ABSTRACT

Objective To build a supervised machine learning-based classifier, which can accurately predict whether Tai Chi practitioners may experience knee pain after years of exercise.

Design A prospective approach was used. Data were collected using face-to-face through a self-designed questionnaire.

Setting Single centre in Shanghai, China.

Participants A total of 1750 Tai Chi practitioners with a course of Tai Chi exercise over 5 years were randomly selected.

Measures All participants were measured by a questionnaire survey including personal information, Tai Chi exercise pattern and Irrgang Knee Outcome Survey Activities of Daily Living Scale. The validity of the questionnaire was analysed by logical analysis and test, and the reliability of this questionnaire was mainly tested by a re-test method. Dataset 1 was established by whether the participant had knee pain, and dataset 2 by whether the participant's knee pain affected daily living function. Then both datasets were randomly assigned to a training and validating dataset and a test dataset in a ratio of 7:3. Six machine learning algorithms were selected and trained by our dataset. The area under the receiver operating characteristic curve was used to evaluate the performance of the trained models, which determined the best prediction model.

Results A total of 1703 practitioners completed the questionnaire and 47 were eliminated for lack of information. The total reliability of the scale is 0.94 and the KMO (Kaiser-Meyer-Olkin measure of sampling adequacy) value of the scale validity was 0.949 (>0.7). The CatBoost algorithm-based machine-learning model achieved the best predictive performance in distinguishing practitioners with different degrees of knee pain after Tai Chi practice. ‘Having knee pain before Tai Chi practice’, ‘knee joint warm-up’ and ‘duration of each exercise’ are the top three factors associated with pain after Tai Chi exercise in the model. ‘Having knee pain before Tai Chi practice’, ‘Having Instructor’ and ‘Duration of each exercise’ were most relevant to whether pain interfered with daily life in the model.

Conclusion CatBoost-based machine learning classifier accurately predicts knee pain symptoms after practicing Tai Chi. This study provides an essential reference for practicing Tai Chi scientifically to avoid knee pain.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The included factors covered most aspects of Tai Chi practice, and their importance in the model was analysed by the SHAPley Additive exPlanations method.
⇒ The Activities of Daily Living Scale has good reliability and validity in evaluating knee function.
⇒ There was no further age stratification, which may affect the results.
⇒ The questionnaire is non-prospective, which may be vulnerable to memory bias.
⇒ The prediction model may be biased when applied to men, as most participants were women.

INTRODUCTION

Tai Chi has become popular worldwide in recent years as a branch of traditional Chinese exercises.1 2 It includes several slow and gentle movements such as knee bending, body rotation and arm swinging. Therefore its safety has been well evaluated.3 4 Recent studies indicate that Tai Chi improves cardio-pulmonary function5 6 and prevents elders from falling.7 8 In patients with knee osteoarthritis (KOA), Tai Chi effectively enhances their strength, balance9 10 and proprioception,11 thereby alleviating knee joint stiffness and pain.12

A growing body of research on the biomechanics of Tai Chi movements suggests that Tai Chi could improve knee extension strength and balance function, which was more pronounced after more than 3 weeks of exercise.13 In addition, Tai Chi improves the knee joint’s motor function by changing the way of muscle force recruitment.14 From the perspective of gait, Tai Chi enhances gait velocity, step length, initial contact angle and maximal angle of flexed knees.15 It activates both active and antagonistic muscles and enhances coordination and control, resulting in better ankle and knee joint proprioception.16 Moreover, Tai Chi attenuates
bone mineral density loss and delays the process of joint degeneration.16

Although Tai Chi improves physical functions, a recent study has shown that Tai Chi exercise may increase the risk of knee pain.17 Zhu et al suggested that Tai Chi practice may accompany knee pain related to patellofemoral syndrome.18 A study found that Tai Chi practitioners used relatively higher knee muscle activation patterns with greater co-contraction during Tai Chi exercise than normal walking, which may reduce sports injuries but bring knee pain.19 Wang et al found that incorrect posture is one of the main factors that directly cause knee injury.20 It is considered that several Tai Chi movements are consisted with lunges and pushdowns in which significant peak knee flexion and extension torque can be observed. Thus, practitioners with weak muscle strength in the lower limb who cannot load excessive torque could be prone to injury and pain during Tai Chi exercise.21

