Computer Vision-Driven Intelligent Assessment System For Physical Education: Automated Evaluation Of Movement Standardization And Performance Scoring

Ming Jiang*

Department of Physical Education, Changzhou Institute of Technology, Changzhou, Jiangsu, 213032, China

* *Corresponding author. E-mail: jiangminggz@126.com*

Received Aug. 11, 2026; accepted Sept. 15, 2026

Accurate and objective assessment of movement quality is essential for enhancing physical education outcomes; however, traditional evaluation methods depend largely on manual observation, leading to subjectivity, inconsistency, and increased assessment time. This study presents a Computer Vision-Driven Intelligent Assessment System for Physical Education that automates movement recognition, standardization evaluation, performance scoring, and feedback generation. The proposed framework integrates Media Pipe Blaze Pose-based pose estimation with a hybrid Random Forest (RF) and Extreme Gradient Boosting (XGBoost) classifier to analyse student activities effectively. The system utilizes the KTH Action Recognition Dataset, containing six human activities: boxing, handclapping, handwaving, jogging, running, and walking. Video frames are processed through preprocessing, landmark extraction, normalization, and noise reduction, followed by feature extraction from 33 detected body landmarks. Movement quality is evaluated using joint angle deviation, pose similarity, body symmetry, and motion consistency. The hybrid classifier achieved superior performance with an accuracy of 0.9759, precision of 0.9764, recall of 0.9759, and F1-score of 0.9761. The proposed system offers an efficient, scalable, and intelligent approach for automated movement assessment and personalized performance improvement in physical education environments.

Keywords: Computer Vision; Physical Education; Media Pipe Blaze Pose; Human Activity Recognition; Movement Standardization; Performance Scoring

© The Author('s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

http://dx.doi.org/10.6180/jase.202701_36.019

1. Introduction

PE is an important part of contemporary education that helps develop physical skills and fitness in students [1]. PE does not only help in improving physical fitness; it also helps to improve cognition, team building, self-control, and good habits throughout life [2]. The increasing use of digital technology in the current era is leading to the use of AI, computer vision (CV), and machine learning for educational purposes [3]. Computer vision allows for the analysis of human motions through video recordings without wearing any sensors [4].

Growing interest in accurate movement analysis has led to the implementation of human pose estimation tools in the fields of sport science, physiotherapy, medicine, and PE [5]. Traditionally, movement skills are evaluated using manual observations made by coaches, which makes this process quite subjective and labor-intensive. Traditional evaluation methods can hardly supply much information on body posture, joint movements, body symmetry, and movement consistency [6]. The latest developments in MediaPipe BlazePose have allowed real-time detection of 33 body landmarks from RGB videos [7].

It is very vital that proper movement assessment be done in order for students to do physical movements in accordance with the required movement patterns, while at the same time minimizing the chances of acquiring improper postures and sporting injuries [8]. The use of intelligent assessment tools provides an objective assessment of joint angles, body positioning, movement paths, symmetry, velocity, and consistency of motion. Transitioning from conventional PE evaluation practices, which rely on subjective, instructor-dependent observation, toward a computer vision-based assessment pipeline requires more than replacing the observer with a camera; it necessitates a structured computational process capable of replicating and extending expert judgment. Although computer vision-based movement analysis has advanced [4, 5], most studies remain limited to action recognition rather than movement standardization in physical education [3]. Action recognition optimizes inter-class discrimination, learning features that separate activity categories rather than quantify execution quality within a class [7].

Existing PE assessment systems rely on aggregate scoring [2] or auto-assessment feedback [6] without skeleton-level biomechanical comparison, leaving action recognition alone insufficient for standardization or corrective feedback. Conventional action recognition systems primarily classify what movement occurred, while the proposed framework evaluates how well it was performed. The framework extends activity classification by using the identified activity to select a standardized reference movement template and then computing biomechanical similarity measures, joint angle deviation, cosine similarity, alignment error, and dynamic time warping (DTW) against the reference. Activity classification identifies the performed activity, whereas physical education scoring evaluates the quality of its execution. In the proposed system, the RF–XGBoost classifier first determines the activity category, after which joint angles, pose similarity, symmetry, motion consistency, alignment, and temporal similarity are used to assess movement quality.

