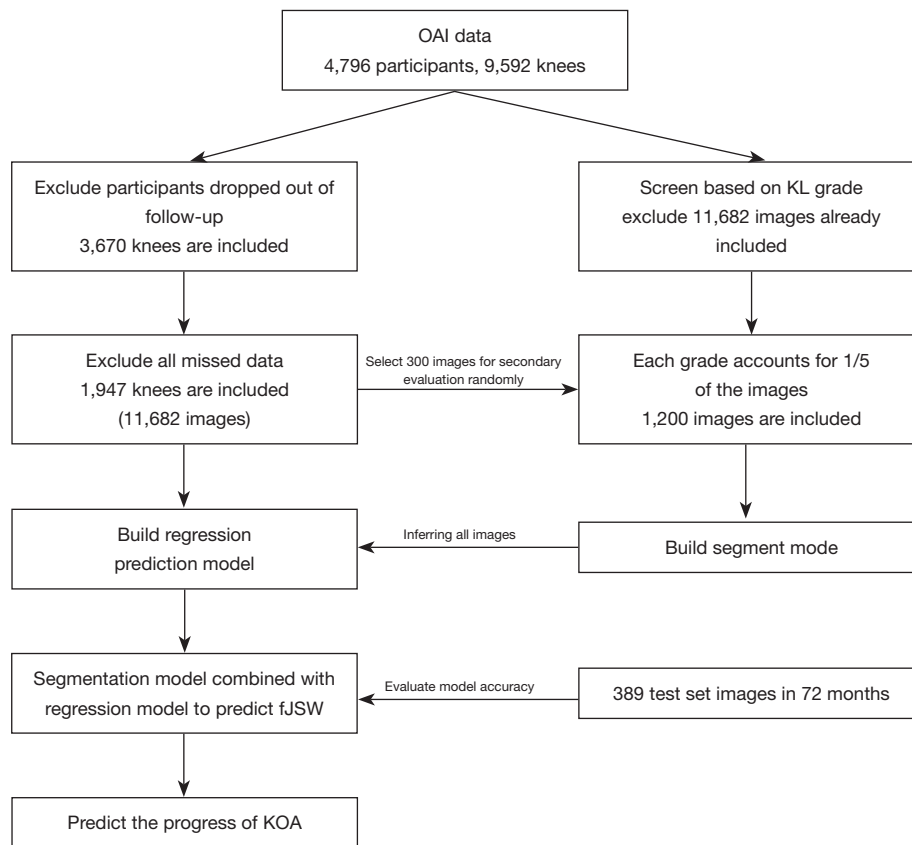


Figure 1 Data sources, quantities, and usage methods. OAI, Osteoarthritis Initiative; KOA, knee osteoarthritis; KL, Kellgren and Lawrence; fJSW, fixed joint space width.

period. In the end, we selected 1,947 knees, and a total of 11,682 images (1,947×6) were used to construct the prediction model (Figure 1).

Additionally, we screened 1,200 more images from the OAI database according to Kellgren and Lawrence (KL) grades to build the joint space segmentation model. These images included KL grades 0–4, with each grade accounting for 1/5 of the dataset to ensure the robustness of the segmentation model when processing different knee joint X-rays. We excluded all images from the 1,947 participants already screened to avoid bias in evaluating the model's accuracy. The OAI project used a standardized fixed flexion method for taking knee X-rays (23) to ensure consistent positioning and angles when measuring fJSW across different patients.

Data labeling and manual measurement of joint space were performed by two orthopedic researchers under the supervision of senior surgeons (Y.Q., with 15 years of orthopedic experience, and S.L., with 30 years of orthopedic experience) using 3D Slicer (Version 5.2.2; National

Institutes of Health, Bethesda, MD, USA). The results were cross-checked to ensure accuracy. Labeling focused on the medial and lateral joint spaces of the knee, with particular attention paid to deviations caused by the edge perspective effects of the femur and tibia. Two labels were generated: joint space (RGB = 127, 127, 127) and background (RGB = 0, 0, 0) (Figure 2).

In total, we manually labeled 1,889 X-ray images, categorized as follows: (I) 1,200 images used to build the segmentation model; (II) 300 images randomly selected from the 1,947 participants for secondary evaluation of the model's segmentation accuracy; and (III) 389 images from the test set (72-month data of the 1,947 participants) used to evaluate the accuracy of the regression model.

Labeling and training the joint space segmentation model

Before using neural networks for learning, we first adjusted the images to a uniform size of 480×480 pixels by cropping and resizing them. We then employed six