Although Tai Chi is popular as an exercise for improving knee function, knee joint pain may also be closely related to a lack of warm-up, wrong movements and knee joint twisting during Tai Chi practice. There is little research on which factors of Tai Chi practice are related to knee pain. It is important to find out what kind of practitioners are likely to have knee pain after years of practice Tai Chi and how severe the knee pain is. If we can use models to solve this problem, more professional advice can be provided for specific practitioners. Machine learning can build models based on past data, make predictions and apply them to the clinic. Some scholars established machine learning-based patient-specific prediction models for KOA to identify and classify KOA patients, which may improve clinical decision-making and precision medicine.22-24 Bansal et al used surface electromyography data to develop a technique for detecting knee movement sustainability through machine learning.25 Tiulpin et al built a model for predicting total knee arthroplasty from ultrasonography through machine learning.26 Liu et al established a prediction model for knee pain in middle-aged and older people.27 However, the current prediction models for knee joints mainly focus on the gait characteristics or demographic characteristics for predicting the risk factors of KOA or knee joint-related surgery, while few relate to sports risk.

Thus, this research aims to establish machine learning-based models to predict whether Tai Chi practitioners will experience knee joint pain after years of exercise and whether knee pain will affect their activities in daily life. In addition, we used the SHAPley Additive exPlanations (SHAP) method to identify relevant influencing factors for knee pain in the model. Our results can provide scientific guidance for Tai Chi practitioners to avoid the risk of knee pain in the future.

**METHODS**

**Participants**

This study was conducted in cooperation with the Shanghai Tai Chi Association. In this study, we conducted a questionnaire survey of Tai Chi practitioners in Shanghai, China, between April 2015 and December 2016. All participants who had practiced Tai Chi for over 5 years were randomly selected, with the native language of Mandarin that could understand the content of the questionnaire and complete the questionnaire with the help of the investigators.

**Patient and public involvement**

Patients were not involved in developing the research question, study design or selection of outcome measures.

**Questionnaire**

The questionnaire was designed under the guidance of professors from the School of Acupuncture-Moxibustion and Tuina, Shanghai University of Traditional Chinese Medicine. It consisted of two parts: basic information and knee function evaluation. The first part collected the personal information of the participants and their situation in practicing Tai Chi, mainly including the following: participant demographics (gender, age, height, body weight, body mass index (BMI)), years of practice, instructor, exercise frequency, duration of exercise, Tai Chi stance, knee joint warm-up, and previous knee pain. ‘The instructor’ in the questionnaire refers to whether the practitioners have had a professional Tai Chi instructor before. Previous knee pain was to confirm knee pain before practicing Tai Chi. In the second part, the knee joint function of the participants was assessed by the Irrgang Activities of Daily Living Scale, which refers to the questionnaire proposed by Irrgang et al in 1998, including 17 questions, with a total score of 100 points. The questionnaire should be completed within 15 min. The scale consists of two parts: assessing knee symptoms and the participant’s daily functional limitations. The assessment of daily living function can also reflect the symptoms of the practitioner’s knee joint. The higher the score in this questionnaire, the fewer knee symptoms and the better joint function the practitioner has. Conversely, a lower score indicates that knee symptoms are more severe and by which daily life is seriously affected. Online supplemental files 1 and 2 show the English and Chinese versions of the questionnaire, respectively.

**Procedure**

The subjects of this research were Tai Chi practitioners in some urban areas of Shanghai, including Hongkou District and Minhang District. Before the formal investigation, we trained the investigators to standardize the investigation methods. Then, a preinvestigation was conducted to uncover the problems existing in the questionnaire and the process. We distributed questionnaires in parks and squares where Tai Chi practitioners gather. The questionnaire is issued in paper form and must be filled out and returned on the spot. Questionnaires were issued and collected in the morning as most Tai Chi practitioners tended to practice in the morning. In addition, considering that some elders may need help understanding the questions in the questionnaire, the investigators explained them in detail during the
investigation process. Afterward, the answers to the questionnaire were filled into a file by recorders.