2. Materials and methods

Explained that computer vision pipeline techniques have been broadly adopted for evaluating the quality of human movements in education and sports fields. Specifically, they showed how pose estimation models like OpenPose and YOLO-based detection systems can be used for detecting the joints, classifying the correctness of the movements and scoring them. Moreover, according to their research, the use of multiple cameras allowed increasing spatial accuracy, making it possible to estimate posture alignment, limb coordination and other movement standards. Simultaneously, according to.[9] vision analysis methods have become popular in terms of assessing movement quality and performance in PE environments.[10] The application of pose estimation, object detection and trajectory tracking models was used to score posture accuracy, identify rule-based events and get automatic scores.[11] found that computer vision technologies had already been successfully employed to assess the quality of movements and performance abilities in rehabilitation and physical education. With the help of pose estimation, object recognition, and motion features analysis, they were able to estimate coordination ability, speed, and accuracy of actions, which would allow automating evaluation and distinguishing the differences between movement patterns of disabled individuals and normal ones.[12] likewise found that markerless pose estimation approaches had been increasingly employed to measure physical performance in educational and clinical practices. It was shown that computer vision algorithms were capable of estimating joint positions and movement accuracy and standardization with low-cost equipment.[13] indicated that the use of lightweight computer vision models for the surveillance of human postures and the assessment of movement quality had been used in the safety, rehabilitation, and PE fields. It is possible to enhance the pose accuracy through the use of 2D to 3D lifting, joint angle optimization, and constraint-based correction approaches to allow detecting the posture errors and movement standards effectively.

Noted that there had been uses of computer vision techniques for quantification of postural alignment and diagnosis of musculoskeletal deviations in both clinical and educational contexts. Smartphone-based measurement of angles, pose tracking, and automatic recognition of landmarks were used to analyze head and neck posture, to measure standardization of movement, and to provide reliable metrics of performance. Similar findings could be observed from another study, where analyzed computer vision systems use for assessing physical fitness and movement quality among school children.

Existing studies reveal three recurring methodological gaps relative to the proposed framework. First, [14] approaches relying on multi-camera setups or pose tracking alone [9] classify or localize movement without computing biomechanical deviation from a reference standard. Second, rehabilitation- and clinical-focused systems [11] emphasize coordination or gait metrics but omit weighted [12], multi-feature performance scoring applicable to PE contexts[15]. Jointly through hybrid classification, biomechanical standardization, and feedback generation. Specifically, the first

gap, absence of reference-based biomechanical deviation in multi-camera and pose-tracking studies, is addressed through the Reference Movement Generation module. The second gap, lack of weighted, multi-feature scoring in rehabilitation-focused systems, is addressed through the Movement Standardization module combining joint angle deviation, cosine similarity, and DTW.

The complete architecture of the proposed intelligent assessment system is shown in Figure 1, where the process begins with the input video data from the KTH Action Recognition Dataset and ends with the feedback provided on the performed activity. The KTH Action Recognition Dataset [16], was employed for designing and evaluation of the proposed Computer Vision-Driven Intelligent Assessment System for Physical Education. Preprocessing steps performed are frame extraction, resizing of images, handling of missing landmarks, normalizing landmarks, and removal of noise through moving average filter. Preprocessing equation are expressed are Eqs. (1)-(3).

$$N_f = F_r \times T \tag{1}$$

Where, $\mathbf{N_f}$ represents as Total number of extracted frames, F_r denotes the Video frame rate (frames per second, fps), T represents the Video duration (seconds).

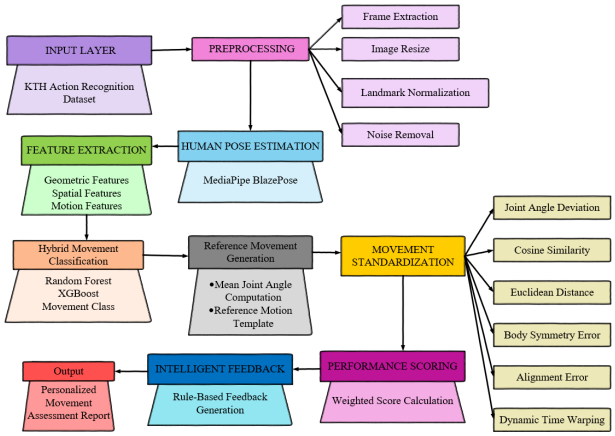

Fig. 1. Workflow Architecture of the Proposed Computer Vision-Driven Intelligent Assessment System

Resizing images by bilinear interpolation ensures that all frames are standardized to a specific resolution prior to estimating the human poses. Linear Interpolation computes the missing landmark coordinates based on the coordinates of the previous and the next frames.

$$L_t = L_{t_1} + \frac{(t - t_1)}{(t_2 - t_1)} (L_{t_2} - L_{t_1}) \tag{2}$$

Where, $\mathbf{L_t}$ represents the Estimated landmark, L_{t_1} denotes the Landmark coordinate in the previous valid frame, L_{t_2} represents the Landmark coordinate in the next valid frame, t denotes the Missing frame index, t_1 *represents the*Previous valid frame index, $\mathbf{t_2}$ represents the Next valid frame index. Consecutive missing landmark frames were handled through linear interpolation between the nearest valid landmark observations before and after the missing sequence, thereby preserving the temporal continuity of the landmark trajectory. When a missing landmark occurred at a boundary frame, where a valid observation was unavailable on one side, the nearest available valid landmark position was propagated to the boundary frame. These procedures ensured that the landmark sequences remained continuous before Min–Max normalization and subsequent geometric, spatial, and motion feature computation. Persistent landmark detection failures that cannot be reliably resolved through interpolation or boundary propagation, the affected frames were excluded from subsequent feature extraction and movement assessment. A sample was retained for assessment only when the resulting 20-frame sequence contained sufficient valid landmark information, thereby preventing unreliable pose estimates from affecting the movement features, performance score, and final feedback.