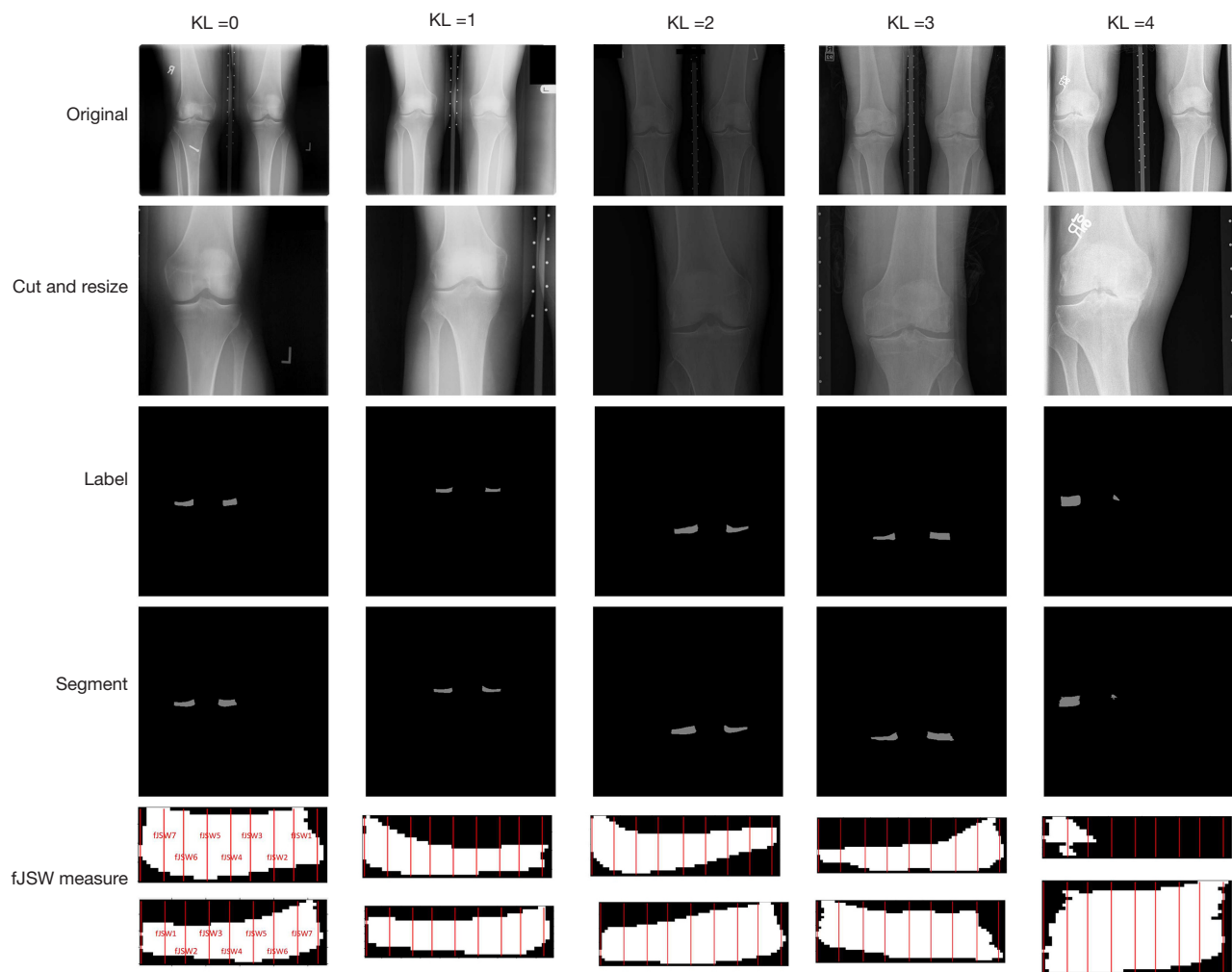


Figure 2 The first row of images displays the original OAI X-ray images, whereas the second row shows the images after cropping and resizing, with the pixel size uniformly adjusted to 480×480. The third row presents the manually annotated images, and the fourth row showcases the images segmented by the neural network (using TransUNet segmented images as the example). The fifth and sixth rows illustrate the visual measurement methods for medial JSW and lateral JSW, respectively, with measurements labeled from edge to center as fJSW 1–7. Columns 1–5 represent KL grades 0–4. OAI, Osteoarthritis Initiative; JSW, joint space width; fJSW, fixed joint space width; KL, Kellgren and Lawrence.

widely used CNN image segmentation models for learning: U-Net (24), UNet++ (25), ResUNet (26), DeepLab V3+ (27), fully convolutional network (FCN) (28), and TransUNet (29). A total of 1,200 labeled images were used for training, with the dataset split into training, validation, and test sets in a 6:2:2 ratio. Additionally, 300 labeled images were reserved for secondary performance evaluation. All six algorithms were implemented under the PyTorch framework (Meta AI, New York City, NY, USA) and computed using a tower server equipped with four

NVIDIA 12GB GPUs (NVIDIA, Santa Clara, CA, USA). To ensure a fair comparison, the hyperparameters of all models were kept consistent (batch size =64, learning rate =0.001, optimizer = Adam, and epochs =100). After the segmentation was completed, we resized the images back to their original scale to maintain the original number of pixels during subsequent gap measurements (*Figure 2*).

FCN

This model consists of both a fully convolutional part

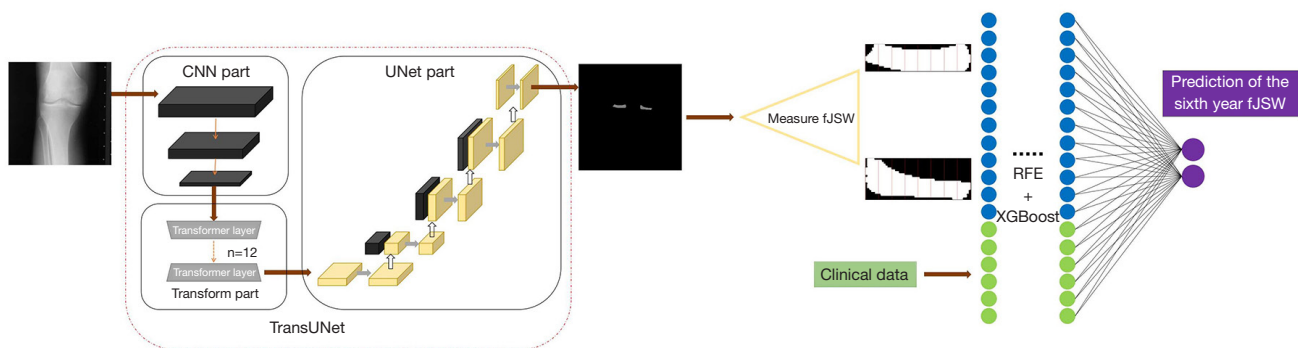


Figure 3 The entire prediction model flowchart. After inputting the image, the fJSW is first segmented using TransUNet that consisting of a CNN part, Transform part, and UNet part decoder module. It then automatically distinguishes and measures the medial and lateral fJSW. Finally, the measurement results are combined with the patient's clinical data to predict fJSW in the sixth year with the help of XGboost. fJSW, fixed joint space width; CNN, convolutional neural network; XGBoost, eXtreme gradient boosting; RFE, recursive feature elimination.

and a deconvolution part. VGG16 serves as the backbone network, where the final fully connected layer is replaced with a 1×1 convolutional layer to extract features and generate a heatmap. Subsequently, the small-sized heatmap is restored to the original size through deconvolution.

U-Net

Widely used in medical image segmentation, U-Net features a U-shaped structure with a symmetric encoder and decoder. The encoder gradually reduces the image size from 480×480 to 30×30 through convolution and pooling operations to extract features. Several tensors are then concatenated, and the decoder generates prediction results through layer-by-layer upsampling and pixel-by-pixel classification.

U-Net++

An enhanced version of U-Net, U-Net++, introduces nested skip pathways for flexible feature fusion in the decoder. It also includes dense skip connections that link feature maps of different layers to improve feature transmission and fusion, capturing more details and contextual information.

ResU-Net

This is an improved version of U-Net that incorporates residual connections (similar to ResNet), allowing features to be passed not only to the next layer but also directly to deeper layers. This effectively mitigates the gradient vanishing problem during training.

DeepLab V3+

This model uses ResNet50 as its backbone and consists of an encoder and decoder. The encoder extracts high-level semantic features, which the decoder then uses to progressively restore spatial resolution, producing finer segmentation results. The atrous spatial pyramid pooling (ASPP) module enhances the model's ability to fuse multi-scale features, improving its understanding of complex scenes.