**Machine learning-based modelling**

In this work, we used machine learning technologies to build binary predictive classifiers, distinguishing the following: (1) whether the Tai Chi practitioner has knee pain after years of exercise; (2) if so, whether the practitioner’s knee pain affects daily life. There are four critical steps in our machine learning-based modelling process: data preprocessing, feature selection, algorithm selection and parameter tuning. Finally, we evaluate the resulting models’ prediction performance and choose the best classifier. The workflow is shown in figure 1. The method was also described in our previous study. Scikit-learn, a widely used Python-based machine learning library, was used to train our predictive models (refer to the website: https://scikit-learn.org/stable/).

**Data preprocessing**

For each dataset, we randomly split the entire dataset into a training and validation dataset (70%) and a test dataset (30%). The training and validation dataset was used to train and validate the prediction model, while the test dataset was applied to evaluate the predictive performance of the trained model. A fivefold cross-validation method was used to prepare the training and validation dataset. Fivefold cross-validation can prevent the model from overfitting and is a classic method to evaluate the generalization ability of the training dataset. The training and validation dataset was divided into five subsets with equal sizes at random. One subset was applied to evaluate the model, and the other four subsets were used for training. The cross-validation process was repeated five times, with each of the five subsets used once for validation.

**Feature selection**

We designed many features to reflect the different outcomes of practitioners after years of Tai Chi exercise. Some features use continuous variables: age, height, body weight, BMI and years of practice. The other features with categorical variables were manually selected: gender (male/female), instructor (negative/positive), exercise frequency (every day/5–6 times per week/1–4 times per week/irregular), duration of exercise (<0.5 hours/0.5–1 hour/1–2 hours/>2 hours), Tai Chi stance (high/medium/low/unclear), knee joint warm-up (no warm-up/5 min/10 min/15 min) and previous knee pain (no/occasionally/often). Table 1 shows the statistical details of selected features.

**Algorithm selection and parameter tuning**

To obtain the best prediction model, we selected six representative machine learning algorithms, including Decision Tree, SVM, Logistic Regression, Random Forest, XGBoost and CatBoost. The workflow shown in figure 1.

---

**Figure 1** The flow-chart of data processing and machine learning based modelling. The original dataset recruited 1750 Tai Chi practitioners. By removing incomplete data entry cases, 1703 Tai Chi practitioners were finally involved in the dataset. Six hundred and forty-three practitioners reported having no knee pain after years of exercise, while 1060 participants had knee pain (dataset 1). The 1060 practitioners were further divided into two groups. Knee pain does not affect daily life group (902) and knee pain affects daily life group (158) (dataset 2). Each dataset was split into a training and validation dataset (70%) and a test dataset (30%). Six different machine learning algorithms were selected for training based on the training and validation dataset. Predictive models were obtained after parameter tuning. The final classifier was determined based on the comparison of the prediction performance of each model. SHAP methods analysed feature ranking in the selected classifier. SHAP, SHApley Additive exPlanations.
Forest, XGBoost and CatBoost for training, and compared their predictive performance. We need to determine an optimal set of parameters for a selected algorithm. Grid search is applied to go through the parameter space based on the training and validation dataset. We select a finite set for values of each parameter to form the parameter space. Grid search iterates through each parameter combination, for which we evaluate the prediction performance. Finally, the parameters leading to the maximum area under the receiver operating characteristic (ROC) curve (AUC) value are recorded based on the training and validation set.

**Model evaluation**

AUC value was used to evaluate the performance of the trained models. AUC represents the probability that the model ranks higher in random positive instances than randomly selected negative instances. The value of AUC is between 0 and 1; the higher the AUC value, the better the model distinguishes practitioners with different knee pain outcomes. The final classifier was determined based on the comparison of the predictive performance of each model.

