Developing An Intelligent System For Real-Time Sports Performance Feedback Using AI And Cloud Technologies

Chenxi Qiao[1] and Qiang Zhang[2*]

[1] College of Physical Education, Henan Institute of Economics and Trade, Zhengzhou 450046, Henan, China
[2] Department of Criminal Science and Technology, Henan Police College, Zhengzhou 450046, Henan, China
[*] Corresponding author. E-mail: qiangzhang7878@outlook.com

Received: Apr. 07, 2026; Accepted: May. 25, 2026

Sports performance analysis is important for improving athlete training and decision-making, but many existing systems lack real-time feedback and efficient handling of complex movements. This study proposes an AI and cloud-based framework for real-time tennis action recognition using labeled images of forehand, backhand, serve, and volley. Preprocessing techniques such as resizing, normalization, and augmentation improve data quality. EfficientNet extracts spatial features, while pose estimation captures movement patterns. These features are fused and processed using a BiLSTM model for accurate classification. Results are stored in the cloud for real-time access and scalability. The model achieves high performance with 0.975 precision, 0.981 recall, 0.978 F1-score, and 0.992 accuracy, outperforming existing methods.

Keywords: Sports Action Recognition, BiLSTM, EfficientNet, Pose Estimation, Cloud Computing

http://dx.doi.org/10.6180/jase.202611_34.038

1. Introduction

Sports performance analysis improves athlete skills, but traditional methods are manual, subjective, and time-consuming [1]. AI and cloud systems enable automatic movement analysis and efficient data handling [2].

Existing studies use ML and DL methods [3] like KNN [4], Decision Tree [5], ResNet [6], VGGNet [7], and MobileNet [8] for action recognition. However, many existing approaches rely on offline or batch processing pipelines, where data is analyzed after collection, resulting in delayed feedback, increased latency, and limited applicability in real-time sports training scenarios.

To address these issues, the proposed framework builds a real-time AI and cloud-based sports feedback system. MediaPipe extracts pose features and EfficientNet learns spatial features, while actions are processed and stored in the cloud for fast, accurate feedback.

- Design an AI-based system to track player movements and improve training efficiency.
- Use the Tennis Player Actions Dataset for model training and testing.
- MediaPipe Pose Estimation to extract key body joint features.
- EfficientNet for spatial feature extraction and cloud computing for efficient action recognition and processing.

The study is organized as follows: Section 2 covers the literature review and problem statement, Section 3 presents the proposed methodology, Section 4 discusses results, and Section 5 provides the conclusion along with future work.

2. Materials and methods

Sports action recognition, pose-based movement analysis, and cloud-based analytics have been extensively studied; however, despite their individual effectiveness, their integration into a unified framework remains limited. In 2025, Jain [9] studied AI in sports using DL to evaluate performance and predict injury risk. Neural networks analyzed movement and physiological data, enabling better training decisions. In 2025, Jiang [10] analyzed youth sports health using cloud computing and gait techniques. ML models identified performance patterns, improving athlete monitoring. In 2024, Hussain et al. [11] introduced a real-time sports system for decision-making. It continuously analyzed athlete data, improving performance evaluation. In 2024, Chang et al. [12] proposed a cloud-based system using SVM and CNN. It improved accuracy in sports activity analysis. In 2025, Yan [13] developed a cloud-edge TinyML model for real-time action recognition. It reduced latency and improved accuracy. In 2025, Jia and Liang [14] built an ML-based model for real-time performance prediction. It enhanced training and performance outcomes. In 2025, Grandhi et al. [15] developed a wearable sports monitoring system using optimized SVM for ECG signal analysis to enable real-time athlete performance and health diagnosis. In 2022, Khater et al. [16] developed a residual inception ConvLSTM model for human activity recognition.

The advancements in sports analytics emphasize real-time performance feedback using AI-driven frameworks. Vec et al. [17] highlight the growing use of machine learning, computer vision, and sensor-based systems for continuous monitoring and low-latency feedback. These developments motivate the need for scalable frameworks integrating spatial, temporal, and pose-based features, as proposed in this study.