TransUNet

This segmentation network is based on the transformer model, employing a hybrid CNN-Transformer-U-Net architecture. It first uses CNN for feature encoding, reducing the image size to 30×30 . The features are then fed into a transformer module that utilizes a self-attention mechanism to extract global context information. Finally, the U-Net decoder module upsamples the encoded features, restoring spatial resolution layer by layer and combining them with CNN features from the encoder path for precise localization (Figure 3).

fJSW measurement

The measurement of JSW was implemented in MATLAB (R2022a; MathWorks, Natick, MA, USA). After segmenting the joint space, the images were resized to their original size, and the two largest regions of interest (RoI) representing the joint spaces were identified, discarding any

smaller regions that might contain segmentation errors. The area ratio and length ratio of these two regions were then analyzed. If the area ratio was less than 0.6 and the length ratio was also less than 0.6, or if the area ratio was less than 0.4 and the length ratio was less than 0.7, the lengths of the medial and lateral RoIs were set equal, with the width being 0 mm. This adjustment ensures accurate fJSW estimation in cases where parts of the joint space are completely lost, which is often observed when the KL grade is 4.

The first step in fJSW measurement involves distinguishing between the medial and lateral joint spaces. To achieve this, we flipped the right knee images so that the left side of each image represents the medial joint space, and the right side represents the lateral. If only one RoI was present, it indicated that the JSW on the other side was absent, corresponding to a KL grade of 4. Next, the two longest horizontal lines within the medial and lateral RoIs were identified and divided into nine vertical lines. The middle seven vertical lines were used to count the number of pixels intersecting with the RoI (Figure 2). The JSW was then calculated by converting the pixel count into actual distances, using the pixel spacing value from the original X-ray's DICOM tag (Figure 2, lower left corner).

Data filtering and fJSW prediction

First, we performed inference on 11,682 images from 1,947 participants and obtained all fJSW results using the segment model and JSW measurement model. Based on clinical experience, we selected 67 potential risk factors from the OAI database that may influence KOA progression, including age, BMI, gender, repeated knee trauma, and kneeling, among others (see Table S1 for specific selection items).

In the regression model, the fJSW results at 0-, 12-, 24-, 36-, and 48-month, along with the selected risk factors, were used as predictors, whereas the 72-month fJSW served as the prediction target. Additionally, we compared six models: random forest (RF) (30), back propagation neural network (BPNN) (31), long short-term memory (LSTM) (32), CNN (33), eXtreme Gradient Boosting (XGBoost) (34), and light gradient boosting machine (LightGBM) (35) to select the best one, and combined these with the recursive feature elimination (RFE) method for feature selection. RFE removes one feature in each iteration to evaluate the independent contribution of each feature and retrains the model on the simplified feature set. This

method not only helps to simplify the model and reduce the risk of overfitting but also retains features that significantly impact the final prediction, thereby enhancing the model's generalization ability and accuracy (36).

We divided the dataset into training and testing sets at an 8:2 ratio, with the training set further split into training and validation sets for cross-validation to identify the important features selected by RFE combined with each regression method. A total of 389 images in the test set were manually annotated as the gold standard for evaluating the accuracy of the regression model. The regression predictions were implemented in Matlab (R2022a), and the hyperparameters of the six regression models were determined using grid search and five-fold cross-validation. The regression prediction results were explained using SHapley Additive exPlanations (SHAP).

RF

A decision tree-based ensemble learning algorithm that constructs models by randomly sampling multiple subsets and training multiple decision trees. The hyperparameters are set as 500 estimators, a maximum depth of 20, a minimum samples leaf of 4, and a minimum sample split of 5.

BPNN

This model computes output results through forward propagation and updates weights using backpropagation in conjunction with expected values, thereby training the model for predictions. BPNN uses the Adam optimizer, with a maximum of 100 iterations, 4 hidden layer nodes, and a learning rate of 0.01.

LSTM

This model performs regression prediction on time series data by progressively passing information layer by layer, incorporating memory gates, forget gates, and cell states to effectively address gradient vanishing and explosion issues in long-term dependencies. LSTM employs the Adam optimizer, with 100 epochs, 4 hidden units, and a learning rate of 0.01.

CNN

Capable of handling both image problems and regression prediction tasks, CNN utilizes its unique convolutional and pooling layers for dimensionality reduction and feature extraction, conducting regression analysis. CNN uses the SGDM optimizer, with a maximum of 100 epochs and a learning rate of 0.01.

XGBoost

A decision tree algorithm based on gradient boosting that makes predictions through multiple trees. Each tree is constructed to correct the prediction errors of all previous trees, resulting in the target value being the difference from previous predictions. In each iteration, XGBoost builds a new decision tree by minimizing the loss function to reduce prediction error as much as possible. The hyperparameters are set as 40 estimators, a maximum depth of 3, 50 leaves, and a learning rate of 0.1 (*Figure 3*).

LightGBM

An improvement on XGBoost that utilizes a histogram-based decision tree algorithm and introduces Gradient-based One-Side Sampling and Exclusive Feature Bundling, enhancing model performance. LightGBM sets the parameters to 40 estimators, a maximum depth of 10, 30 leaves, and a learning rate of 0.1.

Statistical analysis

Data collected in this study were statistically analyzed using Matlab. Cartilage segmentation results were evaluated for accuracy using the dice similarity coefficient (DSC), Intersection over Union (IoU), and average surface distance (ASD); DSC and IoU both ranged from 0 to 1, with higher values indicating better segmentation performance. ASD is used to evaluate the average difference between the segmentation boundary and the true boundary, with values closer to 0 indicating smaller differences, measured in 'mm'.

The accuracy of fJSW measurement was evaluated using the intraclass correlation coefficient (ICC), with values greater than 0.75 considered excellent, 0.75–0.4 moderate, 0.11–0.4 low, and below 0.1 indicating no consistency ($P < 0.05$).

Regression prediction models were assessed using root mean squared error (RMSE), mean absolute error (MAE), and R-square (R^2). RMSE measures the deviation between predicted and actual values, whereas MAE assesses the average absolute deviation, with both being better when they are closer to 0. R^2 ranges from 0 to 1, with values closer to 1 indicating better model fit. RMSE and MAE measure the differences between predicted and observed values, whereas R^2 assesses the explanatory power of the model. To assess the statistical significance of the regression predictions, the Kruskal-Wallis test was conducted, followed by Dunn's test for pairwise comparisons. A P value of less than 0.05 was considered to indicate statistically significant

differences among the results.

Results

Segment and measure results

The results from the six deep neural network segmentation models tested indicated that TransUNet achieved the best segmentation performance on the test set, with an overall DSC of 0.889, an IoU of 0.802, and an ASD of 0.498 mm. In the secondary evaluation conducted on 300 images, TransUNet again demonstrated superior performance, achieving a DSC of 0.885, an IoU of 0.795, and an ASD of 0.474 mm (*Table 1*). The pixel spacing of the images ranged from 0.1 to 0.194 mm, with an average value of 0.15 mm, making the ASD approximately 3.24 times the average pixel spacing. Furthermore, when comparing the ASD values with the average JSW of 5.72 mm, it is evident that the ASD values are substantially smaller than the JSW, indicating high segmentation precision relative to anatomical structures.