**Statistical analysis**

Statistical analysis was performed using Python programming software. The numerical variables were represented as mean±SD (Table 1); categorical variables were described in numbers and percentages. Independent Student’s t-test was used to compare the means of the continuous variables with normal distribution; Welch’s t-test was used if the data were not normally distributed; $\chi^2$ test was used to compare categorical variables. A p-value less than 0.05 indicates statistically significant. To make the prediction model interpretable, we used the SHAP method to quantify the importance of each feature in the prediction model. SHAP is a representative method to explain the predictions of supervised machine learning-based classifiers. We applied this method to quantify the importance of each feature to the categories of the practitioners (with or without knee pain; knee pain affecting or not affecting daily life). For each feature, we used the metric mean (|SHAP value|), the average value of the absolute values of the SHAP values of all the practitioners, to obtain the value of the feature importance. When the |SHAP value| is higher, the feature contributes more to the prediction model.

**RESULTS**

**Tai Chi practitioners**

A total of 1750 practitioners filled out the questionnaire and 47 with incomplete information were excluded from the analysis. As a result, 1703 practitioners who had practiced Tai Chi for over 5 years were included for further analysis. 1433 (85.14%) were female, and 270 (15.85%) were male. Participants were aged 26–94, with an average age of 68.58 years (SD=8.72 years). The mean Tai Chi practicing period was 12.54 years (SD=6.33 years). There were 639 practitioners (37.32%) who had never experienced knee pain before practicing Tai Chi, 836 practitioners (49.09%) had knee pain occasionally, and 228 participants (13.39%) often had knee pain. The database comprises 12 features belonging to two categories: five

<table>
<thead>
<tr>
<th>Feature</th>
<th>n (%)/(mean±SD)</th>
<th>P value (1)</th>
<th>P value (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>0.961 0.817</td>
<td>0.961 0.817</td>
<td>0.961 0.817</td>
</tr>
<tr>
<td>Female</td>
<td>1433 (85.14%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>270 (15.85%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>68.58±8.72</td>
<td>0.370 0.142</td>
<td>0.370 0.142</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>162.66±37.04</td>
<td>0.258 0.157</td>
<td>0.258 0.157</td>
</tr>
<tr>
<td>Body weight (kg)</td>
<td>61.07±8.99</td>
<td>0.851 0.105</td>
<td>0.851 0.105</td>
</tr>
<tr>
<td>Body mass index (BMI)</td>
<td>23.26±2.68</td>
<td>0.345 0.377</td>
<td>0.345 0.377</td>
</tr>
<tr>
<td>Years of Tai Chi practice</td>
<td>12.54±6.33</td>
<td>0.087 0.546</td>
<td>0.087 0.546</td>
</tr>
<tr>
<td>Have instructor?</td>
<td>0.030* 0.704</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1163 (68.29%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>540 (31.71%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercise frequency</td>
<td>0.333 0.569</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Every day</td>
<td>678 (39.81%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5–6 times/week</td>
<td>361 (21.20%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–4 times/week</td>
<td>277 (16.27%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irregular</td>
<td>387 (22.72%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of each exercise</td>
<td>0.307 0.159</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;0.5 hours</td>
<td>226 (13.27%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5–1 hour</td>
<td>853 (50.09%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–2 hours</td>
<td>558 (32.76%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;2 hours</td>
<td>66 (3.88%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tai Chi stance</td>
<td>0.047* 0.809</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>404 (23.72%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>709 (41.63%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>410 (24.08%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unclear</td>
<td>180 (10.57%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knee joint warm-up</td>
<td>0.001*** 0.599</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No warm-up</td>
<td>445 (26.13%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 min</td>
<td>737 (43.28%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 min</td>
<td>207 (12.16%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 min</td>
<td>314 (18.44%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have knee pain before Tai Chi practice?</td>
<td>&lt;0.001*** &lt;0.001***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>639 (37.52%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occasionally</td>
<td>836 (49.09%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Often</td>
<td>228 (13.39%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P value (1): with/without knee pain group; p value (2): does/not affect daily life group.

