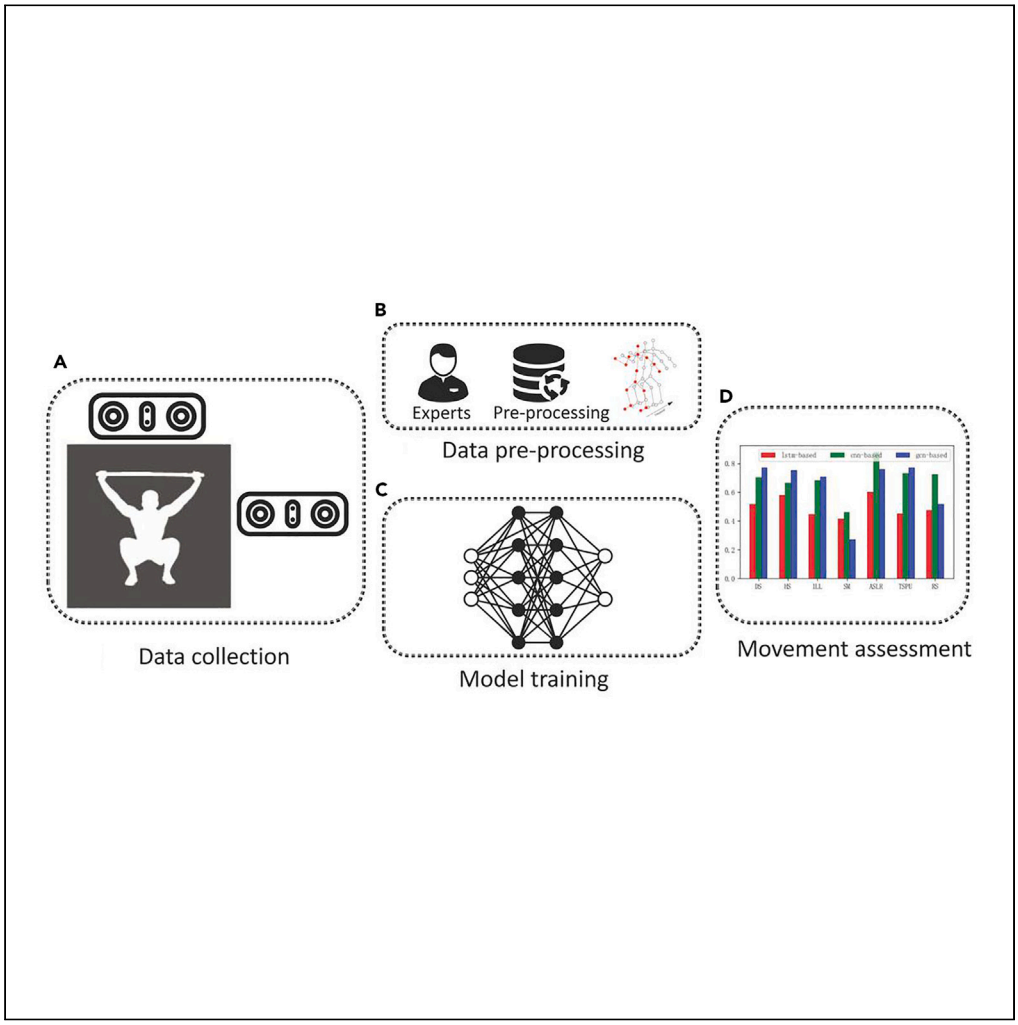
iScience

CellPress
OPEN ACCESS

Article

Markerless vision-based functional movement screening movements evaluation with deep neural networks



Yuan-Yuan Shen,
Qing-Jun Xing,
Yan-Fei Shen

syf@bsu.edu.cn

Highlights

Markerless vision-based sensors are used for computer-aided assessment of FMS tests

Three kinds of deep learning models are designed and compared

Multi-view features are fused to provide better representation learning

Shen et al., iScience 27, 108705
January 19, 2024 © 2023
https://doi.org/10.1016/
j.isci.2023.108705

iScience

CellPress
OPEN ACCESS

Article

Markerless vision-based functional movement screening movements evaluation with deep neural networks

Yuan-Yuan Shen,[1] Qing-Jun Xing,[2] and Yan-Fei Shen[1,3,*]

SUMMARY

The functional movement screen (FMS) test is a seven-test battery used to assess fundamental movement abilities of individuals. It is commonly used to predict sports injuries but relies on clinical expertise and is not suitable for self-examination. This study presents an automatic FMS movement assessment framework using a multi-view deep neural network called MVDNN. The framework combines automatic skeleton extraction with manual feature selection to extract 3D trajectory features of human skeleton joints from two different directions. Three mainstream methods of time-series modeling are then used to learn high-level feature representation from skeleton sequences, and motion features from two views are fused to provide complementary information. Results of public FMS movements dataset demonstrate that our MVDNN outperforms current state-of-the-art methods with an average miF1 score of 0.857, maF1 score of 0.768, and Kappa score of 0.640 over ten runs.

INTRODUCTION

Sports injuries[1,2] have become one of the most common injuries in modern society, and many people who engage in a variety of daily sports are plagued by sports injuries. To identify the risk factors of sports injuries in advance, the functional movement screen (FMS) tests[3,4] have been designed to identify deficits in individuals' movement patterns that may increase the likelihood of sports injuries. Due to its simplicity, efficiency, and low cost, FMS tests[5,6] have been widely used by sport practitioners such as personal trainers, athletic trainers, physical therapists, and strength coaches. For example, the FMS has become one useful tool[7,8] during a pre-participation physical examination to screen athletes for risk of injury in organizations such as the National Hockey League (NHL), the National Football League (NFL), and the United States military for several years.[9] By using the FMS tests, coaches can gain a better understanding of an athlete's ability to participate in physical activity based on their final FMS scores. This can help identify risk for injury in advance and lead to decisions related to interventions for performance enhancement.

At present, the FMS tests rely mainly on human experts. Seven functional movement test tasks including deep squat (DS), hurdle step (HS), inline lunge (ILL), shoulder mobility (SM), active straight leg raise (ASLR), trunk stability pushup (TSPU), and rotary stability (RS) constitute an individual FMS test. In the FMS test, participants perform all seven movements in sequence, and a score of 0–3 is assigned for each task by FMS experts based on FMS scoring criteria. A score of 3 points is awarded for perfect form, a score of 2 points is given for completing the test with compensations, a score of 1 point is awarded for not completing the test accurately, and a score of 0 point is given if the subjects feel any pain during the test. Obviously, the assessment scores are more prone to biases with experts' subjective evaluation. Furthermore, it is inconvenient to offer FMS expert for every single FMS evaluation. Thus, an automatic FMS movement analysis system with advanced computer-assisted techniques becomes a promising idea.

For computer-aided technical FMS analysis system, various sensor data (e.g., motion data, video data, depth data) are being collected during FMS tests. These data are then analyzed to determine the test score of the participants. Although there are several studies that utilize wearable inertial measurement unit (IMU) technology to automate FMS tests,[10,11] IMU-based FMS tests are not friendly to the home-based crowd because of expensive inertial sensors and inconvenient to wear. Moreover, there is an abundance of literature on computer-based human motion analysis, but most of existing studies cannot be directly applied to other application scenarios due to the significant difference of data distribution.