Min–Max Normalization is used for landmark normalization and rescales the coordinates of the landmarks extracted into a fixed range of 0 to 1. It eliminates the impact of varying image sizes, camera locations, and subject distances on the landmarks.

$$L_{\text{norm}} = \frac{L - L_{\text{min}}}{L_{\text{max}} - L_{\text{min}}} \tag{3}$$

Here, L_{norm} represents the normalized landmark coordinate obtained after applying MinMax Normalization.

The Human Pose Estimation process estimates the position of key body joints for each processed video frame through the use of the MediaPipe BlazePose model. MediaPipe uses the lightweight MediaPipe BlazePose architecture that detects 33 body anatomical landmarks, including the head, shoulders, elbows, wrists, hips, knees, ankles, and feet. Geometric features, joint angles, Euclidean distance, and limb length, quantify static postural correctness and body proportion consistency at a given instant. Spatial features, skeleton alignment, body symmetry, and orientation, capture positional accuracy and balance relative to the reference posture. Motion features, velocity, acceleration, and trajectory, characterize the temporal dynamics and smoothness of execution across frames. Together, these categories provide a comprehensive representation spanning static posture, spatial configuration, and movement dynamics, rather than relying on any single descriptor. Movement

trajectory describes the path travelled by a landmark during an activity expressed in Eq. (4).

$$T = \sum_{t=1}^{N-1} \|P_{t+1} - P_t\| \tag{4}$$

Here, T represents the total movement trajectory length, where P_t and P_{t+1} denote landmark positions in consecutive frames. The variable N represents the total number of frames. The extracted descriptors were grouped into geometric features (joint angles, distances, and limb lengths), spatial features (alignment, symmetry, and orientation), and motion features (velocity, acceleration, and trajectory). These groups represent structural, positional, and temporal movement characteristics and were jointly provided to the RF-XGBoost classifier. Before classification, the descriptors were checked for consistency and redundancy to maintain a uniform feature representation the processing pipeline.

The architecture of the RF model is demonstrated in Figure 2, where the original dataset is subjected to bootstrap sampling to create diversified subsets of data for training individual decision trees. Finally, the activity class with the maximum probability is chosen as a final classification label. The function of the XGBoost is described in Eq. (5).

$$C_{Hybrid} = argmax\left((1-\lambda)P_{RF} + \lambda P_{XGB}\right) \tag{5}$$

Where, C_{Hybrid} represents the final predicted movement class, P_{RF} and P_{XGB} denote the posterior probability vectors generated independently by the Random Forest and XGBoost classifiers, respectively, and λ is the weighting coefficient $(0 \leq \lambda \leq 1)$. The weighting coefficient (λ) controlling the contribution of the XGBoost classifier in the soft-voting fusion was selected through validation-based parameter tuning. Candidate values of λ = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 were evaluated using the validation folds of the training data. The λ value producing the highest validation accuracy and F1-score was selected as the final fusion weight. The independent 20% test set was not used during this selection process and was retained exclusively for final performance evaluation. Probability-level fusion lies in the complementary error characteristics of RF and XGBoost: RF reduces variance through bootstrap aggregation across decorrelated trees. Reference Movement Generation module aims at building the standardized movement model, which is used as the reference to evaluate the performance of the students.

The reference videos were screened based on joint-angle consistency, body alignment, symmetry, movement trajectory, and temporal consistency. Sequences showing incomplete execution, abnormal posture, unstable landmarks, or substantial movement deviations were excluded. Only

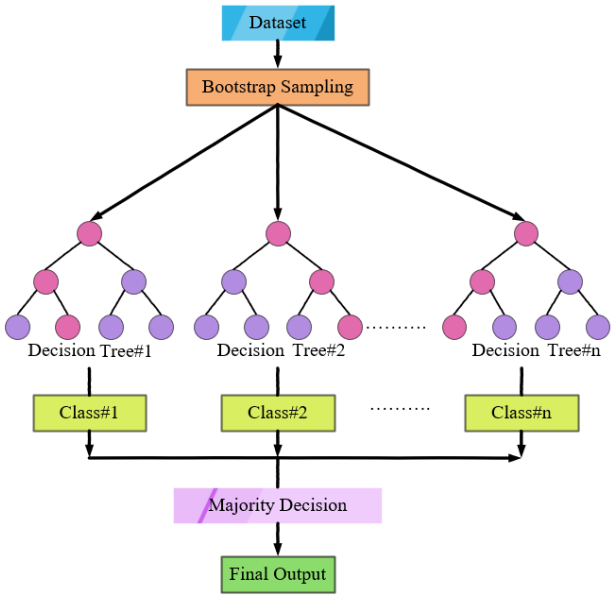

Fig. 2. Random Forest Model Architecture

consistent and representative sequences satisfying these criteria were retained to construct the Reference Movement Library, thereby improving the reliability of the reference movement templates.