*p<0.05 was considered statistically significant; **p<0.01; ***p<0.001.
from demographic features and seven from Tai Chi practice features. The number of practitioners in each feature is shown in table 1.

Reliability and validity of the questionnaire
The validity of the questionnaire was analysed by logical analysis and tests. The score of each dimension of the questionnaire is greater than 0.7, and the total reliability of the questionnaire is 0.94, indicating that it has good reliability and meets the requirements of the questionnaire analysis. The internal consistency reliability of the total questionnaire and each factor score was 0.940 and 0.899–0.907, respectively. The score of each factor is related to the total questionnaire, suggesting that the questionnaire has good structural validity.

The reliability of the questionnaire was mainly tested by the re-test method. Twenty Tai Chi practitioners in Shanghai were randomly selected for re-test, and the interval was 4 weeks from the last test. The reliability coefficient of the test is R=0.87 (p<0.01). The KMO (Kaiser-Meyer-Olkin measure of sampling adequacy) value of the questionnaire validity was 0.949 (>0.7), indicating that the structural validity of the questionnaire is satisfactory.

Use machine learning to predict knee pain after Tai Chi exercise
In our dataset, 643 practitioners (37.76%) had no knee pain after practicing Tai Chi for years (non-knee pain group), while 1060 practitioners (62.24%) experienced knee pain (knee pain group) (dataset 1). In the knee pain group, 902 practitioners’ knee pain did not interfere with their daily life (52.96%). In comparison, 158 practitioners’ knee pain affected their daily life (9.28%) (knee pain affects daily life group vs. knee pain does not affect daily life group, dataset 2). First, we adopted supervised machine learning algorithms to build a binary classifier that distinguished practitioners who had knee pain or not after Tai Chi exercise. Second, we further built a classifier to predict the severity of knee joint symptoms (knee pain affects daily life or not).

To achieve the first goal, six different machine learning algorithms, namely Decision Tree, SVM, Logistic Regression, Random Forest, XGBoost and CatBoost, were selected for training based on dataset 1 (figure 1). The ROC curve of each trained model in predicting knee pain after years of Tai Chi practice. Six different machine learning algorithms, namely Decision Tree, SVM, Logistic Regression, Random Forest, XGBoost and CatBoost, were selected for training. The area under the curve value for each model was presented in the lower right corner of the graph. ROC, receiver operating characteristic.
and 0.895, respectively. The variance of each model is presented in online supplemental table 1. By comparing the predictive performance of these six models, both XGBoost and CatBoost-based models achieved the highest AUC value of 0.895, indicating the best predictive performance in distinguishing practitioners with or without knee pain after Tai Chi exercise.

There are 12 features included in our dataset to drive the classification model, but the importance of each feature in the model is different. Next, we quantified each feature’s SHAP value in the trained CatBoost-based classifier (figure 3). According to the SHAP value of each feature, ‘Have knee pain before Tai Chi practice’ ranked first among all the analyzed features, thus contributing most to distinguishing practitioners with or without knee pain after Tai Chi exercise.

Next, to visualise the difference between the two groups of participants in dataset 1, we compared the top three features based on the SHAP value in the CatBoost-based classifier. We found that if practitioners did not have knee pain initially, they experienced less knee pain after practicing Tai Chi. However, if they had occasional or frequent knee pain before practicing Tai Chi, knee pain after practice increases. Regarding knee warm-up, warming up for more than 10 min before practicing Tai Chi is less likely to cause knee pain. Likewise, people who exercised for more than an hour were less probably to have knee pain (figure 4).

Use machine learning to predict further the impact of knee pain on participants’ daily life

Clinically, we are more concerned about whether practitioners’ knee pain affects daily life. If it affects practitioners’ daily life, they need to adjust their practice.

---

Figure 3  SHAP summary plot of the CatBoost-based classifier in distinguishing whether Tai Chi practitioners have knee pain. The relative importance for each feature in the CatBoost-based classifier is obtained by taking the average absolute value of each feature’s SHAP value. SHAP, SHApley Additive exPlanations.