With recent developments in computer vision[12–14] and machine learning,[15–17] markerless vision-based sensors have become increasingly significant in the sport domain.[18–20] These sensors provide a fast and effective way to acquire real-time movement data for analysis. In this paper, we present a home-based FMS analysis system that utilizes markerless vision-based sensors.[21,22] The system is designed to

[1]School of Sport Engineering, Beijing Sport University, Beijing 100084, China
[2]School of Sport Science, Beijing Sport University, Beijing 100084, China
[3]Lead contact
*Correspondence: syf@bsu.edu.cn
https://doi.org/10.1016/j.isci.2023.108705

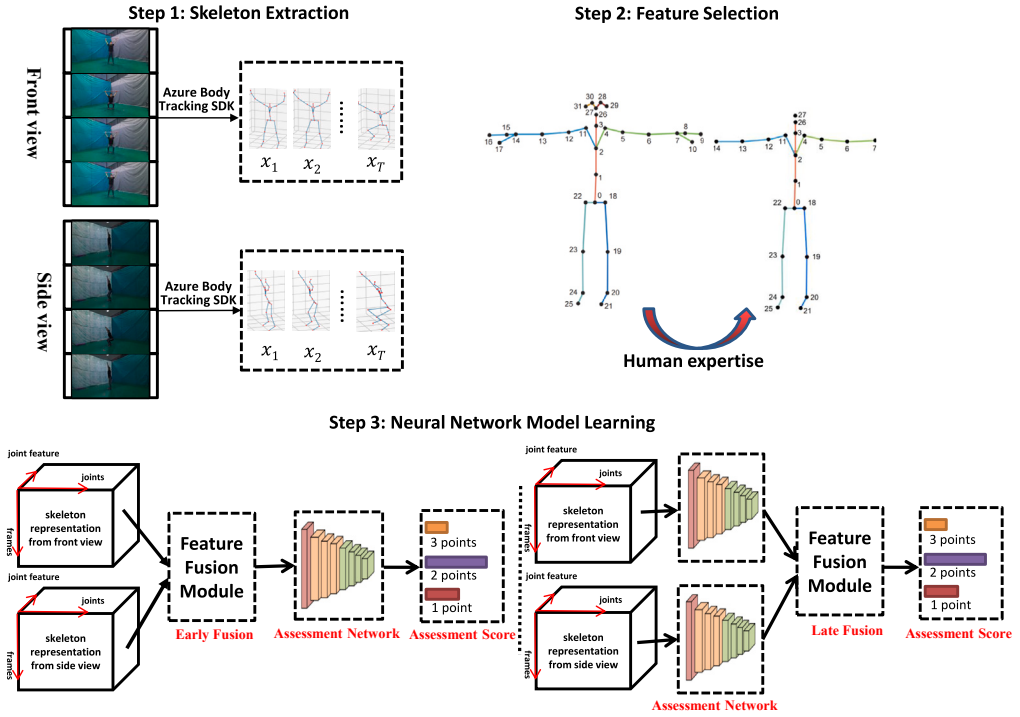


Figure 1. The pipeline of markerless vision-based FMS movement assessment methods

automatically handle FMS movement assessment. To capture the 3D skeleton trajectories, we employ Microsoft Azure Kinect depth sensors, which are capable of accommodating dynamic environments and complex backgrounds. Each person's 3D skeleton trajectory contains 32 body joints. Considering that not all joints extracted from original Azure SDK are informative for FMS test, we select 22 most informative joints for each FMS test movement. In order to improve the representation of motion features, a deep neural network[23–25] is employed to analyze the 3D skeleton data firstly from both the front and side views. The resulting spatiotemporal feature representation is then fused with complementary information from both views to achieve better results. The working principle of the proposed MVDNN method is shown in Figure 1.

To summarize, the main contributions of the paper are three-fold:

1. Markerless vision-based sensors are used for computer-aided assessment of FMS tests;
2. Based on the fine-grained characteristic of FMS test, three typical categories of deep models are designed and compared;
3. Multi-view features from both front view and side view are fused together to provide better representation learning.

RESULTS

Dataset

We used the FMS dataset proposed by Xing et al.,[26] which contains 1,812 action clips from 45 subjects executing all seven FMS tests. The 3D skeletal sequence data for each action clip are obtained using the official Microsoft Kinect SDK. This includes 128-dimensional sequences of joint orientations (four dimensions per joint) and 96-dimensional sequences of joint positions (three dimensions per joint) from two views (i.e., front view and side view). Due to differences in FMS test movements and movement speed across different subjects, the number of frames in an action sequence varies from 34 to 545 frames. In addition, an annotation json file is provided, which contains the FMS test scores for each episode rated by three FMS experts on a scale of 0–3 points. We evaluated the inter-rater reliability among the three experts and presented the final results in Figure S1. Several examples of the movements in FMS test are shown in Figure 2.

The FMS movements assessment task aims to predict an individual's FMS score by analyzing their movement quality in a 3D video clip. Using the same evaluation criteria as the original FMS test, a neural network assigns a score of 1–3 points (0 point cannot be determined solely from visual data) according to the quality of the individual's movements.

Experimental setting

To minimize the impact of intraclass variations on estimated skeleton data, normalization and alignment are performed first. Each experiment is conducted 10 times to help yield stable average results, with 30% of randomly selected used as test set. Model

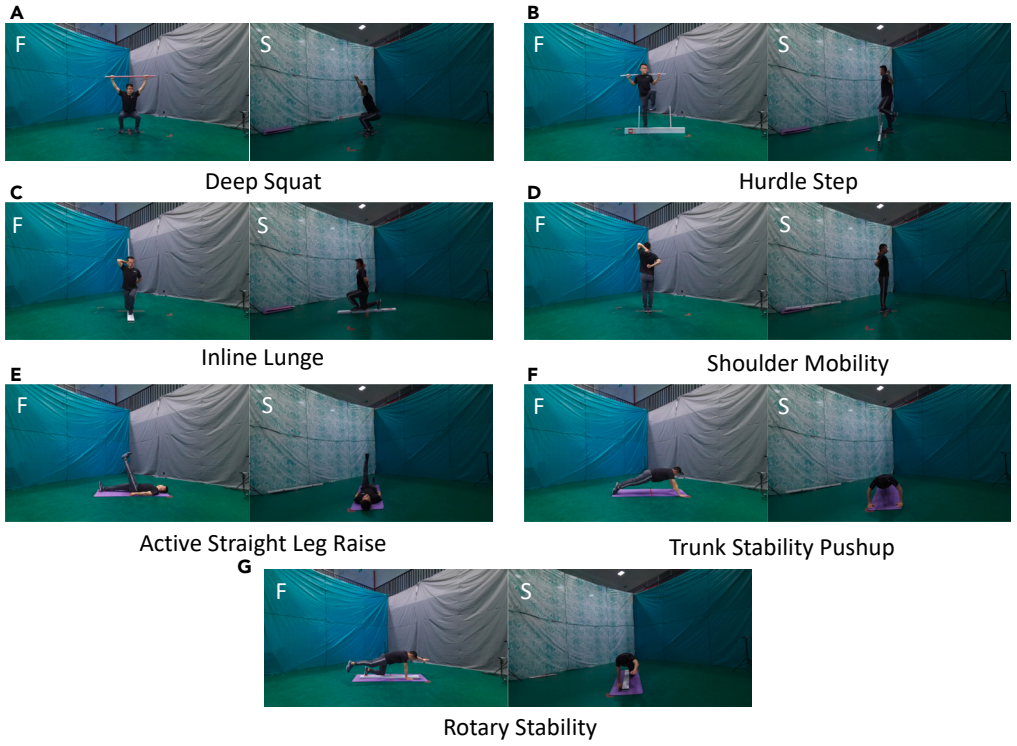


Figure 2. Samples of seven FMS tests

Each FMS test movement contains a pair of images from the front view and the side view.

parameters are trained using the Adam optimizer with a momentum of 0.9 and a learning rate of 0.001. In order to guarantee convergence, we adopt different training epochs for each kind of network structure. That is, 50 epochs for CNN-based and LSTM-based models, and 500 epochs for GCN-based model. The batch size for all experiments is set to 64. The structures of the CNN-base model, LSTM-based model, and GCN-based model that we utilized are presented in Figures 3, 4, and 5, respectively. The feature fusion method we employed is illustrated in Figure 6.