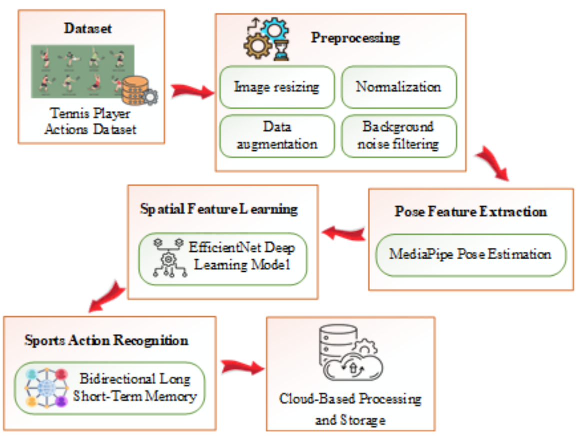


Fig. 1. Proposed methodology

Fig. 1 shows a real-time AI and cloud-based sports feedback system integrating MediaPipe, EfficientNet, BiLSTM, and a preprocessed tennis dataset to address data, computation, latency, and reliability issues. EfficientNet extracts spatial features, MediaPipe captures pose features, and BiLSTM fuses both for action recognition.

Tennis Player Actions Dataset [18] is used to train and test the proposed system, containing labeled images of actions like forehand, backhand, serve, and volley. The dataset is partitioned into 70% training, 15% validation, and 15% testing subsets using a stratified sampling strategy to ensure that each action class is proportionally represented across all splits. This division enables effective model training, hyperparameter tuning, and unbiased performance evaluation on unseen data.

Preprocessing prepares the collected tennis player images to improve input quality before training. All images are resized to 224 × 224 and normalized to [0,1] for EfficientNet processing. Gaussian filtering is applied for noise reduction and augmentation is used to improve dataset diversity. Normalization scales pixel values from 0-255 to 0-1, improving model stability and faster convergence. The values are converted to this range using a standard normalization equation.

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

Data augmentation increases dataset size and diversity by generating new images from existing ones. Techniques like rotation, flipping, and cropping are used, along with adaptive augmentation methods such as dynamic brightness variations and noise addition to simulate varying environmental conditions, thereby improving model robustness and consistency.

Rotation: Rotates images at different angles to simulate player orientations.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{2}$$

Horizontal Flipping: Mirrors images to represent opposite movement directions.

$$x' = W - x \tag{3}$$

Cropping: Focuses on key regions of the player and action area.

$$I_{\text{crop}} = I\left(x_1 : x_2, y_1 : y_2\right) \tag{4}$$

Background noise removal eliminates unwanted visual effects using Gaussian filtering while preserving player posture for better learning.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{5}$$

Pose feature extraction uses MediaPipe to detect key joints and convert images into structured pose data, which is combined with spatial features for action recognition (Equation 6).

$$P = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)\} \quad (6)$$

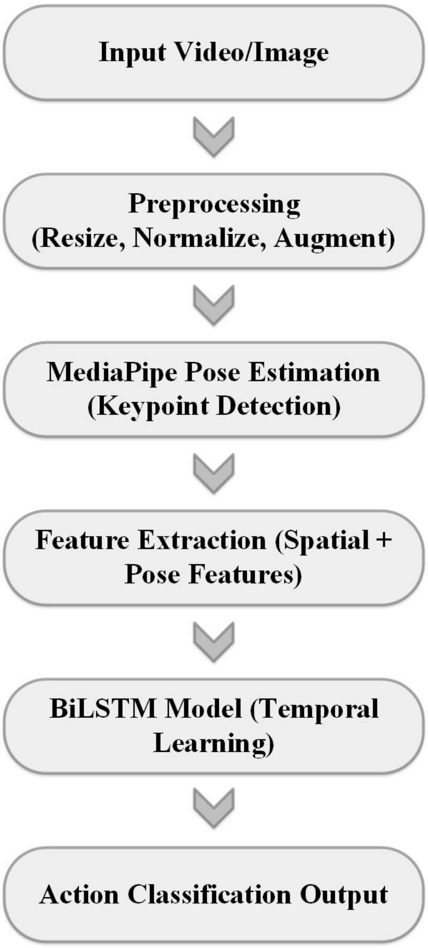


Fig. 2. Workflow of MediaPipe-based pose estimation and feature extraction process.