The results of automatic fJSW measurement after segmentation were statistically analyzed using the ICC consistency test with our manually labeled and measured results. The ICC of fJSW segmented by TransUNet was 0.927, consistent with the best network results of the segmentation model (*Table 1*).

We also analyzed the correlation between our manually segmented measurement results and another scholar: Duryea developed a customized software (37) for fJSW measurement on OAI knee X-rays. The ICC between our manually labeled measurements of the test set images and the measurements by this customized software yielded a consistency result of 0.611, which suggested that there was a certain correlation between the customized software's results and our observations. The differences were visualized using a Bland-Altman plot (*Figure 4*).

Prediction results

A total of 70 features of fJSW data obtained from measurement result (14 fJSW data * 5 years) were combined with 67 clinical features selected based on clinical experience (38). Six regression prediction models—RF, BPNN, LSTM, CNN, XGboost, and LightGBM—were used to predict the fJSW of 1,947 participants at 72 months. Five-fold cross validation RFE was combined with prediction models to remove features with low correlation.

Table 1 Result of six segmentation models

Evaluation metrics	TransUNet	UNet	ResUNet	UNet++	FCN	Deeplab V3+	Customized software
Test set of DSC	0.889	0.877	0.87	0.838	0.847	0.846	—
Test set of IoU	0.802	0.782	0.772	0.74	0.737	0.736	
Test set of ASD	0.498	0.535	0.561	0.577	0.73	0.732	
Secondary evaluation of DSC	0.885	0.87	0.866	0.857	0.842	0.835	
Secondary evaluation of IoU	0.795	0.773	0.768	0.76	0.73	0.719	
Secondary evaluation of ASD	0.474	0.502	0.546	0.604	0.681	0.723	
ICC	0.927	0.908	0.853	0.801	0.791	0.855	0.611

Results of ICC between manually labeled measurement of fJSW and automatic measurement. As well as the ICC between manually labeled measurement of JSW and measurement by Duryea's customized software, ICC consistency test, $P < 0.01$. “—” means this part was not statistically analyzed. DSC, dice similarity coefficient; IoU, Intersection over Union; ASD, average surface distance, measured in ‘mm’; ICC, intraclass correlation coefficient; fJSW, fixed joint space width; FCN, fully convolutional network.

The RFE method indicated that cross-validation accuracy peaked with 43 features, and further increases in features did not significantly enhance accuracy and instead added complexity or slightly led to overfitting (*Figure 5*). All selected features included 13 clinically relevant features and 30 fJSW measurement features, with similar results merged to yield eight essential clinical features necessary for predicting fJSW using the model (*Table 2*, features before merging are in *Table S2*).

The results indicated that XGBoost achieved the best predictive performance, with an R^2 of 0.804, an MAE of 0.48, and an RMSE of 0.697 on the test set of 389 cases (*Table 3*). The Kruskal-Wallis test followed by Dunn's post-hoc test showed that XGBoost had statistically significant differences ($P < 0.05$) in predictive performance compared to the other five machine learning models. The predictive results of all six models were visualized using scatter plots to demonstrate the accuracy of the regression models (*Figure 5*). A comparative analysis of the manual fJSW measurements on the test set against the predictions from the regression model was conducted using ICC consistency tests, where XGBoost displayed the best performance with an accuracy of 0.887 (*Table 3*).

Based on the prediction results, SHAP values were calculated to quantify the importance of each feature, with the y-axis reflecting their significance in the prediction model. The SHAP values clarified each feature's individual contribution to the model's prediction, where the impact of each feature was marked by colored points to distinguish between high and low feature values (*Figure 6*).

Reducing data

Predicting the 72-month fJSW with data from five consecutive years yielded high accuracy. However, this is impractical for real clinical applications, as most patients cannot undergo annual follow-up for KOA progression. Knee data of only 1 year cannot predict future fJSW, so we attempted to use data with minimal follow-up, predicting joint space changes in the sixth year based solely on 2-year results (0- + 12-month prediction, 0- + 24-month prediction, etc.). The results showed that using 0- + 48-month for predicting the 72-month fJSW achieved accuracy closest to that of using all data, with an R^2 of 0.803, RMSE of 0.74, MAE of 0.516, and ICC of 0.846. Additionally, even using 0- + 12-month for prediction indicated relatively high accuracy, with an R^2 of 0.776, RMSE of 0.85, MAE of 0.585, and ICC of 0.805. This suggests that with our model, patients only need 2 years of follow-up to achieve reasonably accurate 6-year knee JSW change predictions (*Table 4*).

Discussion

KOA is a slowly progressive disease characterized by irreversible joint damage. In this study, we combined segmentation models, which excel in image processing, with predictive models, which excel in handling sequential data, to develop a model capable of predicting KOA fJSW through X-ray images and clinical data. We proposed a new approach to predict patients' future knee fJSW and identified the best combination, achieving the highest