Performance evaluation

We employ the F1 measure, kappa statistic and confusion matrix to evaluate the performance of our model performance for scoring FMS movements. To account for the limitations of F1 score in representing overall performance, we calculate the micro-averaged F1(abbr. mi$F1$) and the macro-averaged F1(abbr. ma$F1$). Assume the true positive, false positive, false negative and true negative for the j-th class are tp_j, fp_j, fn_j and tn_j, respectively, then the mi$F1$ and the ma$F1$ can be computed as follows:

$$\mathrm{mi}F1 = \frac{2\,PR}{P+R},$$

$$P = \frac{\sum_{j=1}^{C} tp_j}{\sum_{j=1}^{C}\left(tp_j + fp_j\right)}, \qquad \text{(Equation 1)}$$

$$R = \frac{\sum_{j=1}^{C} tp_j}{\sum_{j=1}^{C}\left(tp_j + fn_j\right)},$$

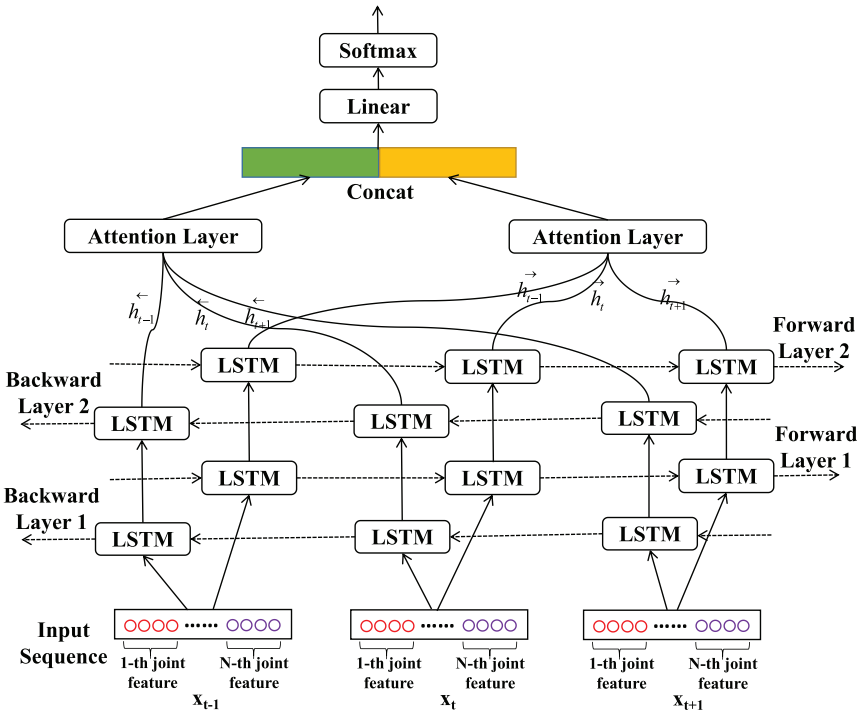


Figure 3. The structure of LSTM-based neural network

$$\text{ma}F1 = \sum_{j=1}^{C} \frac{2P_j R_j}{P_j + R_j} / C,$$

$$P_j = \frac{tp_j}{tp_j + fp_j}, \qquad \text{(Equation 2)}$$

$$R_j = \frac{tp_j}{tp_j + fn_j},$$

where C is the number of categories, and C is set to 3 in our case. The kappa statistic assesses the measurement consistency between human experts and neural network prediction results. The Cohen's Kappa used in our experiment is defined as:

$$\kappa = \frac{(p_o - p_e)}{(1 - p_e)}, \qquad \text{(Equation 3)}$$

where p_o is the actual observed agreement ratio between two kinds of the raters, and p_e is the expected agreement ratio when both annotators assign labels randomly. The level of agreement was further evaluated using the scales in Table 1[27]

Model comparison results

In this experiment, we evaluated the performance of the proposed method from two perspectives. First, we compared the performance of our proposed framework with two baseline method (i.e., AdaBoost.M1[28] and FCN[26]) which have been used in recent studies for FMS evaluation.

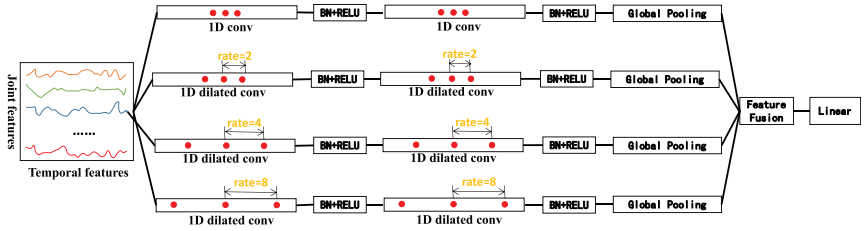


Figure 4. The structure of CNN-based neural network

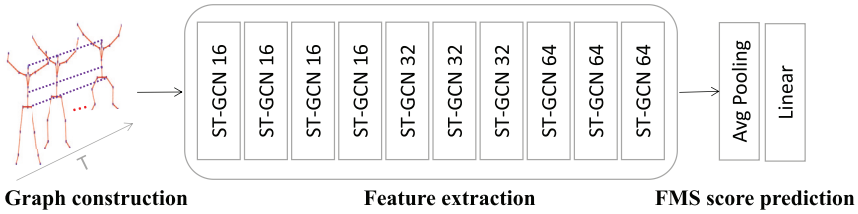


Figure 5. The structure of ST-GCN network

Second, we investigated how different network architectures and viewpoints affected the performance of our method. The comparison results are shown in Table 2.

From the results in Table 2, we can see that our proposed framework clearly outperforms the two baseline methods. In fact, the CNN-based method with single view is a modified version of the FCN method that controls the receptive field to achieve multi-scale feature fusion. Specifically, four branches with distinct receptive fields are used to achieve multi-scale perception, and the size of each branch's receptive field is controlled through different dilation rates. Compared with FCN and CNN-based methods, we found that this multi-scale strategy achieves significantly better performance on all three kinds of evaluation metrics.

We also investigate the performance comparison of the proposed framework for the task of FMS movement assessment using three mainstream networks. In Table 2, the results for LSTM-based, CNN-based, and GCN-based methods are presented from the fifth row to the end, evaluated independently for front view, side view, and multi-view fusion. The best performance of both single-view and multi-view models are highlighted in bold. Overall, the GCN-based model outperforms the others in both single-view and multi-view scenarios, followed by the CNN-based and LSTM-based models.

To summarize, we can draw three conclusions. First, the CNN-based and GCN-based methods outperform the LSTM-based methods in FMS movements assessment for both single-view and multi-view input. This is because the CNN-based and GCN-based methods excel at capturing subtle differences in spatial information, which is a crucial aspect in the task of FMS movement assessment, compared to the sequence modeling approach of the LSTM-based methods. Second, front-view feature data generally produces higher miF1, maF1 and Kappa scores than side-view feature data due to severe vision occlusion from the latter. While LSTM-based methods exhibit similar results with both views, significant differences exist for CNN-based and GCN-based methods. Lastly, multi-view fusion models that capture complementary information from different views achieve better results than the compared methods for both front and side views. Furthermore, early fusion and late fusion can achieve similar performance in the LSTM-based methods and the CNN-based methods, while late fusion is proved to be better in GCN-based methods.

In addition, we have plotted the learning curve in Figure S2. The learning curve demonstrates that the CNN-based and GCN-based approaches exhibit superior convergence compared to the LSTM-based method. Consistent with this, our experiments reveal that the CNN-based and GCN-based methods achieve better performance than the LSTM-based approach.

Per-action assessment results

Since that the assessment complexity of seven FMS movement tests varies, we further investigate the model performance for each FMS test individually. The micro-averaged and macro-averaged F1-measure for seven FMS movements are depicted in Figure 7.

Figure 7 displays the F1-measure results for the single-view and the multi-view models. For the single view, the CNN-based methods outperform the LSTM-based and the GCN-based methods in most tests. Specifically, CNN-based methods outperform other methods in four out of seven FMS movements including DS test, SM test, ASLR test and RS test, while GCN-based methods achieve the highest

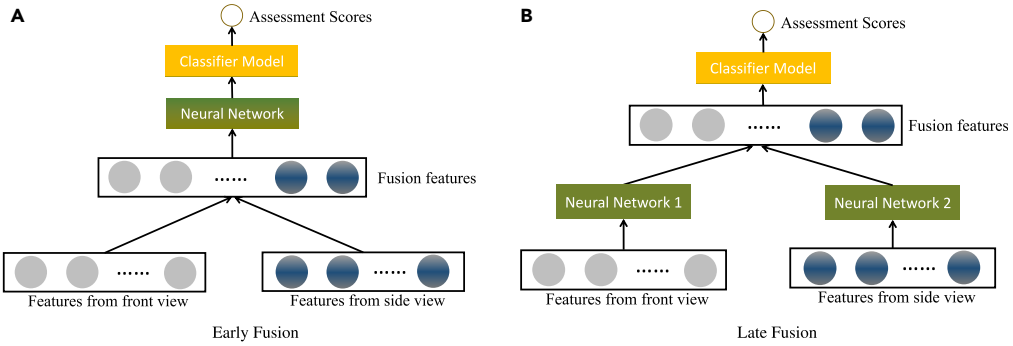


Figure 6. The early fusion strategy and the late fusion strategy

Table 1. The correspondence between the κ values and the levels of agreement

κ value	<0.20	0.21–0.40	0.41–0.60	0.61–0.80	>0.80
Level of Agreement	Poor	Fair	Moderate	Good	Very Good

performance only in the HS test. Furthermore, different feature fusion strategies affect FMS movement tests differently. Early fusion improves the CNN-based methods' performance in six FMS movements except the RS test, but has limited effect on GCN-based methods. Late fusion significantly improves the GCN-based methods' performance in five of the seven tests, except for the SM and ASLR tests.