The Reference Movement Generation equations are represented as Eqs. (6) and (7). The mean joint angle is calculated by averaging the corresponding joint angles from all correctly performed reference movements.

$$\bar{\theta}_j = \frac{1}{N} \sum_{i=1}^{N} \theta_{ij} \tag{6}$$

Here, $\bar{\theta}_j$ represents the mean angle. The variable θ_{ij} denotes the angle, while N represents the total number of reference movement samples. The Reference Motion Template, the selected reference sequences were temporally aligned to account for differences in movement duration and execution speed. Each sequence was normalized to a common temporal scale, and corresponding movement phases were synchronized across the reference samples. The temporally aligned joint-angle, spatial, and motion features were then integrated across sequences to obtain a representative reference template, reducing the effect of variations in execution speed and sequence duration.

After generating the standardized reference skeletons, they are arranged sequentially to represent the complete movement pattern.

$$R = \{S_1, S_2, \ldots, S_T\} \tag{7}$$

Here, **R** denotes the reference motion template. The

variables $S_1, S_2, \ldots, S_T$ represent the standardized reference skeletons, where T is the total number of frames in the movement sequence. Joint angles are inherently scale- and proportion-invariant, the averaged reference template remains valid across participants of differing height, limb length, and build, as angular relationships rather than absolute landmark distances are compared. Coordinates are additionally normalized prior to averaging, removing dependency on body size or camera distance. This ensures the reference template reflects standardized movement geometry rather than participant-specific anthropometry, preserving consistent scoring criteria across activities and body types.

DTW evaluates the temporal similarity between the student movement sequence and the reference movement sequence expressed in Eq. (8).

$$DTW(i,j) = d(i,j) + min \begin{cases} DTW(i-1,j), \\ DTW(i,j-1), \\ DTW(i-1,j-1) \end{cases} \quad (8)$$

Here, $DTW(i,j)$ represents the cumulative alignment cost between the student and reference movement sequences. The term $d(i,j)$ denotes the distance between the frame. Student sequences vary in frame count relative to the fixed-length reference template, dynamic time warping is applied first to establish an optimal frame-to-frame correspondence between the student and reference sequences, compensating for differences in execution speed and duration. The resulting DTW-aligned frame pairs are then used as the basis for computing all subsequent frame-wise measures, joint angle deviation, cosine similarity, Euclidean distance, and alignment error, ensuring that comparisons are made between temporally corresponding movement phases rather than arbitrarily indexed frames.

The weight distribution was determined based on the relative biomechanical significance of each feature to movement correctness in physical education contexts. Joint angle accuracy was assigned the highest weight (35%) as it directly reflects fundamental postural correctness and is most strongly associated with injury risk when deviated. Pose similarity (25%) reflects overall conformity to the reference movement pattern, while body symmetry and motion consistency (15% each) capture secondary balance and temporal regularity considerations. Stability (10%) was weighted lowest as a supplementary indicator. This weighting further reflects that joint angle accuracy captures localized, segment-level postural deviation, which pose similarity's holistic skeleton-wide comparison can mask when isolated errors are averaged out, justifying the higher priority given to granular joint-level correctness over aggregated pose conformity. Algorithm 1 describes the overall process flow of the intelligent assessment system based on computer vision.

Algorithm 1. Pre-Assessment Processing

Input: Student video V_{student}, Reference Movement Library RML

Output: Aligned student-reference sequence pair

1: Extract frames from V_{student}
2: Detect landmarks and handle preprocessing (resize, missing-landmark handling, normalization, noise filtering)
3: Extract geometric, spatial, and motion features
4: Classify activity using hybrid RF-XGBoost classifier
5: Select reference template from RML based on classified activity
6: Temporally align student sequence with selected reference template
7: **return** Aligned sequences for Movement Standardization

The selected preprocessing settings included 20 frames, 640 × 480 resolution, BlazePose complexity of 2, confidence threshold of 0.5, and a 5 × 5 moving-average filter. The final classifiers used 100 Random Forest trees with maximum depth 15 and 100 XGBoost estimators with maximum depth 6 and learning rate 0.1.