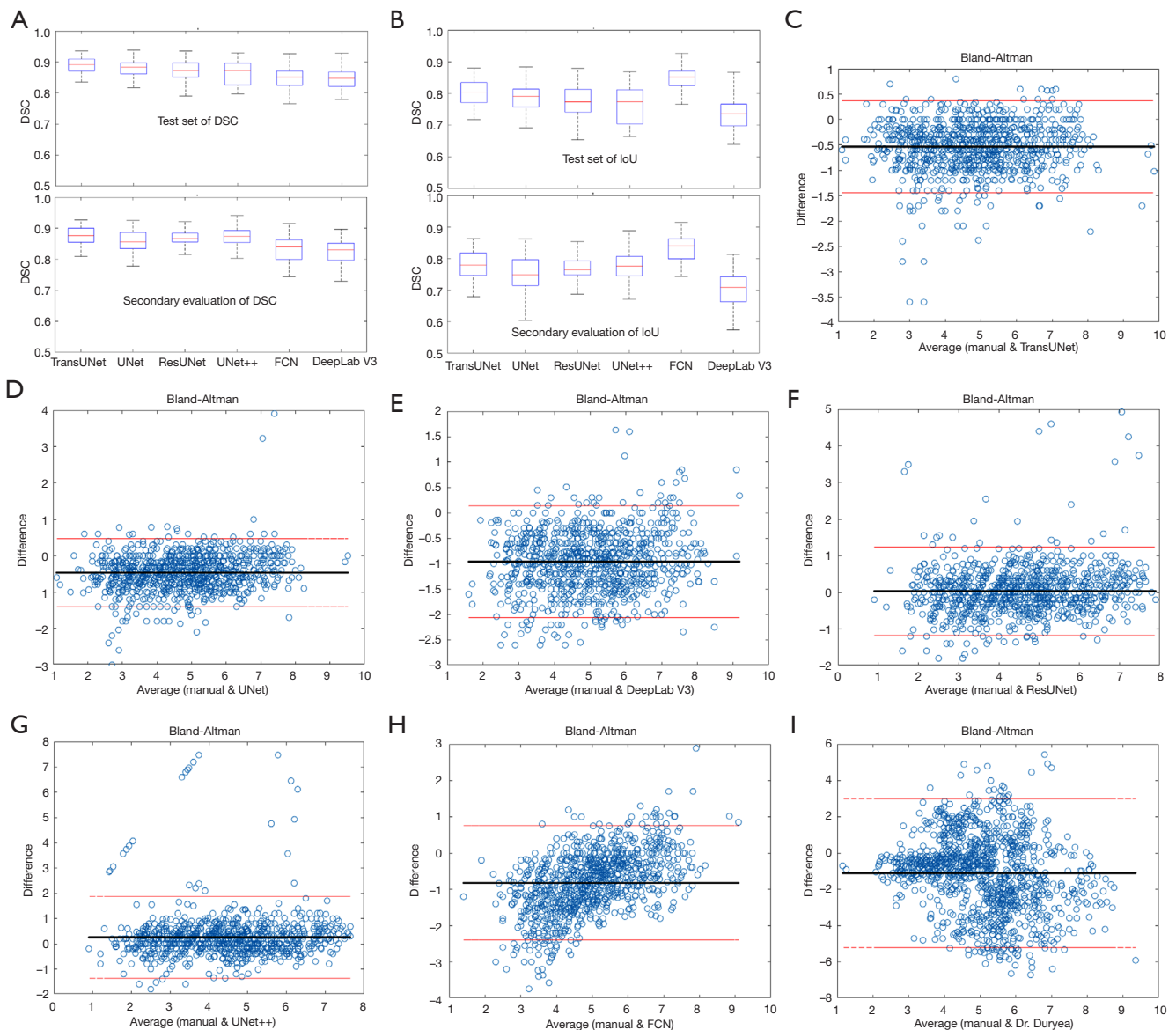


Figure 4 Result of segment models. (A,B) The box plots of total DSC and IoU for six different networks; (C-H) Bland-Altman plots comparing the fJSW measurement results of six different networks; (I) customized software with manually labeled and measured fJSW. DSC, dice similarity coefficient; IoU, Intersection over Union; fJSW, fixed joint space width; FCN, fully convolutional network.

prediction accuracy with TransUNet and XGboost.

Observing knee JSW through X-ray is the most commonly used imaging method for diagnosing KOA in clinical settings. Early scholars used ML to identify joint spaces. Marijnissen *et al.* (7) and Oka *et al.* (10) developed JSW measurement models in 2007 and 2008, respectively. Marijnissen *et al.*'s model measured JSW by identifying the boundaries of the femur and tibia on X-rays, with an average correlation of 0.71 compared to KL grade. Oka

et al.'s program not only automatically measured JSW but also measured the femoral-tibial angle and analyzed the presence of osteophytes, which our model lacks. Although these methods achieved good reliability through automatic measurement, the process was very time-consuming and required a digital viewer (39). In recent years, some researchers have used DL to measure JSW by labeling and segmenting the regions of the femur and tibia (40). After segmentation, they extracted the RoIs of the joint space and

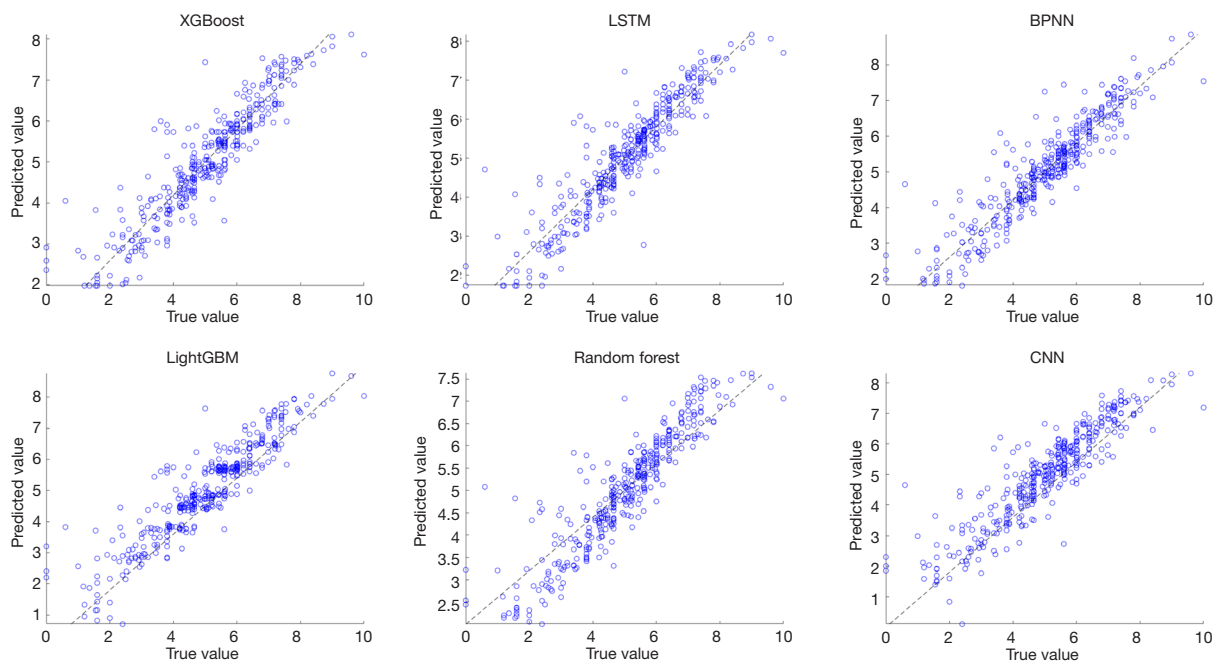


Figure 5 Comparison of fJSW scatter plots predicted by six regression models. XGBoost, eXtreme gradient boosting; LSTM, long short-term memory; fJSW, fixed joint space width; BPNN, back propagation neural network; LightGBM, light gradient boosting machine; CNN, convolutional neural network.

Table 2 The necessary clinical features of XGBoost prediction model

Feature code	OAI variable name	Description	Type
2	P02SEX	Patient’s sex (1: male, 2: female)	Binary
3	P01FAMKR	Blood relative ever had knee replacement surgery for arthritis	Binary
13	V01WOMTSR	WOMAC total score	Continuous
16	V01CHNFQCV	Chondroitin sulfate frequency of use	Ordered categorical
31	V05KOOSKPR	KOOS pain score	Continuous
50	V05KOOSYMR	KOOS symptoms score	Continuous
51	V06SF1	In general, how is health	Ordered categorical
52	V06LKSX	Knee symptom status	Ordered categorical

XGBoost, eXtreme gradient boosting; OAI, Osteoarthritis Initiative; WOMAC, Western Ontario and McMaster Universities Arthritis Index; KOOS, Knee Injury and Osteoarthritis Outcome Score.