Per-level assessment results

Each FMS movement test can be predicted as a score of 1–3 by models automatically. To analyze the frequency of incorrect scoring, we visualize the confusion matrix based on the manual and automatic scoring. We plot the confusion matrix in Figure 8 obtained by the LSTM-based, CNN-based and GCN-based methods. From Figure 8, we can observe that the misclassified samples tend to be predicted with a score close to their true scores. For example, a 3-point sample is more likely to be predicted incorrectly as 2 points rather than 1 point, and a 1-point sample is more likely to be predicted incorrectly as 2 points rather than 3 points. The evaluation methods make the most errors when grading a 3-points sample as a 2-points sample.

DISCUSSION

Select the most representative features

As mentioned above, the original Azure Kinect cameras track 32 joints for each person. However, not all of these joints are informative for FMS movements test and the irrelevant joints may introduce excessive noise. To address this issue, we analyze the scoring criteria of the FMS movements test by Minick et al.,[29] and select the most informative joints for each movement. The original skeleton joints from Kinect SDK and the extracted skeleton joints used in our experiments are illustrated in Figures 9A and 9B, respectively. Figure 9C enumerates the joints index and joints name. Table 3 summarizes the joint information for all FMS movements. Finally, we adopted 22 skeleton joints in our movements evaluation experiments by gathering all important skeleton joints from all FMS movements.

We compared the experimental results of 32 original skeleton joints with our carefully chosen 22 skeleton joints to validate the efficiency of our feature selection process. Figure 10 shows the results, with red bars representing the original joints and blue bars representing the selected joints. The models consistently performed better after feature selection. For example, GCN-based methods with late fusion showed a performance boost of approximately 9.73%, 17.43%, and 30.61% for micro-average f1, macro-average f1, and kappa, respectively.

Furthermore, we conducted two sets of controlled experiments, utilizing the GCN-based (Front) method to model the original 32 skeleton joints and the filtered 22 skeleton features, respectively. Each experiment was conducted 10 times, and the results were recorded for three metrics: micro-average F1, macro-average F1, and kappa. By applying the t-test, we calculated the p-values for these three metrics. All resulting p-values were less than 0.05, indicating a significant difference in model performance between the feature-filtered and original feature-based models across these evaluation metrics. For a more detailed experimental process, please refer to Table S2.

Table 2. Overall comparisons for FMS movement assessment with sample-based training and test set splits

Method	mi*F1*	ma*F1*	Kappa	Level of Agreement
Adaboost.M1(Front)[28]	0.568	0.387	0.106	Fair
Adaboost.M1(Side)[28]	0.708	0.338	0.074	Fair
FCN(Front)[26]	0.794	0.642	0.47	Moderate
FCN(Side)[26]	0.773	0.600	0.41	Moderate
LSTM-based(Front)	0.726	0.560	0.35	Fair
LSTM-based(Side)	0.731	0.545	0.33	Fair
LSTM-based(Early Fusion)	0.754	0.591	0.40	Fair
LSTM-based(Late Fusion)	0.758	0.594	0.41	Moderate
CNN-based(Front)	0.827	0.721	0.58	Moderate
CNN-based(Side)	0.812	0.684	0.53	Moderate
CNN-based(Early Fusion)	0.855	0.760	0.64	Good
CNN-based(Late Fusion)	0.854	0.759	0.64	Good
GCN-based(Front)	**0.833**	**0.731**	**0.60**	**Good**
GCN-based(Side)	0.776	0.643	0.48	Moderate
GCN-based(Early Fusion)	0.812	0.699	0.56	Moderate
GCN-based(Late Fusion)	**0.857**	**0.768**	**0.64**	**Good**

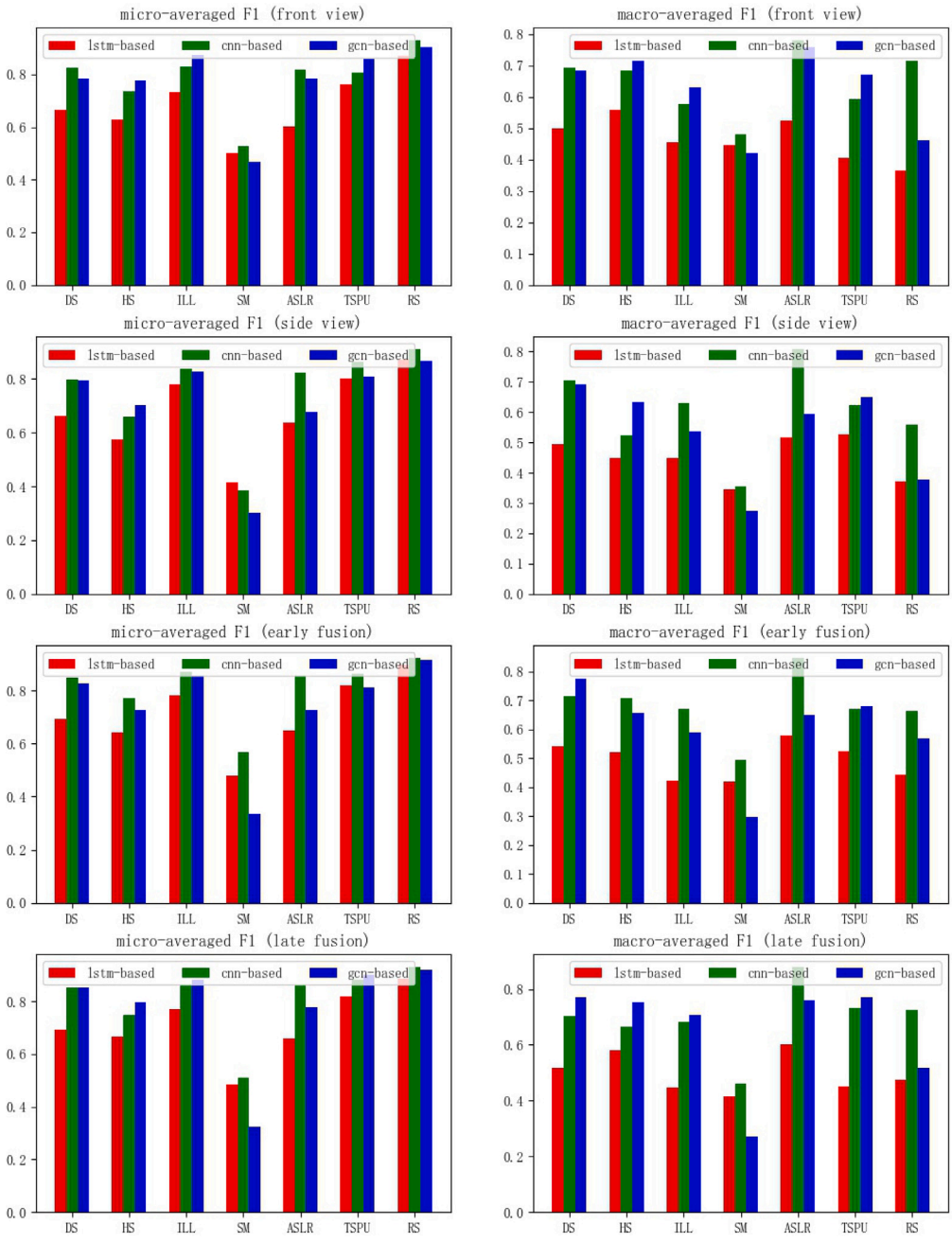


Figure 7. The micro-averaged and macro-averaged F1-measure for seven FMS movements

Ablation study on body joints

Taking into account the interdependence of various body joints, especially those in close proximity, in influencing the final FMS evaluation results (for instance, the knee and hip joints both play a crucial role in assessing the thigh angle during the HS test), we organized all body joints into four distinct groups: facial joints, hand joints, upper limb joints, and lower limb joints (refer to Table S1). Through a series of ablation experiments, we investigated the impact of these distinct joint groupings on the outcomes of the FMS evaluation, as presented in Table 4.

In Table 4, we examine the impact on the GCN-based (Front) model's performance by removing one set of joints from the 32 skeletal joints extracted using Microsoft's Kinect Azure. We observations reveal that the exclusion of the facial joints or hand joints has a minimal or even a positive effect on the final result. However, removing the upper or lower limb joints distinctly leads to a significant decrease in the model's performance.