Figure 4  Comparison of the top three features between Tai Chi practitioners with or without knee pain based on graph metrics. (A) Knee pain before Tai Chi exercise; (B) knee joint warm up; (C) the duration of each exercise. Blue column: Tai Chi practitioner without knee pain; orange column: Tai Chi practitioner with knee pain.
strategies or seek the help of doctors. Next, we used dataset 2 to train machine learning algorithms to predict further whether practitioners’ daily lives with knee pain would be affected. Dataset two contains two groups; knee pain affects the daily life group, and knee pain does not affect the daily life group as we described before. According to the AUC values, the CatBoost-based classifier (AUC 0.840) can better predict the severity of knee joint symptoms after practicing Tai Chi. See figure 4 for details of all models. The variance of each model is presented in online supplemental table 2.

Then, we used the lSHAP value to sort the features’ relative importance in the CatBoost-based classifier. The top three features were ‘Have knee pain before Tai Chi practice’, ‘Have Instructor’, and ‘Duration of each exercise’ (figure 5). Our data suggested that the knee joint’s original function significantly influenced the knee function after practicing Tai Chi. Compared with other features, BMI and the height of practitioners contributed less to the classifier.

Finally, we compared the top three features based on the ranking of lSHAP value in the CatBoost-based classifier to distinguish whether participants’ knee pain affects daily life (figure 6). We found that practitioners who initially had knee pain had a more significant impact on their daily life from Tai Chi practice. Compared with ‘Have knee pain before Tai Chi practice’, ‘Have instructor’, and ‘Duration of each exercise’ have relatively more minor impacts on the daily life of practitioners. However, our analysis also showed that participants with an instructor and more than 30 min of practice had less impact on their daily lives (figure 7).

**DISCUSSION**

This study adopted six machine learning algorithms to predict knee pain and the severity of knee pain after practicing Tai Chi for over 5 years. Based on our results, we conclude the following: (1) CatBoost-based machine learning model achieved the highest AUC value in predicting whether Tai Chi practitioners had knee pain; (2) the feature ‘Have knee pain before Tai Chi practice’ has the most significant dependence to the practitioner category (knee pain/never pain); (3) the CatBoost-based model also offered the best prediction of knee pain severity; (4) the feature ‘Have knee pain before Tai Chi practice’ similarly has the most considerable dependence to the practitioner category (knee pain/affects daily life or not). In short, our models accurately predict Tai Chi practitioners’ knee joint conditions. Our study also demonstrated what factors should be considered to protect the knee joint when practicing Tai Chi.
The US Arthritis Foundation recommends Tai Chi for treating osteoarthritis, while it is still controversial whether Tai Chi exercise affects knee joint function. Scholars believe that Tai Chi relieves stiffness and pain, improves the physical function of KOA patients, and to some extent, reduces analgesic intake.\(^{37,38}\) Li et al established a finite element model by a three-dimensional kinematic analysis of Tai Chi movements. They concluded that practicing Tai Chi may reduce the risk to the knee joint compared with walking and jogging as the intra-articular pressure is less concentrated during Tai Chi practice.\(^{39}\) However, some studies found that knee joint pain might have a close relationship with Tai Chi practice as some Tai Chi movements bring a heavy load inside knee joints.\(^{19}\) Therefore, we designed this study to investigate the correlation between Tai Chi and knee pain using machine learning and built a classification model to predict knee pain after Tai Chi practice.

Different Tai Chi practitioners have different knee joint functions due to their specific conditions. For a particular practitioner, conventional methods cannot tell whether knee joint pain and functional changes will occur after exercise Tai Chi. The research method of artificial intelligence can establish a model by computing existing data to predict the situation of new practitioners after practice. By applying six supervised machine learning-based algorithms and comparing the trained models’ prediction performance, our results showed that CatBoost-based models could accurately predict the knee pain symptom (with an AUC value of 0.895) and the severity of knee pain after practicing Tai Chi (with an AUC value of 0.840). CatBoost is an implementation of Gradient