Table 3 Evaluation of prediction results of regression models

Evaluation metrics	XGBoost	LSMT	BPNN	LightBGM	RF	CNN
R ²	0.804	0.792	0.786	0.756	0.725	0.717
MAE	0.48	0.498	0.501	0.585	0.588	0.594
RMSE	0.697	0.708	0.728	0.813	0.822	0.839
ICC	0.887	0.875	0.872	0.857	0.837	0.846

Evaluation of the accuracy of predicting 72-month fixed joint space width using different regression models. XGBoost, eXtreme gradient boosting; LSTM, long short-term memory; BPNN, back propagation neural network; LightGBM, light gradient boosting machine; RF, random forest; CNN, convolutional neural network; R², R-square; MAE, mean absolute error; RMSE, root mean squared error; ICC, intraclass correlation coefficient, P<0.01.

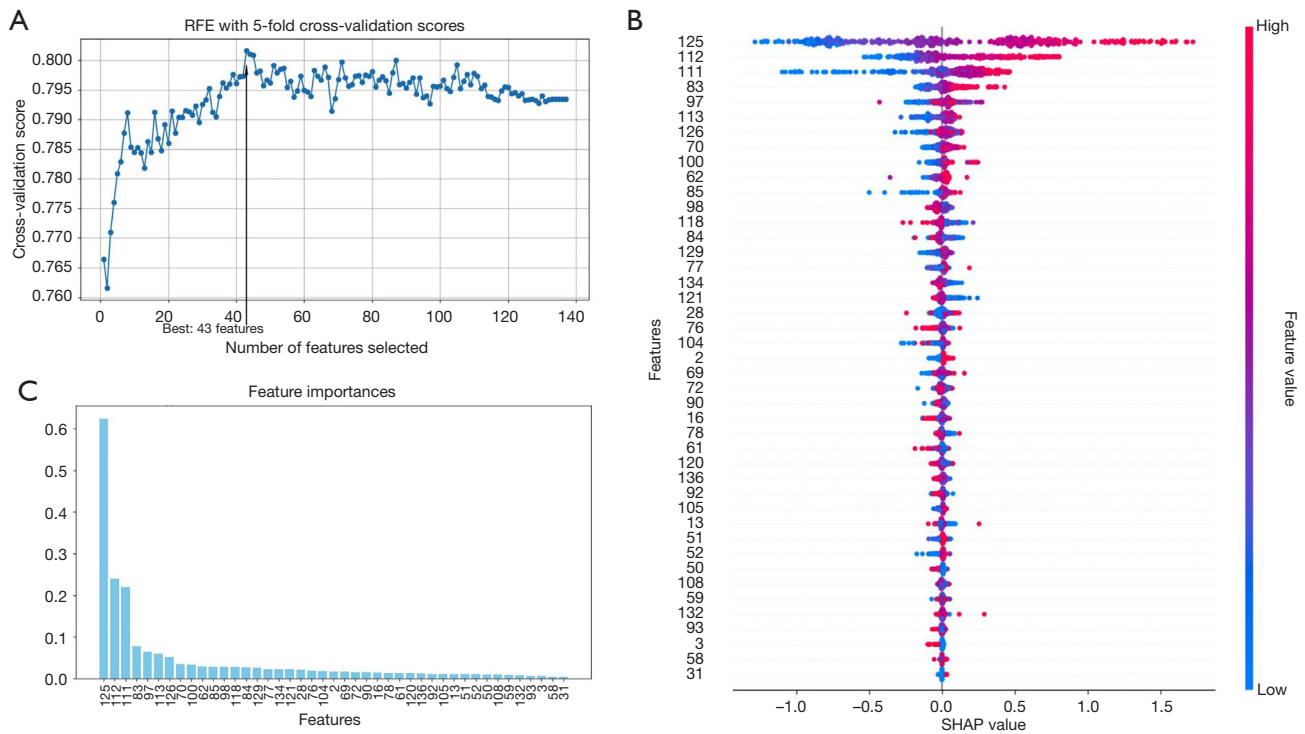


Figure 6 The XGBoost regression prediction model predicts 72-months fJWS. (A) RFE uses five-fold cross validation to select the best features; (B) SHAP quantifies the importance of features; (C) sort the importance of features. XGBoost, eXtreme gradient boosting; fJWS, fixed joint space width; RFE, recursive feature elimination; SHAP, SHapley Additive exPlanations.

Table 4 Result of reducing data

Predict factors	R ²	RMSE	MAE	ICC
72-month prediction result				
All image results	0.804	0.697	0.48	0.887
Reduced image results				
0- + 48-month	0.803	0.74	0.516	0.846
0- + 36-month	0.797	0.769	0.523	0.839
0- + 24-month	0.78	0.805	0.545	0.822
0- + 12-month	0.776	0.85	0.585	0.805

Results between manually measurement and predicted results when data and images are reduced on the 72-month test set. R², R-square; MAE, mean absolute error; RMSE, root mean squared error; ICC, intraclass correlation coefficient, P<0.01.

measured them, achieving an R² of 0.6086 and a Pearson correlation coefficient of 0.7801 (41), which was lower than our results. We further developed this approach by focusing solely on the joint space, reducing errors in the mask conversion process and thereby improving accuracy, which obtained an R² of 0.804, and an ICC of 0.887.

Many researchers have previously developed predictive

models for KOA. Some have focused on predicting KOA pain. For example, Wang *et al.* (42) analyzed the relationship between baseline knee pain and the onset and progression of KOA, whereas Guan (43) used X-ray images alongside clinical data to predict whether KOA pain would progress, obtaining an AUC accuracy of 0.807. Other researchers have focused on the likelihood and timing of total knee

replacement in KOA patients in the fifth year, obtaining an AUC accuracy of 0.873 (44). Additionally, several studies have investigated the progression and future severity of KOA. Hu *et al.* demonstrated KOA progression through changes in KL grade (20). Most researchers use binary or multivariate classification to determine accuracy through area under the curve/receiver operating characteristic (AUC/ROC) analysis. Ahmed achieved an accuracy of 74.57% on the test set by predicting changes in JSW that may or may not occur in the future through KOA (14), whereas another researcher predicted whether future JSW changes would exceed a fixed threshold, achieving an ROC accuracy of 0.7 in internal data and 0.6 in external data (15). Cheung *et al.* (41) used XGboost to predict KOA progress, achieving a best AUC accuracy of 0.609. In our study, we conducted a more detailed analysis of changes in fJSW at multiple points in the knee joint, providing comprehensive and direct results for researchers and clinicians and get a high accuracy with ICC test of 0.887. However, by quantifying the calculation results rather than simply classifying them, the prediction error can be reduced and the accuracy can be higher. This approach can also facilitate analysis of KL grade or fJSW progression easily, offering nuanced insights into KOA development.