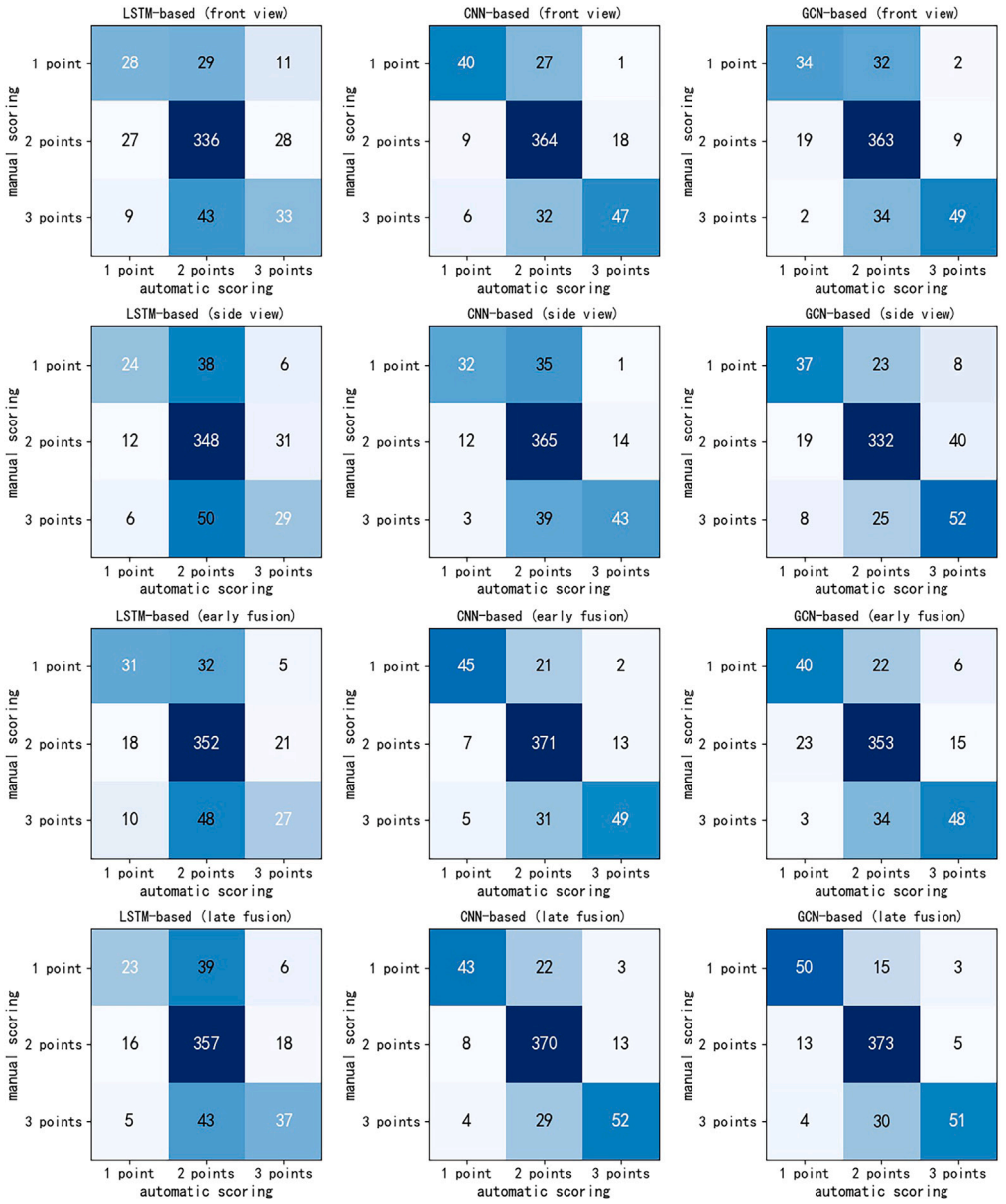
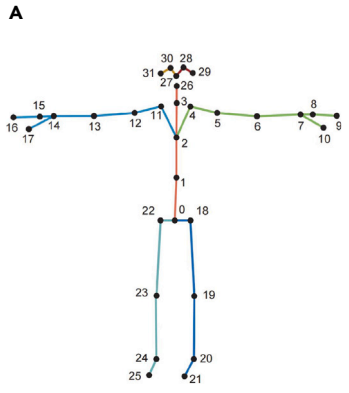


Figure 8. Confusion matrix for per-level assessment in FMS test

Splitting the training and test sets by individuals

To further validate the generalization performance of our proposed model on new subjects, we adopted an individual-based partitioning of the training and test sets. More precisely, we randomly selected samples from 40 individuals for the training set, setting aside samples from the remaining 5 individuals for the test set. Considering that there are only 7 standard actions with 2 scores in the IL test, all action clips from the other 6 FMS movement tests are used in our experiments. We conducted experiments using the CNN-based, LSTM-based, and GCN-based models, with the experimental results summarized in Table 5.

Overall, we observe relatively lower performance with the individual-based dataset splits. This can be attributed to the inherent challenges posed by this validation scheme, stemming from differences in data distribution between the training and test data, particularly when dealing with a small dataset. Among the three types of models, the GCN-based methods continue to achieve the best performance. This further emphasizes that this graph-based model is more suitable for modeling human skeletal joints. Additionally, unlike the sample-based dataset splits, feature fusion strategies did not yield improved performance in this scenario. This could be due to the increased model complexity introduced by the feature fusion process, potentially leading to overfitting issues, particularly in the context of small datasets.



C

Joint index	Joint name	Joint index	Joint name
0	Pelvis	16	Right hand tip
1	Spine navel	17	Right thumb
2	Spine chest	18	Left hip
3	Neck	19	Left knee
4	Left clavicle	20	Left ankle
5	Left shoulder	21	Left foot
6	Left elbow	22	Right hip
7	Left wrist	23	Right knee
8	Left hand	24	Right ankle
9	Left hand tip	25	Right foot
10	Left thumb	26	Head
11	Right clavicle	27	Nose
12	Right shoulder	28	Left eye
13	Right elbow	29	Left ear
14	Right wrist	30	Right eye
15	Right hand	31	Right ear

Figure 9. The skeleton structure and relevant information used in our paper

(A) The skeleton structure extracted by the Azure Kinect camera.

(B) The skeleton structure used in our methods.

(C) The joints index and joints name.

Comparison with existing studies

Due to the time-consuming nature of manual FMS movements test and the potential for bias in subjective evaluation, researchers have turned to automatic FMS tests in recent years. The study in 2014 by Whiteside et al.[10] is the first attempt in this direction. They aimed to evaluate the criterion validity of manual grading by a certified FMS tester using an objective grading system based on IMU motion capture data. Their study revealed poor consistency between manual FMS scores and inertial-based motion capture system results, suggesting that manual grading may not be a valid measurement tool.

To develop an automatic image-capture and angle tracking system for assisting the FMS test, Chang et al.[30] developed an automatic angle tracking system using motion capture software to assist in the FMS test. Mrozek et al.[12] proposed a mechanical model for assessing physical performance in the trunk stability push-up exercise using BTS motion capture data, and the final experiment results show that even the players who obtained the highest marks in the test do not always perform the exercises flawlessly. For all above three studies, assessment model is built based on human dynamics, and manual threshold values are set for distinguishing different levels of FMS test. Different threshold value is appropriate to different FMS test. Each threshold value should be chosen separately and carefully, which makes the threshold setting process very difficult. To develop a universal automatic FMS scoring system for all FMS exercises, Wu et al.[28] view this problem as a classification task. Combined subset feature selection with ensemble learning classifier, they develop the automatic scoring system based on full-body inertial measurement unit sensors for six FMS exercises. However, this work treats feature learning and classifier model as two independent stages, thus can hardly achieve a good performance.

Limitations of the study

There are two limitations in this study. First, model interpretability[31–33] is crucial for practical applications of deep learning methods. If predictive models lack interpretability, users may lose trust in them. Therefore, additional research is necessary to develop interpretable machine learning models that can explain automatic scoring from neural networks. Second, it is important to have access to large datasets in order to train models that can be generalized to real-world scenarios. To test the proposed architectures, more substantial datasets, including FMS tests from a more extensive range of subjects, will be required.