Figure 6  SHAP summary plot of the CatBoost-based classifier in distinguishing whether the daily life of Tai Chi practitioners with knee pain is affected. The relative importance for each feature in the CatBoost-based classifier is obtained by taking the average absolute value of each feature’s SHAP value. SHAP, SHApley Additive exPlanations.
Boosted Decision Trees, which Yandex researchers and engineers developed. It has been applied in various fields due to its excellent performance for classification and regression tasks. Besides, we used automatic parameter tuning to achieve the best set of parameters, ensuring our model can achieve the best prediction performance. For a specific practitioner, by collecting the 12 features, our model can calculate whether the practitioner will have knee pain after practicing Tai Chi; if there is knee pain, our model can also judge whether this knee pain can affect their daily life. Therefore, our model can distinguish different situations of knee joint function in Tai Chi practitioners. Our method can provide valuable reference suggestions for specific practitioners, especially for identifying practitioners with knee pain that affects their daily lives. Clinicians can provide valuable advice or prescriptions, slowing the practitioner’s knee joint injury.

Among the 12 features we have selected, it is essential to identify which are more relevant to the classification of practitioners. We have used the SHAP value to sort the importance of these 12 features. ‘Have knee pain before Tai Chi practice’ ranked first in both classification models, indicating that practitioners who had knee pain before Tai Chi practice were susceptible to knee pain after practice. Knee pain has a more significant impact on the daily life of practitioners. This may be related to joint instability caused by pain. Studies have shown that the ratio of leg muscle mass to the whole body is negatively correlated with pain. The more severe the pain, the weaker the quadriceps muscle strength; exercise will increase the knee joint load and aggravate the pain. Fulton et al. suggested that post-injury changes were present in strength, proprioception, and kinematics, which may have led to overall motor control and function changes. Loss of muscle strength and proprioception increases the risk of sports injuries. However, other studies show that Tai Chi can positively affect knee extensor muscle strength, pain, proprioception, and knee function in patients with partial anterior cruciate ligament injuries. Schmid et al. found that Tai Chi exercise could improve knee proprioception in people with KOA, improving joint stability and function. Shen et al.’s study recovered that Tai Chi improves resting-state functional and structural connectivity between the amygdala and medial prefrontal cortex, which is closely associated with chronic pain. Therefore, further research is needed to explore the influence of post-knee pain on Tai Chi exercise and whether it may enhance the risk of knee joint injury during Tai Chi exercise.

‘Knee joint warm-up’ and ‘Duration of each exercise’ were in the second and third positions in the CatBoost-based classifier in distinguishing whether Tai Chi practitioners have knee pain. Our results suggest that the warm-up and the exercise duration affect the development of knee pain. Figure 4 shows that participants with no warm-up had a high proportion of knee pain, and an introduction of 10–15 min of warm-up can reduce the proportion of knee pain participants. Our result suggested that the shorter the warm-up duration, the higher the risk of knee pain. This result is consistent with previous studies that effective neuro-muscle warm-up strategies could reduce the risk of lower limb injuries during exercise. In the present study, we divided the duration of each exercise into four levels, namely <0.5 hours, 0.5–1 hour, 1–2 hours, and >2 hours. Among them, the duration of each exercise with the highest proportion was 0.5–1 hour (50.09%), and practicing for more than an hour was less likely to experience knee pain. Studies have shown that Tai Chi can help strengthen lower limb muscles and improve balance. A recent systematic review found that most of the included literature on Tai Chi practice duration ranged from 60 to 90 min. All these studies have a good effect on improving elders’ lower limb strength and balance ability. However, we cannot rule out the possibility that participants with short exercise duration are due to pre-existing knee pain. Moreover, our study found that whether a practitioner has an instructor also impacts the occurrence of knee pain and knee joint function. Without the guidance of an instructor, self-taught Tai Chi movements can easily be substandard, resulting in changes in the kinematics and biomechanics of the lower limbs. It may increase knee joint contact load, change muscle activation mode, and ultimately affect lower limb function. Based on the above findings, it is recommended that the warm-up time before practicing should be longer than 10 min, each exercise should be longer than 1 hour, and practicing under the guidance of a professional instructor.