The MediaPipe BlazePose model (model complexity=2) estimates 33 body landmarks with detection and tracking confidence of 0.5. The landmark coordinates are normalized using Min-Max normalization and noise is filtered using 5x5 Moving Average Filter. The processed KTH dataset comprised 6,880 samples, including 5,504 training samples (80%) and 1,376 testing samples (20%) distributed across boxing, handclapping, handwaving, jogging, running, and walking. The 80:20 train-test split was performed at the participant level rather than at the frame or clip level, ensuring that all video samples belonging to a given participant were assigned exclusively to either the training or the testing set.

The system under development was designed on a 12th Gen Intel(R) Core(TM) i9-14900 processor that had 32 GB of RAM and ran Windows 11 (64-bit) operating system. The system was designed using Python 3.11.9 in a CPU-based computational environment. MediaPipe BlazePose served for real-time estimation and extraction of 33 body landmarks from images, NumPy was employed for numerical computations and features manipulation. The Random Forest algorithm was developed using the Scikit-learn library, while the XGBoost classifier was built with the help of the XGBoost library. TensorFlow 2.20.0 was used as a founda-

tion of machine learning algorithms implementation.

3. Results and discussion

This section gives the experimental outcomes of the intelligent assessment system with computer vision.

Confusion Matrix of Proposed Hybrid RF–XGBoost Classifier is shown in Figure 3. The number of correct classifications were 384 for boxing, 383 for hand clapping, 382 for hand waving, 76 for jogging, 75 for running, and 76 for walking with slight errors between jogging and running.

The calibration curves of the suggested hybrid model of RF-XGBoost are shown in Figure 4 for six different activity classes.

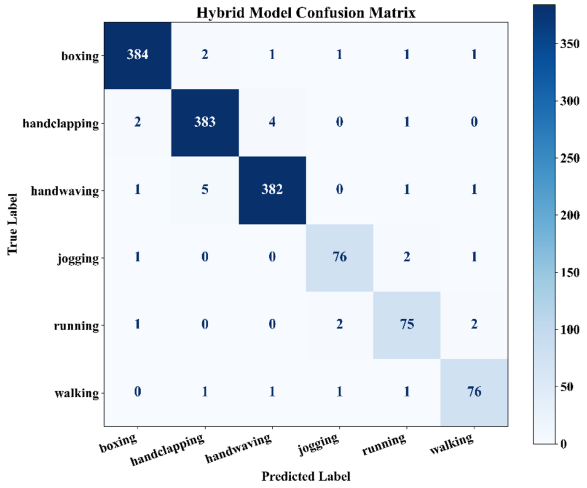

Fig. 3. Hybrid Model Confusion Matrix

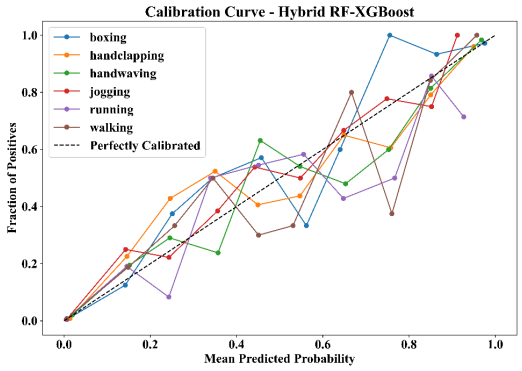

Fig. 4. Calibration Curve for Hybrid RF+XGBoost

Figure 5 shows the ROC curves of the proposed hybrid RF-XGBoost classifier for six activity types[cite: 2, 3]. The high AUC scores (0.97–1.00) suggest good discriminative ability, but the overlapping region between the two activities (jogging and running) implies the similarity in their motion patterns.

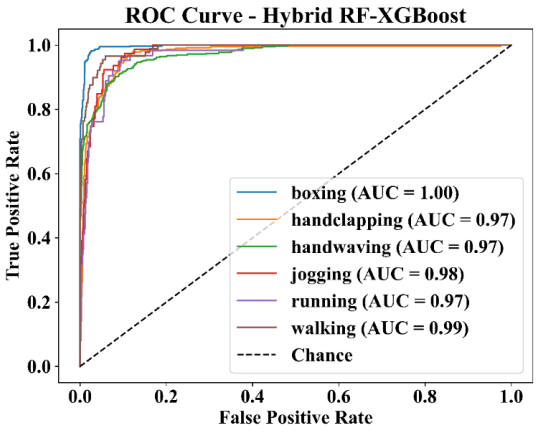

Fig. 5. ROC Curve for Hybrid RF+XGBoost

Comparison of FPR and FNR for the six activity classes is shown in Figure 6. The minimum FPR was found to be 0.012 for the walking class whereas maximum FPR of 0.061 was for handwaving. The maximum FNR of 0.397 was for running, and 0.354 for jogging. The comparatively high false-negative rates for running (0.397) and jogging (0.354) can be attributed to the biomechanical similarity between these two locomotion patterns, both involving comparable joint angle trajectories, stride symmetry, and limb velocity profiles, which the pose-based feature vector does not fully disambiguate. This is consistent with the confusion matrix and ROC overlap observed between these classes. Future improvements may incorporate temporal motion descriptors such as stride frequency or cadence-based features to enhance discrimination between kinematically similar activities.