Conclusions

In this paper, we propose a markerless vision-based automatic FMS motion analysis methods. The proposed method can handle FMS movements assessment without human intervention. By extensive experiments, we verified that it was feasible to conduct FMS test using vision sensor, and FMS movement tests with different manual scores can be distinguished by models. We compared the performance achieved by three mainstream neural network architectures and found that CNN-based and GCN-based models have a clear advantage over the LSTM-based models. Additionally, late fusion strategy which use two different neural networks to extract features from front view and side view separately can boost the performance of models more effectively.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

Table 3. A summary of the FMS scoring criteria[29]

	Scoring criteria				
Tests	3 points	2 points	1 point	0 point	Related skeleton joints
DS	Upper torso is parallel with tibia or toward vertical. Femur is below horizontal. Knees are aligned over feet. Dowel is aligned over feet.	Meet criteria of 3 points with compensation.	Tibia and upper torso are not parallel. Femur is not below horizontal. Knees are not aligned over feet. Lumbar flexion is noted.	Complete with pain in any part of the body.	0,1,2,3,7,14,18,19, 20,21,22,23,24, 25,26
HS	Hips, knees, and ankles remain aligned in sagittal plane. Minimal to no movement is noted in lumbar spine. Dowel and hurdle remain parallel. Foot remains dorsiflexed.	Alignment lost between hips, knees, and ankles. Movement is noted in lumbar spine. Dowel and hurdle do not remain parallel.	Contact between foot and hurdle occurs. Loss of balance is noted.	Complete with pain in any part of the body.	2,4,5,7,11,12,14,18, 19,20,22,23,24
ILL	Minimal to no torso movement is noted. Feet remain in sagittal plane on 2 x 6 board. Knee touches 2 x 6 board behind heel of front foot.	Movement is noted in torso. Feet do not remain in sagittal plane. Knee does not touch behind heel of front foot.	Loss of balance is noted.	Complete with pain in any part of the body.	0,1,2,3,4,5,6,7,11,12,13, 14,18,19,20,22,23,24
SM	Fists are within 1 hand length.	Fists are within 1.5 hand length.	Fists are not within 1.5 hand length.	Complete with pain in any part of the body.	0,1,2,3,4,5,6,7,11,12,13,14
ASLR	Dowel resides between mid-thigh and anterior superior iliac spine. Opposite hip remains neutral, toes remain pointing up. Knees remain in contact with board.	Dowel resides between mid-thigh and mid-patella.	Dowel resides below mid-patella.	Complete with pain in any part of the body.	18,19,20,22,23,24
TSPU	Males perform 1 repetition with thumbs aligned with top of head. Females perform 1 repetition with thumbs aligned with chin. Body is lifted as one unit (no lag in lumbar spine). Feet remain dorsiflexed.	Subjects perform 1 repetition in modified position. Male-thumbs aligned with chin Female-thumbs aligned with chest.	Subjects are unable to perform 1 repetition in modified position.	Complete with pain in any part of the body.	0,1,2,3,4,5,6,7,11,12, 13,14,18,19,20,22,23,24,26,27
RS	Subjects perform 1 correct repetition while keeping torso parallel to board. Knee and elbow touch in line over the board.	Subjects perform 1 correct diagonal flexion and extension lift while maintaining torso parallel to board and floor.	Subjects are unable to perform diagonal repetition.	Complete with pain in any part of the body.	0,1,2,3,6,13,19,23

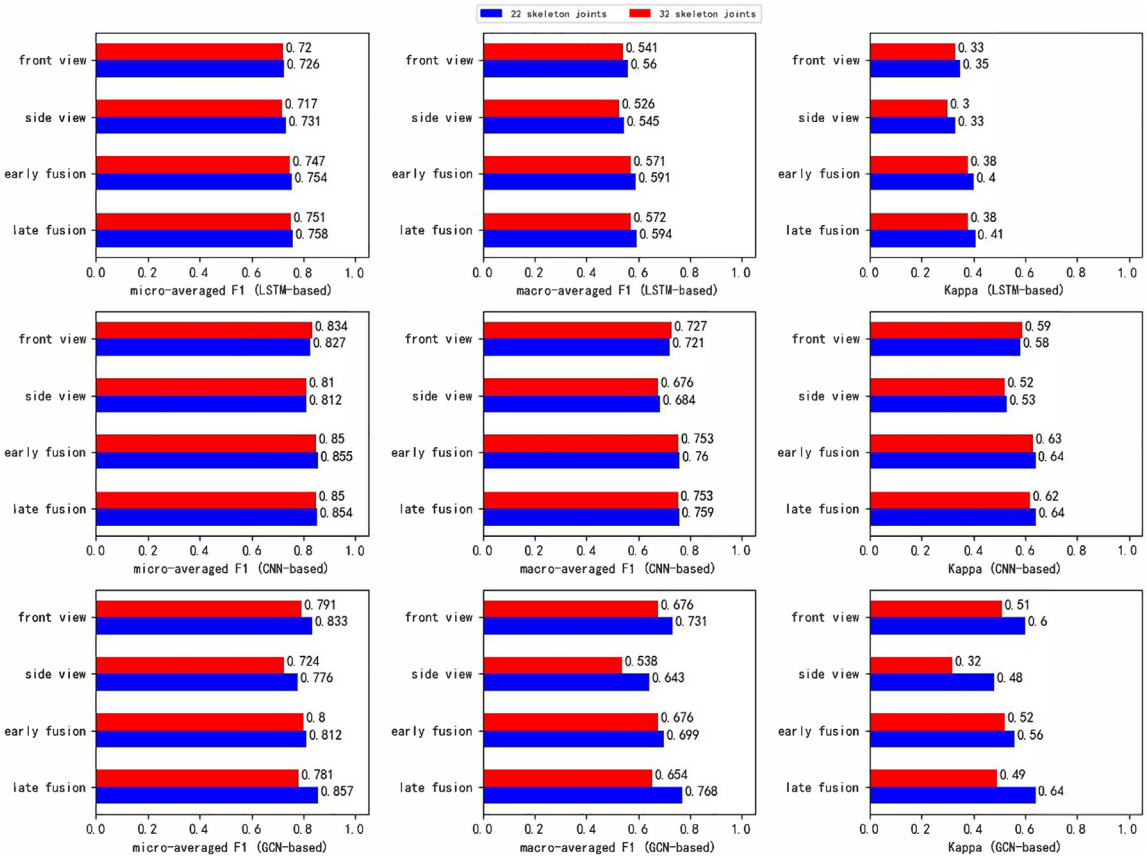


Figure 10. The comparison results of the original 32 skeleton joints with 22 artificial selected skeleton joints

- Lead contact
- Materials availability
- Data and code availability

- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participants
- METHOD DETAILS
  - Problem formulation
  - Neural network architectures
  - LSTM-based neural network
  - CNN-based neural network
  - GCN-based neural network
  - Multi-view fusion models
  - Early fusion
  - Late fusion
- QUANTIFICATION AND STATISTICAL ANALYSIS

Table 4. The effect of different joint groups on FMS movement assessment with the GCN-based (Front) method

Measures	Without facial joints	Without hand joints	Without upper limb joints	Without lower limb joints	All joints
mi*F*1	0.803	0.814	0.788	0.756	0.804
ma*F*1	0.689	0.716	0.665	0.626	0.698
Kappa	0.534	0.572	0.501	0.437	0.543

Table 5. Overall comparisons for the FMS movement assessment with individual-based training and test set splits

Method	mi*F*1	ma*F*1	Kappa	Level of Agreement
LSTM-based(Front)	0.632	0.412	0.122	Fair
LSTM-based(Side)	0.656	0.430	0.181	Fair
LSTM-based(Early Fusion)	0.672	0.448	0.171	Fair
LSTM-based(Late Fusion)	0.655	0.407	0.119	Fair
CNN-based(Front)	0.681	0.499	0.285	Fair
CNN-based(Side)	0.688	0.409	0.12	Fair
CNN-based(Early Fusion)	0.666	0.519	0.263	Fair
CNN-based(Late Fusion)	0.615	0.39	0.08	Fair
GCN-based(Front)	0.725	0.540	0.358	Fair
GCN-based(Side)	0.596	0.452	0.205	Fair
GCN-based(Early Fusion)	0.690	0.516	0.270	Fair
GCN-based(Late Fusion)	0.643	0.509	0.286	Fair

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108705.