Currently, most KOA prediction models are based on ML. Some researchers have used logistic regression for analysis (15). Alexos (45) tried six ML techniques—K-nearest neighbors, SVM, decision trees, XGBoost, Naïve Bayes, and RF—for comparison, finding that RF performed best among the six. Recently, it has been found that RNN, especially LSTM, is more efficient in handling time-series data. Wang *et al.* (33) compared LSTM, RF, SVM, and Naïve Bayes, finding that LSTM and RF had the best accuracy with an AUC accuracy >0.8. This result was also confirmed in kidney transplant survival rate models (46), where LSTM outperformed Cox regression and RF models (0.661 *vs.* 0.644, 0.646). However, in our attempt with six models, the results differed, with XGboost performing the best, followed by LSTM; researchers had not made the same comparisons in previous studies.

There are some limitations to this study. Firstly, since we only used the OAI dataset, we are unsure if our model performs equally well on all data, posing a generalization issue that requires validation with more datasets. Secondly, we trained the segmentation model with unified parameters without separately tuning each model's parameters, which we believe could improve accuracy with parameter adjustments. Furthermore, our measurement

model currently only measures joint space and does not analyze other important indicators in knee X-ray, such as the femoral–tibial angle and the presence of osteophytes. Adding these indicators might improve the prediction model's results, which needs further analysis.

Conclusions

We have built a new automatic joint space measurement model with a high accuracy compared to previous methods. The use of numerical fJSW as a predictor provides a new way to predict the development of KOA in patients. The optimal combination prediction model by combining TransUNet and XGboost has been developed. By quantifying the calculation results, the prediction error can be reduced and the accuracy can be higher. In addition, under this model, the prediction of fJSW in the sixth year can also achieve a high accuracy through X-ray images of the knee joint only for 2 years, which is of great clinical significance.

Acknowledgments

None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD+AI reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-24-1397/rc>

Funding: This study was supported by the National Orthopedic and Exercise Rehabilitation Clinical Medical Research Center, China (No. 2021-NCRC-CXJJ-ZH-11).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-1397/coif>). All authors report that this study was supported by the National Orthopedic and Exercise Rehabilitation Clinical Medical Research Center, China (No. 2021-NCRC-CXJJ-ZH-11). The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was

conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Silverwood V, Blagojevic-Bucknall M, Jinks C, Jordan JL, Protheroe J, Jordan KP. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. *Osteoarthritis Cartilage* 2015;23:507-15.
2. Lespasio MJ, Piuze NS, Husni ME, Muschler GF, Guarino A, Mont MA. Knee Osteoarthritis: A Primer. *Perm J* 2017;21:16-183.
3. Vongsirinavarat M, Nilmar P, Somprasong S, Apinonkul B. Identification of knee osteoarthritis disability phenotypes regarding activity limitation: a cluster analysis. *BMC Musculoskelet Disord* 2020;21:237.
4. Minciullo L, Parkes MJ, Felson DT, Cootes TF. Comparing image analysis approaches versus expert readers: the relation of knee radiograph features to knee pain. *Ann Rheum Dis* 2018;77:1606-9.
5. Conaghan PG, Hunter DJ, Maillefert JF, Reichmann WM, Losina E. Summary and recommendations of the OARSI FDA osteoarthritis Assessment of Structural Change Working Group. *Osteoarthritis Cartilage* 2011;19:606-10.
6. Bartlett SJ, Ling SM, Mayo NE, Scott SC, Bingham CO 3rd. Identifying common trajectories of joint space narrowing over two years in knee osteoarthritis. *Arthritis Care Res (Hoboken)* 2011;63:1722-8.
7. Marijnissen AC, Vincken KL, Vos PA, Saris DB, Viergever MA, Bijlsma JW, Bartels LW, Lafeber FP. Knee Images Digital Analysis (KIDA): a novel method to quantify individual radiographic features of knee osteoarthritis in detail. *Osteoarthritis Cartilage* 2008;16:234-43.
8. Neumann G, Hunter D, Nevitt M, Chibnik LB, Kwok K, Chen H, Harris T, Satterfield S, Duryea J. Location specific radiographic joint space width for osteoarthritis progression. *Osteoarthritis Cartilage* 2009;17:761-5.
9. Duryea J, Zaim S, Genant HK. New radiographic-based surrogate outcome measures for osteoarthritis of the knee. *Osteoarthritis Cartilage* 2003;11:102-10.
10. Oka H, Muraki S, Akune T, Mabuchi A, Suzuki T, Yoshida H, Yamamoto S, Nakamura K, Yoshimura N, Kawaguchi H. Fully automatic quantification of knee osteoarthritis severity on plain radiographs. *Osteoarthritis Cartilage* 2008;16:1300-6.
11. Goodfellow IJ, Courville A, Bengio Y. Scaling up spike-and-slab models for unsupervised feature learning. *IEEE Trans Pattern Anal Mach Intell* 2013;35:1902-14.
12. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 2021;8:53.
13. Li W, Xiao Z, Liu J, Feng J, Zhu D, Liao J, Yu W, Qian B, Chen X, Fang Y, Li S. Deep learning-assisted knee osteoarthritis automatic grading on plain radiographs: the value of multiview X-ray images and prior knowledge. *Quant Imaging Med Surg* 2023;13:3587-601.
14. Ahmed SM, Mstafa RJ. Identifying Severity Grading of Knee Osteoarthritis from X-ray Images Using an Efficient Mixture of Deep Learning and Machine Learning Models. *Diagnostics (Basel)* 2022.
15. Zhang W, McWilliams DF, Ingham SL, Doherty SA, Muthuri S, Muir KR, Doherty M. Nottingham knee osteoarthritis risk prediction models. *Ann Rheum Dis* 2011;70:1599-604.
16. Heidari B. Knee osteoarthritis prevalence, risk factors, pathogenesis and features: Part I. *Caspian J Intern Med* 2011;2:205-12.
17. Peterson JA, Meng L, Rani A, Sinha P, Johnson AJ, Huo Z, Foster TC, Fillingim RB, Cruz-Almeida Y. Epigenetic aging, knee pain and physical performance in community-dwelling middle-to-older age adults. *Exp Gerontol* 2022;166:111861.
18. Tong B, Chen H, Wang C, Zeng W, Li D, Liu P, Liu M, Jin X, Shang S. Clinical prediction models for knee pain in patients with knee osteoarthritis: a systematic review. *Skeletal Radiol* 2024;53:1045-59.
19. Leung K, Zhang B, Tan J, Shen Y, Geras KJ, Babb JS, Cho K, Chang G, Deniz CM. Prediction of Total Knee Replacement and Diagnosis of Osteoarthritis by Using Deep Learning on Knee Radiographs: Data from the Osteoarthritis Initiative. *Radiology* 2020;296:584-93.
20. Hu J, Zheng C, Yu Q, Zhong L, Yu K, Chen Y, Wang Z, Zhang B, Dou Q, Zhang X. DeepKOA: a deep-learning model for predicting progression in knee osteoarthritis using multimodal magnetic resonance images from