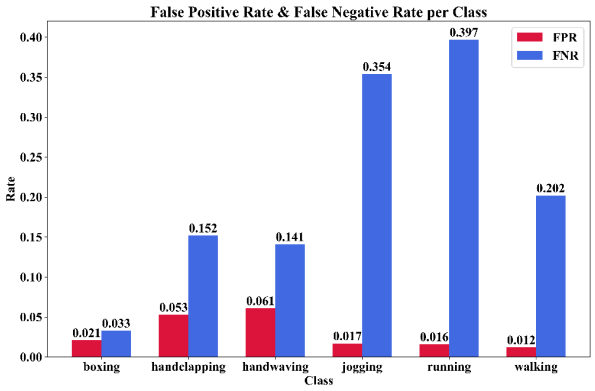

Fig. 6. False Positive and False Negative Rates per Class

Table 1. Human Instructor Evaluation of Generated Feedback Strategies

Activity	No. of Feedback Cases	Correctly Identified Strategy	Strategy Accuracy (%)	Instructor Relevance (%)	Instructor Usefulness (%)
Boxing	20	19	95.0	95.0	90.0
Handclapping	20	20	100.0	95.0	95.0
Handwaving	20	19	95.0	95.0	90.0
Jogging	20	18	90.0	90.0	90.0
Running	20	18	90.0	90.0	85.0
Walking	20	19	95.0	95.0	95.0
Overall	120	113	94.2	93.3	90.8

Pose Cosine Similarity is shown for six activities in Figure 7. The disparity between the high cosine similarity values (0.98–0.99) and the comparatively moderate standardization scores (60–75) arises because cosine similarity measures only the directional alignment of pose feature vectors, which remains high even under proportional or systematic deviations, whereas the standardization score aggregates absolute joint angle deviation, alignment error, symmetry, and DTW-based temporal misalignment.

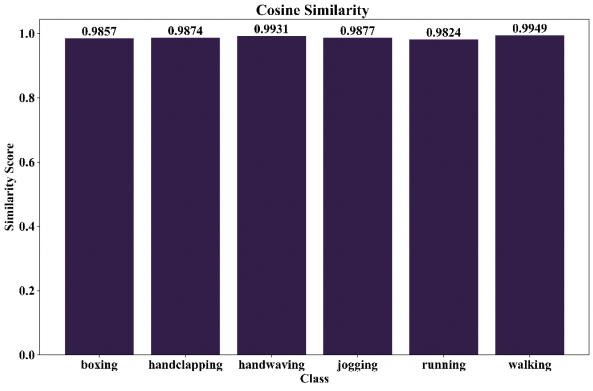

Fig. 7. Pose Cosine Similarity Across Activities

Performance results of the six activity groups are shown in Figure 8. The highest performance values are seen in Handclapping (≈65-66) and Handwaving (≈67-68). Walking (≈64-65) and Standing (≈63-64) have moderate performance values. Jogging (≈54-55) and Running (≈51-52) have low values. Boxing (≈59-60) falls into mid-range. The performance scores translate biomechanical similarity measures, joint angle deviation, cosine similarity, symmetry, and alignment error, into a single interpretable metric aligned weighting scheme, linking computational output to movement quality rather than raw classification confidence. Lower-scoring activities (jogging, running) show elevated joint angle deviation and reduced cosine similarity, consistent with execution inconsistency, while higher-scoring activities (handclapping, handwaving) show tighter biomechanical conformity.

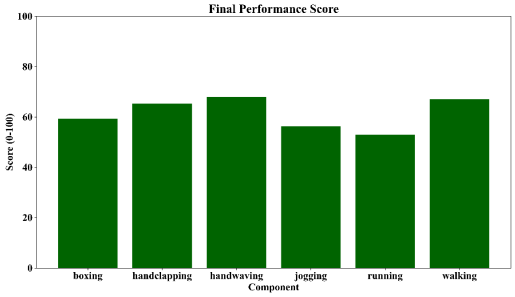

Fig. 8. Final Performance Score Across Activities

Figure 9 shows the feedback dashboard developed from the system that has been proposed. The dashboard provides information on the performance score.

Table 1 evaluates the accuracy and relevance of generated feedback strategies. The system achieved 94.2% strategy identification accuracy, while instructors rated 93.3% of recommendations as relevant and 90.8% as useful. Overall, the results demonstrate effective and practical feedback for movement correction.

Table 2 presents the instructor-based validation of individualized feedback generated by the proposed system.

Table 3 shows the baseline comparison between six different models. RF+XGBoost is always better than any other model.

Table 4 is the ablation study where individual modules were eliminated to see their effect on performance.