ACKNOWLEDGMENTS

This work has been supported by the Natural Science Foundation of Beijing, China under Grant No.9234029 and the National Natural Science Foundation of China under Grant No.72071018.

AUTHOR CONTRIBUTIONS

Conceptualization, Y.Y.S. and Y.F.S.; Methodology, Y.Y.S. and Q.J.X.; Software, Y.Y.S. and Q.J.X.; Validation, Y.Y.S.; Formal Analysis, Y.Y.S.; Writing – Original Draft, Y.Y.S. and Q.J.X.; Writing – Review and Editing, Y.Y.S. and Q.J.X.; Resources, Y.Y.S. and Y.F.S.; Funding Acquisition, Y.Y.S. and Y.F.S.; Supervision, Y.Y.S. and Y.F.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 5, 2023
Revised: October 3, 2023
Accepted: December 7, 2023
Published: December 14, 2023

REFERENCES

1. Hopkins, W.G., Marshall, S.W., Quarrie, K.L., and Hume, P.A. (2007). Risk factors and risk statistics for sports injuries. Clin. J. Sport Med. *17*, 208–210.
2. Abou Elmagd, M. (2016). Common sports injuries. Int. J. Phys. Educ. Sports Health *3*, 142–148.
3. Bardenett, S.M., Micca, J.J., DeNoyelles, J.T., Miller, S.D., Jenk, D.T., and Brooks, G.S. (2015). Functional movement screen normative values and validity in high school athletes: can the FMS be used as a predictor of injury? Int. J. Sports Phys. Ther. *10*, 303–308.
4. Warren, M., Lininger, M.R., Chimera, N.J., and Smith, C.A. (2018). Utility of FMS to understand injury incidence in sports: current perspectives. Open Access J. Sports Med. *9*, 171–182.
5. Cook, G., Burton, L., and Hoogenboom, B. (2006). Pre-participation screening: the use of fundamental movements as an assessment of function-part 1. N. Am. J. Sports Phys. Ther. *1*, 62–72.
6. Cook, G., Burton, L., and Hoogenboom, B. (2006). Pre-participation screening: the use of fundamental movements as an assessment of function-part 2. N. Am. J. Sports Phys. Ther. *1*, 132–139.
7. Shultz, R., Anderson, S.C., Matheson, G.O., Marcello, B., and Besier, T. (2013). Test-retest and interrater reliability of the functional movement screen. J. Athl. Train. *48*, 331–336.
8. Frost, D.M., Beach, T.A.C., Callaghan, J.P., and McGill, S.M. (2015). FMS scores change with performers' knowledge of the grading criteria—are general whole-body movement screens capturing "dysfunction". J. Strength Condit Res. *29*, 3037–3044.
9. Bonazza, N.A., Smuin, D., Onks, C.A., Silvis, M.L., and Dhawan, A. (2017). Reliability, validity, and injury predictive value of the functional movement screen: a systematic review and meta-analysis. Am. J. Sports Med. *45*, 725–732.
10. Whiteside, D., Deneweth, J.M., Pohorence, M.A., Sandoval, B., Russell, J.R., McLean, S.G., Zernicke, R.F., and Goulet, G.C. (2016). Grading the functional movement screen: A comparison of manual (real-time) and objective methods. J. Strength Condit Res. *30*, 924–933.
11. Mrozek, A., Sopa, M., Myszkowski, J., Bakiera, A., Budzisz, P., Kuliberda, A., Białecka, M., Walczak, T., and Grygorowicz, M. (2020). Assessment of the functional movement screen test with the use of motion capture system by the example of trunk stability push-up exercise among adolescent female football players. Vib. Phys. Syst. *31*.
12. Çeliktutan, O., Akgul, C.B., Wolf, C., and Sankur, B. (2013). Graph-based analysis of physical exercise actions. The 1st ACM

International Workshop on Multimedia Indexing and Information Retrieval for Healthcare, pp. 23–32.
13. Doughty, H., Mayol-Cuevas, W., and Damen, D. (2019). The pros and cons: Rank-aware temporal attention for skill determination in long videos. CVPR, 7862–7871.
14. Xiang, X., Tian, Y., Reiter, A., Hager, G.D., and Tran, T.D. (2018). S3d: Stacking segmental p3d for action quality assessment. ICIP, 928–932.
15. Pan, J.H., Gao, J., and Zheng, W.S. (2019). Action Assessment by Joint Relation Graphs. ICCV, 6331–6340.
16. Levin, M., McKechnie, T., Khalid, S., Grantcharov, T.P., and Goldenberg, M. (2019). Automated methods of technical skill assessment in surgery: a systematic review. J. Surg. Educ. *76*, 1629–1639.
17. Lei, Q., Du, J.X., Zhang, H.B., Ye, S., and Chen, D.S. (2019). A survey of vision-based human action evaluation methods. Sensors *19*, 4129–4155.
18. Debnath, B., O'Brien, M., Yamaguchi, M., and Behera, A. (2022). A review of computer vision-based approaches for physical rehabilitation and assessment. Multimed. Syst. *28*, 209–239.
19. Pirsiavash, H., Vondrick, C., and Torralba, A. (2014). Assessing the quality of actions. ECCV, 556–571.
20. Parmar, P., and Morris, B.T. (2017). Learning to score olympic events. CVPRW, 20–28.
21. Liao, Y., Vakanski, A., and Xian, M. (2020). A deep learning framework for assessing physical rehabilitation exercises. IEEE Trans. Neural Syst. Rehabil. Eng. *28*, 468–477.
22. Vamsikrishna, K.M., Dogra, D.P., and Desarkar, M.S. (2016). Computer-vision-assisted palm rehabilitation with supervised learning. IEEE Trans. Biomed. Eng. *63*, 991–1001.
23. Li, H.Y., Lei, Q., Zhang, H.B., and Du, J.X. (2021). Skeleton Based Action Quality Assessment of Figure Skating Videos (ITME), pp. 196–200.
24. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., and Hovy, E. (2016). Hierarchical attention networks for document classification. NAACL, 1480–1489.
25. Yan, S., Xiong, Y., and Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. AAAI *32*, 7444–7452.
26. Xing, Q.J., Shen, Y.Y., Cao, R., Zong, S.X., Zhao, S.X., and Shen, Y.F. (2022). Functional movement screen dataset collected with two azure kinect depth sensors. Sci. Data *9*, 104–120.
27. Kisner, C., Colby, L.A., and Borstad, J. (2017). Therapeutic Exercise: Foundations and Techniques (Fa Davis).
28. Wu, W.L., Lee, M.H., Hsu, H.T., Ho, W.H., and Liang, J.M. (2020). Development of an automatic functional movement screening system with inertial measurement unit sensors. Appl. Sci. *11*, 96–106.
29. Minick, K.I., Kiesel, K.B., Burton, L., Taylor, A., Plisky, P., and Butler, R.J. (2010). Interrater reliability of the functional movement screen. J. Strength Condit Res. *24*, 479–486.
30. Chang, H., Hsueh, Y.H., and Lo, C. (2018). Automatic Image-capture and angle tracking system applied on functional movement screening for athletes. ICKII, 106–107.
31. Ismail, A.A., Gunady, M., Corrada Bravo, H., and Feizi, S. (2020). Benchmarking deep learning interpretability in time series predictions. NIPS (News Physiol. Sci.) *33*, 6441–6452.
32. Papadimitroulas, P., Brocki, L., Christopher Chung, N., Marchadour, W., Vermet, F., Gaubert, L., Eleftheriadis, V., Plachouris, D., Visvikis, D., Kagadis, G.C., and Hatt, M. (2021). Artificial intelligence: deep learning in oncological radiomics and challenges of interpretability and data harmonization. Phys. Med. *83*, 108–121.
33. Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., Bian, J., and Dou, D. (2022). Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. Knowl. Inf. Syst. *64*, 3197–3234.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Skeleton Data	Public access	https://doi.org/10.25452/figshare.plus.c.5774969.v1
Source Code	This paper	https://github.com/shenyuanyuan/FMS_evaluation
Software and algorithms		
Python (version 3.6)	Python software	https://www.python.org/
Cuda (version 11.8.0)	Nvidia	https://developer.nvidia.com/
PyTorch (version 1.10.2)	Pytorch software	https://pytorch.org/
Numpy (version 1.19.2)	Numpy package	https://scipy.org/install/
Scikit-learn (version 0.24.2)	Sklearn package	https://scikit-learn.org/stable/install.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed and will be fulfilled by the Lead contact, Yanfei Shen (syf@bsu.edu.cn).