In the next step, we will recruit more Tai Chi practitioners to increase the accuracy of the classifier and build prediction software. The software can automatically predict the outcome of practicing Tai Chi on the knees based on the corresponding indicators, which will help Tai Chi practitioners prevent knee injuries. Therefore, the machine learning-based methods described in this study may become a powerful tool to guide people in practicing Tai Chi scientifically and provide evidence for whether Tai Chi will hurt the knee.

However, our study still has some limitations. First, the practitioners included in the study were all from Shanghai, Tai Chi originated in China, and many Chinese people have learned Tai Chi. Our classifier may not represent all regions of China and other countries. Second, most participants in this study were women (85.14%). Therefore, prediction models may be biased when applied to men. Third, we used a subjective and non-prospective questionnaire and collected those data once with no follow-up, which may be vulnerable to memory bias. In the future, we will include more objective features to quantify conclusions and set up more follow-up time points. Moreover, we focused on knee symptoms rather than a specific disease in this study. In the future, we might pay attention to specific diseases, particularly degenerative joint diseases like KOA. It is expected that our predictive classifier will become more accurate and can be clinically promoted.
CONCLUSION

In this study, we proposed two supervised machine learning-based models to predict whether Tai Chi practitioners would experience knee pain and whether knee pain would affect their daily lives. The CatBoost-based classifiers could predict the outcomes of practitioners well. Furthermore, we use the SHAP method to investigate the feature importance in the models. Having knee pain before Tai Chi practice is the most discriminating factor. In addition, the duration of each exercise, knee joint warm-up, and whether there is an instructor are all crucial factors that affect the practitioner’s knee joint function. Our solution relies only on computers and open-source software and is highly accurate. It can quickly identify whether Tai Chi practitioners have knee pain after exercise. The information provided by our classifier is an essential reference to formulate reasonable exercise recommendations for specific practitioners, thereby preventing Tai Chi practitioners from hurting their knees.

Author affiliations

1Department of Tuina, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China
2Shanghai Key Lab of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai, China
3School of Acupuncture and Tuina, Shanghai University of Traditional Chinese Medicine, Shanghai, China
4Clinical Research Center for Mental Disorders, Shanghai Pudong New Area Mental Health Center, Tongji University School of Medicine, Shanghai, China

Contributors LG participated in the study design and data collection. LY participated in data analysis and model building. HX and XS participated in the study design and data collection. YL and YC contributed to the data analysis. YJ, ZK, WS and FY contributed to the data collection. LG is responsible for the overall content as the guarantor and accepts full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish. All authors contributed to the manuscript writing. All authors have read and approved the final version of the manuscript.

Funding This work was supported by the National Natural Science Foundation of China (Grant Number: 8197151854); Shanghai Key Clinical Specialty Construction Project (Grant Number: Shslczx004001); the Sailing programme of Shanghai Science and Technology Commission (Grant Number: 22YF1444300); Projects within the budget of Shanghai University of Traditional Chinese Medicine (Grant Number: 2021UK091); Clinical Research Project of Shanghai Municipality Health Commission (Grant Number: 202140037); and Open Research Programme of Shanghai Key Lab of Intelligent Information Processing (Grant Number: IIP201911).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was reviewed by the Ethics Committee of Shanghai University of Traditional Chinese Medicine, Yueyang Hospital of Integrated Traditional Chinese Medicine and Western Medicine (No. 2015-034). This was an observational study, and there was no clinical intervention for the participants. Therefore, the study was not registered clinically. All data collected will be used for this study only. The study data were in the form of questionnaires and did not include clinical specimens or human genetic information. All participants were informed and consented before the study. Informed consent includes the study purpose, the content of the study, the security of data and the possible benefits of the study to individuals and society. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

30 Noble WS. What is a support vector machine NAT BIOTECHNOL 2006;24:1565-7.
32 Breiman L. Random forests. MACH LEARN 2001;45:5-32.