4. Conclusion

The proposed framework demonstrates three distinct outcomes that should be interpreted separately. First, activity classification performance was rigorously validated, with

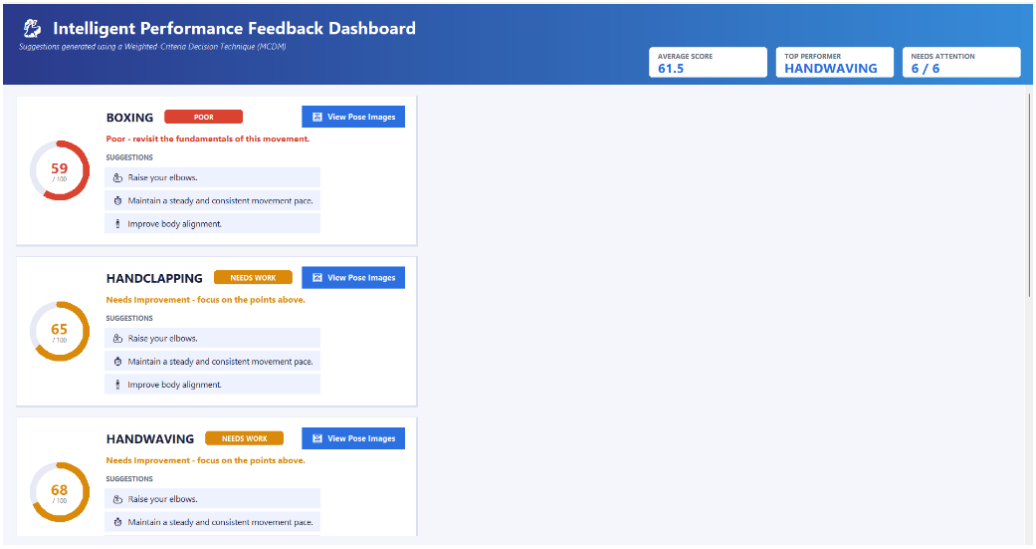

Fig. 9. Intelligent Performance Feedback Dashboard

Table 2. Instructor-Based Validation of Individualized Feedback

Activity	No. of Cases	Correct Recommendations	Correctness (%)	Relevant Recommendations	Relevance (%)
Boxing	20	19	95.0	19	95.0
Handclapping	20	20	100.0	19	95.0
Handwaving	20	19	95.0	19	95.0
Jogging	20	18	90.0	18	90.0
Running	20	18	90.0	18	90.0
Walking	20	19	95.0	19	95.0
Overall	120	113	94.2	112	93.3

the hybrid RF-XGBoost model achieving 0.9759 accuracy, 0.9764 precision, 0.9759 recall, and 0.9761 F1-score, outperforming baseline classifiers and prior literature. Second, movement standardization scores and performance scores were computed through biomechanical deviation measures, joint angle deviation, cosine similarity, alignment error, and DTW, relative to the generated reference templates, but were not independently benchmarked against expert-assigned quality ratings. The proposed architecture was tested using the KTH Action Recognition dataset under controlled settings that could restrict the application of the model to the wider range of PE environments. In future research, efforts would be made to validate the architecture through testing it on more extensive and diverse data sets obtained from real-world PE environments, extending activity recognition ability to complicated movements performed in sports activities, as well as multi-person analysis.

Declaration

Data Availability

https://www.kaggle.com/datasets/vafaeii/kth-action-recognition-dataset

Conflicts of Interest: The author declares that there are no conflicts of interest regarding the publication of this manuscript.

Funding Statement: The author received no specific funding for this research from any public, commercial, or not-for-profit funding agency.

Author Contribution: Ming Jiang: Conceptualization, methodology, software development, data curation, formal analysis, investigation, validation, visualization, writing—original draft preparation, writing—review and editing, and supervision.

Ethical Approval: This study did not involve human participants, human biological materials, or animal subjects. The proposed framework was developed and evaluated using the publicly available KTH Action Recognition Dataset.

Table 3. Baseline Model Comparison

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.9132	0.914	0.9132	0.9134
RF	0.9481	0.9492	0.9481	0.9485
XGBoost	0.9627	0.9634	0.9627	0.9629
SVM	0.9418	0.9425	0.9418	0.942
KNN	0.9285	0.9294	0.9285	0.9288
RF + XGBoost (Proposed)	**0.9759**	**0.9764**	**0.9759**	**0.9761**

Table 4. Ablation Study Results

Configuration	Accuracy	Precision	Recall	F1-Score
Without Motion Features	0.9182	0.9215	0.9182	0.9198
Without Spatial Features	0.9274	0.9296	0.9274	0.9285
Without Geometric Features	0.9341	0.9367	0.9341	0.9353
Without Reference Movement Library	0.9426	0.9444	0.9426	0.9435
Without Movement Standardization Module	0.9512	0.9529	0.9512	0.9520
Without Performance Scoring Module	0.9608	0.9621	0.9608	0.9614
Full Proposed Hybrid Model	**0.9759**	**0.9764**	**0.9759**	**0.9761**

Therefore, ethical approval was not required.