Materials availability

This study did not generate new unique reagents or materials.

Data and code availability

- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.
- Original code have been deposited at Github, and are publicly accessible as of the date of publication. Open access link is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

The study involved 45 healthy participants (22 females and 23 males) aged between 18 and 59 years (with an average age of 28.7 years and a standard deviation of 10.76 years). They were recruited from the campus of Beijing Sport University. The participants had a weight range of 46-116 kg (with an average weight of 68.8 kg and a standard deviation of 13.01 kg), and a height range of 158-190 cm (with an average height of 171.9 cm and a standard deviation of 7.98 cm). Their body mass index (BMI) varied from 17.1 to 32.82, with an average BMI of 23.14 and a standard deviation of 3.1. Prior to the experiment, all participants confirmed that they did not have any known movement disorders or other health issues that could affect their ability to exercise. Each participant was fully informed about the procedure, introduced to the experimental instruments, and made aware of any potential risks. Data collection took place between May 2021 and June 2021. All the patients were Han Chinese from mainland China and provided written informed consent for the use of the FMS tests data.

METHOD DETAILS

Problem formulation

For all FMS movements assessment, we adopt a unified neural network framework which takes the 3D skeleton sequence data of FMS tests collected from the Azure Kinect sensor as input signals, and predict the categories information of the performance levels. Given a fixed number of training samples N, we define the commonly-used categorical cross entropy (CE) loss as cost function for training our network:

$$J(\theta) = -\sum_{i=1}^{N}\sum_{c=1}^{C} I\left(y^{(i)} = c\right)\log p\left(y^{(i)} = c\middle|x^{i};\theta\right), \quad \text{(Equation 4)}$$

where C is the level number of FMS movements. $I(.)$ is the indicator function and θ is the network parameters to be learned.

Neural network architectures

Different from the task of the action type determination, FMS movements assessment is a fine-grained classification problem. For this kind of problem, the critical issue is how to learn discriminative global features. In this section, we will discuss three types of network structures to extract feature from original 3D skeletonal data.

LSTM-based neural network

LSTM has the characteristic of capturing long-term dependencies, and has been widely used in time series analysis. Attention has the capability of paying attention to the most obvious features in the line of sight, so it is widely used in feature engineering. According to the characteristics of LSTM and attention, a time-series model which combine LSTM and attention is established. The architecture of the LSTM-based model is illustrated in Figure 3.

Taking the joints data for each frame as a feature vector, FMS test assessment based on 3D skeleton sequence can be treated as a task of multivariate temporal classification. To efficiently incorporate both past and future features within a specific time range, we adopt a bi-directional long short-term memory (Bi-LSTM) layer to extract temporal features. A Bi-LSTM layer consists of two LSTM layers in opposite directions: one for forward processing and the other for backward processing. Two Bi-LSTM layers in our models are stacked to create a deeper network architecture.

Different frames of input sequence provide different importance for FMS test assessment. To better capture the temporal context and automatically learn the importance of each frame for FMS test assessment, we utilize an attention mechanism that assigns higher weights to important frame features and reduces the weight of redundant ones. The resulting feature vectors from different frames are then aggregated to form a weighted FMS test movement feature representation. Two attention layers are placed after the last Bi-LSTM layer, with temporal attention computed using the method proposed by Yang et al.[24] Taking the forward LSTM as an example, suppose $H \in R^{d \times T}$ be a matrix consisting of hidden vectors $[h_1, h_2, ..., h_T]$ that the LSTM produced, where d is the size of hidden layers and T is the length of the given action sequence, then attention score can be computed as follows:

$$u_t = \tanh(Wh_t + b),$$
$$\alpha_t = \frac{\exp(u_t^T v)}{\sum_t \exp(u_t^T v)}, \quad \text{(Equation 5)}$$
$$s = \sum_t \alpha_t h_t,$$

where $W \in R^{d \times d}$, $b \in R^d$ and $v \in R^d$ are the weight matrices to be learned in our model. Finally, the forward feature and backward feature from the attention layer are concatenated to form the feature representation of each FMS movement.

CNN-based neural network

The architecture of the CNN-based model is illustrated in Figure 4. To obtain a more powerful representation learning, we adopt a multi-branch CNN with similar network structures to extract different features from input skeleton data in this section. The primary difference among the branches is the dilated rate used in the convolution operation, allowing for the aggregation of temporal information from receptive fields of different sizes. In other words, each branch in our network architecture can capture temporal dependencies at different temporal levels. Our network architecture consists of four branches, with dilated rates set as 1, 2, 4 and 8, respectively.

Each branch in our network architecture is represented by a single composite function: CONV1D-BN-RELU-CONV1D-BN-RELU-POOLING. Here, CONV1D, BN, RELU, and POOLING represent the 1D convolutional layer, batch normalization operator, ReLU activation, and global average pooling, respectively. Temporal local features from different frames within a single FMS test episode are first extracted via Conv1D operation with a kernel size of 3 and stride of 1. Joint features from different body parts are simultaneously fused to generate global features through input-output channels of the Conv1D operation. A batch normalization layer is added after the convolutional calculation to avoid the covariate shift problem caused by changes in data distribution. At the end of each network branch, a global average pooling layer reduces the dimension of the feature vector by averaging the activation values across the temporal dimension. Finally, CNN features from all four branches are concatenated into a single feature vector that represents each FMS test movement.

GCN-based neural network

As the geometric properties of the human skeleton lend themselves well to graph-based representation algorithms, we adopt the spatial temporal graph convolutional network (ST-GCN) model for human action and interaction recognition. Inspired by the success of Yan et al.,[25] our GCN-based network structure is illustrated in Figure 5.

The GCN-based model is composed of three parts: graph construction, feature extraction, and FMS score prediction. For the graph construction, each FMS test movement is represented as a graph $G(V, E)$ over time, where V and E are used to represent the joint points and the connected relationship between them, respectively. The graph includes intra-body edges defining natural connections in human bodies and inter-frame edges connecting the same joints between consecutive frames. For feature extraction, we utilize ten consecutive ST-GCN blocks

consisting of graph convolutions for extracting spatial features and temporal convolutions for learning the transition between frames. By the process of iterative messages passing and information fusion over a joint graph, the GCN-based model can effectively extract discriminative spatial and temporal features based on human skeleton sequence data. Considering the complexity of our task, the numbers of output channels for each ST-GCN blocks are set as 16, 16, 16, 16, 32, 32, 32, 64, 64 and 64. The detailed ST-GCN block structure refer to the work of Yan et al.[25] For FMS score prediction, we perform global average pooling after the ST-GCN blocks to fuse features through temporal average pooling, then use a softmax classifier to obtain the FMS score prediction.

Multi-view fusion models

A specific FMS test action is captured simultaneously by the front and side Azure Kinect sensors, each providing diverse and complementary information. To improve model performance using both views, the central problem is how to properly fuse information from multiple views. In our experiments, we investigate two general feature fusion techniques: early fusion and late fusion.

Early fusion

The early fusion strategy fuses information at the beginning of processing and involves training a single model to extract features from different data sources. In our case, 3D skeletal features from the front view and side view are first concatenated into a joint representation, followed by adoption of a single-stream deep neural network to estimate FMS action quality. An overview of the early fusion strategy is shown in Figure 6A. Since data fusion begins at the outset, this method can more fully integrate the features of different views.

Late fusion

In contrast to early fusion, where features from different data sources are integrated into a multimodal input, the late fusion strategy begins with representation learning of unimodal features. Two separate neural networks are adopted to extract features from the front view and side view, followed by concatenation of these features to obtain better representation learning. Compared to the early fusion strategy, late fusion has a higher computation cost since every modality requires a separate supervised learning stage. A general scheme for late fusion is illustrated in Figure 6B.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical analysis was carried out utilizing Python software (version 3.6; https://www.python.org/). Inter-annotator agreement among three expert annotators was assessed using Spearman's rho and Fleiss' Kappa coefficient (see Figure S1). The performance of the learning model in scoring FMS movements was evaluated using the F1 measure (see Tables 2, 4, 5, Figures 7 and 10). Measurement consistency between human experts and neural network prediction results was examined through the kappa statistic (see Tables 1, 2, 4, and 5). To delve deeper into the accuracy of scoring at each level, a confusion matrix was constructed, where rows corresponded to manual scoring results and columns represented the automatically predicted levels by the learning models (see Figure 8).