- the osteoarthritis initiative. *Quant Imaging Med Surg* 2023;13:4852-66.
21. Liu Q, Chu H, LaValley MP, Hunter DJ, Zhang H, Tao L, Zhan S, Lin J, Zhang Y. Prediction models for the risk of total knee replacement: development and validation using data from multicentre cohort studies. *Lancet Rheumatol* 2022;4:e125-34.
 22. Nevitt M, Felson D, Lester G. The Osteoarthritis Initiative: Protocol for the cohort study. 2006; 1-74. Available online: <https://oai.epi-ucsf.org/datarelease/docs/StudyDesignProtocol.pdf>. Accessed April 5, 2015.
 23. Peterfy C, Li J, Zaim S, Duryea J, Lynch J, Miaux Y, Yu W, Genant HK. Comparison of fixed-flexion positioning with fluoroscopic semi-flexed positioning for quantifying radiographic joint-space width in the knee: test-retest reproducibility. *Skeletal Radiol* 2003;32:128-32.
 24. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans Pattern Anal Mach Intell* 2017;39:640-51.
 25. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation 6M. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 234-241.
 26. Xiao X, Lian S, Luo Z, Li S. Weighted Res-UNet for High-Quality Retina Vessel Segmentation. 2018 9th *International Conference on Information Technology in Medicine and Education (ITME)*. Hangzhou, China: 2018;327-331.
 27. Ibtihaz N, Rahman MS. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Netw* 2020;121:74-87.
 28. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation. *ArXiv, abs/1706.05587* (2017).
 29. Guo J, Yan P, Qin Y, Liu M, Ma Y, Li J, Wang R, Luo H, Lv S. Automated measurement and grading of knee cartilage thickness: a deep learning-based approach. *Front Med (Lausanne)* 2024;11:1337993.
 30. Zhou Q, Zhou H, Li T. Cost-sensitive feature selection using random forest: Selecting low-cost subsets of informative features. *Knowl Based Syst* 2016;95:1-11.
 31. Ke G, Qi M, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* 2017;3149-57.
 32. Kuch J, Chakraborty S, Tang H, Luo R, Song J, Sabharwal A, Ermon S. Belief Propagation Neural Networks. *arXiv:2007.00295* (2007).
 33. Wang Y, You L, Chyr J, Lan L, Zhao W, Zhou Y, Xu H, Noble P, Zhou X. Causal Discovery in Radiographic Markers of Knee Osteoarthritis and Prediction for Knee Osteoarthritis Severity With Attention-Long Short-Term Memory. *Front Public Health* 2020;8:604654.
 34. Jernelv IL, Hjelme DR, Matsuura Y, Aksnes A. Convolutional neural networks for classification and regression analysis of one-dimensional spectral data. *ArXiv abs/2005.07530* (2020).
 35. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016;785-94.
 36. Idris NE, Ismail MA, Jaya MIM, Ibrahim AO, Abulfaraj AW, Binzagr F. Stacking with Recursive Feature Elimination-Isolation Forest for classification of diabetes mellitus. *PLoS One* 2024;19:e0302595.
 37. Duryea J, Li J, Peterfy CG, Gordon C, Genant HK. Trainable rule-based algorithm for the measurement of joint space width in digital radiographic images of the knee. *Med Phys* 2000;27:580-91.
 38. Misra D, Fielding RA, Felson DT, Niu J, Brown C, Nevitt M, Lewis CE, Torner J, Neogi T; MOST study. Risk of Knee Osteoarthritis With Obesity, Sarcopenic Obesity, and Sarcopenia. *Arthritis Rheumatol* 2019;71:232-7.
 39. Komatsu D, Hasegawa Y, Kojima T, Seki T, Ikeuchi K, Takegami Y, Amano T, Higuchi Y, Kasai T, Ishiguro N. Validity of radiographic assessment of the knee joint space using automatic image analysis. *Mod Rheumatol* 2016;26:761-6.
 40. Flynn BI, Javan EM, Lin E, Trutner Z, Koenig K, Anighoro KO, Kun E, Gupta A, Singh T, Jayakumar P, Narasimhan VM. Deep learning based phenotyping of medical images improves power for gene discovery of complex disease. *NPJ Digit Med* 2023;6:155.
 41. Cheung JC, Tam AY, Chan LC, Chan PK, Wen C. Superiority of Multiple-Joint Space Width over Minimum-Joint Space Width Approach in the Machine Learning for Radiographic Severity and Knee Osteoarthritis Progression. *Biology (Basel)* 2021.
 42. Wang Y, Teichtahl AJ, Abram F, Hussain SM, Pelletier JP, Cicuttini FM, Martel-Pelletier J. Knee pain as a predictor of structural progression over 4 years: data from the Osteoarthritis Initiative, a prospective cohort study. *Arthritis Res Ther* 2018;20:250.
 43. Guan B, Liu F, Mizaian AH, Demehri S, Samsonov A,

- Guermazi A, Kijowski R. Deep learning approach to predict pain progression in knee osteoarthritis. *Skeletal Radiol* 2022;51:363-73.
44. Mahmoud K, Alagha MA, Nowinka Z, Jones G. Predicting total knee replacement at 2 and 5 years in osteoarthritis patients using machine learning. *BMJ Surg Interv Health Technol* 2023;5:e000141.
45. Alexos A, Kokkotis C, Moustakidis S, Papageorgiou E, Tsaopoulos D. Prediction of pain in knee osteoarthritis patients using machine learning: data from osteoarthritis initiative. In: 2020 11th International Conference on Information, Intelligence, Systems and Applications IISA. Piraeus, Greece; 2020.
46. Paquette FX, Ghassemi A, Bukhtiyarova O, Cisse M, Gagnon N, Della Vecchia A, Rabearivelo HA, Loudiyi Y. Machine Learning Support for Decision-Making in Kidney Transplantation: Step-by-step Development of a Technological Solution. *JMIR Med Inform* 2022;10:e34554.

Cite this article as: Guo J, Yan P, Luo H, Ma Y, Jiang Y, Ju C, Chen W, Liu M, Lv S, Qin Y. Predicting joint space changes in knee osteoarthritis over 6 years: a combined model of TransUNet and XGBoost. *Quant Imaging Med Surg* 2025;15(2):1396-1410. doi: 10.21037/qims-24-1397