Consent to Participate: Not applicable.

Consent to Publication: Not applicable.

Competing Interests: The author declares that there are no competing interests.

References

[1] V. M. Valle-Muñoz, M. Mendoza-Muñoz, and E. Villa-González, (2025) *"Physical Literacy as a Pedagogical Model in Physical Education"* **Children** ***12***(8): 1008. DOI: 10.3390/children12081008.

[2] H. Liu, H. Min, Y. Wu, Y. Guo, T. Hu, and Y. Shi, (2025) *"Construction of an evaluation index system for core literacy in physical education and health for Chinese college students"* **Scientific Reports** ***15***(1): 41829. DOI: 10.1038/s41598-025-25852-1.

[3] L. Zhao, G. Wu, W. Shao, and X. Ma, (2024) *"Conceptual understanding and cognitive patterns construction for physical education teaching based on deep learning algorithms"* **Scientific Reports** ***14***(1): 31409. DOI: 10.1038/s41598-024-83028-9.

[4] X. Suo, W. Tang, L. Mao, and Z. Li, (2025) *"Digital human and embodied intelligence for sports science: advancements, opportunities and prospects"* **The Visual Computer** ***41***(4): 2477–2493. DOI: 10.1007/s00371-024-03547-4.

[5] S. Edriss et al., (2024) *"The Role of Emergent Technologies in the Dynamic and Kinematic Assessment of Human Movement in Sport and Clinical Applications"* **Applied Sciences** ***14***(3): 1012. DOI: 10.3390/app14031012.

[6] L.-H. Hsia, G.-J. Hwang, and J.-P. Hwang, (2024) *"AI-facilitated reflective practice in physical education: an auto-assessment and feedback approach"* **Interactive Learning Environments** ***32***(9): 5267–5286. DOI: 10.1080/10494820.2023.2212712.

[7] C. Ferraris, G. Amprimo, S. Cerfoglio, L. Vismara, and V. Cimolin, (2025) *"A Deep Dive Into MediaPipe Pose for Postural Assessment: A Comparative Investigation"* **IEEE Access** ***13***: 211055–211074. DOI: 10.1109/ACCESS.2025.3643126.

[8] D. Koźlenia and K. Kochan-Jacheć, (2024) *"The Impact of Interaction between Body Posture and Movement Pattern Quality on Injuries in Amateur Athletes"* **Journal of Clinical Medicine** ***13***(5): 1456. DOI: 10.3390/jcm13051456.

[9] I. Ćirić, N. Ivačko, M. Milić, P. Ristić, and D. Krstić, (2025) *"Computer Vision-Driven Framework for IoT-Enabled Basketball Score Tracking"* **Computers** ***14***(11): 469. DOI: 10.3390/computers14110469.

[10] S. Khater, M. Hadhoud, and M. B. Fayek, (2022) *"A novel human activity recognition architecture: using residual inception ConvLSTM layer"* **Journal of Engineering and Applied Sciences** ***69***(1): 45. DOI: 10.1186/s44147-022-00098-0.

[11] Y. Fan et al., (2025) *"Computer Vision-Driven Digitalization of the Nine Hole Peg Test Assessment Method: A Pilot Study"* **Journal of Medical and Biological Engineering *45***(5): 720–737. DOI: 10.1007/s40846-025-00980-1.

[12] D. Rode, A. Dunkel, R. Willi, P. Wolf, M. Xiloyannis, and R. Riener, (2025) *"Assessment of monocular human pose estimation models for clinical movement analysis"* **Scientific Reports *15***(1): 38767. DOI: 10.1038/s41598-025-22626-7.

[13] J.-W. Kim, J.-Y. Choi, E.-J. Ha, and J.-H. Choi, (2023) *"Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model"* **Applied Sciences *13***(4): 2700. DOI: 10.3390/app13042700.

[14] G. L. Goh, G. D. Goh, J. W. Pan, P. S. P. Teng, and P. W. Kong, (2023) *"Automated Service Height Fault Detection Using Computer Vision and Machine Learning for Badminton Matches"* **Sensors 23**(24): 9759. DOI: 10.3390/s23249759.

[15] A. Kryeem et al., (2025) *"Action assessment in rehabilitation: Leveraging machine learning and vision-based analysis"* **Computer Vision and Image Understanding *251***: 104228. DOI: 10.1016/j.cviu.2024.104228.

[16] A. Vafaei. *KTH Action Recognition Dataset*. 2026. URL: https://www.kaggle.com/datasets/vafaeii/kth-action-recognition-dataset (visited on 07/11/2026).