The workflow of MediaPipe-based pose estimation and feature extraction is illustrated in Fig. 2, providing a clear representation of the sequential processing steps involved in the proposed system.

The algorithm reliably tracks body keypoints even under occlusion and rapid motion, maintaining consistent performance with an observed accuracy of approximately $93-95\%$ and minimal degradation in keypoint detection. Feature extraction captures key patterns like posture, racket position, and movement in sports images. EfficientNet learns deep spatial features using pretrained weights, while BiLSTM uses 128 units with 0.5 dropout. Convolution operations extract high-level features, as defined in Equation (7).

$$F(i,j) = \sum_m \sum_n I(i-m, j-n) K(m,n) \quad (7)$$

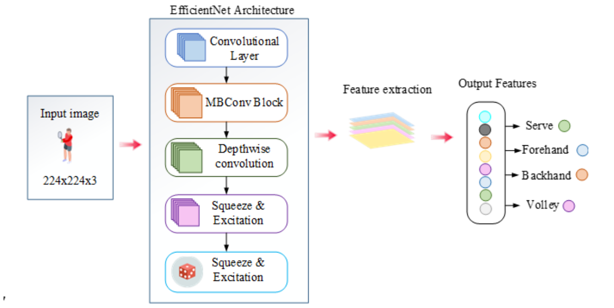


Fig. 3. EfficientNet Deep Learning Model

Fig. 3 shows the EfficientNet model, where images pass through convolutional layers, MBConv blocks, and squeeze-and-excitation modules.

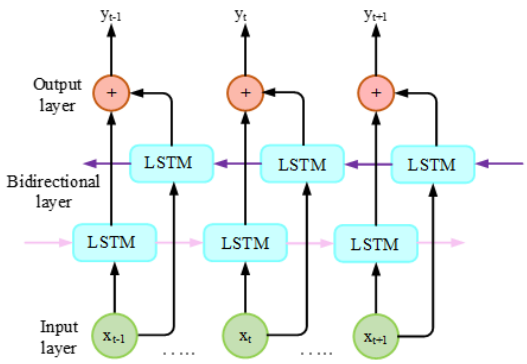


Fig. 4. Bidirectional Long Short-Term Memory

The sports action recognition process identifies tennis movements such as forehand, backhand, and serve. BiLSTM learns forward and backward temporal features, capturing past and future context efficiently. It performs well on medium-sized datasets without needing large training data. BiLSTM processes sequences in both directions, improving contextual understanding of player movements. It captures bidirectional dependencies better than GRU and TCN, though they offer higher efficiency and parallel processing. The BiLSTM architecture for sports action recognition is shown in Fig. 4. The model has two bidirectional layers with 128 units each, followed by a 0.5 dropout layer.

The forward hidden state is computed as equation (8):

$$\overrightarrow{h_t} = f\left(W_x x_t + W_h \overrightarrow{h_{t-1}} + b\right) \quad (8)$$

The backward hidden state is computed as equation (9):

$$\hbar_t = f\left(W_x x_t + W_h \hbar_{t+1} + b\right) \tag{9}$$

The final BiLSTM output is obtained by combining both directions shows equation (10):

$$h_t = \left[\overrightarrow{h_t}; \hbar_t\right] \tag{10}$$

Cloud processing manages the model and results through a pipeline where data is preprocessed, analyzed, and stored for dashboard access. Deployment on AWS and Google Cloud ensures scalability, security, and real-time analytics. Edge devices handle initial processing tasks such as image preprocessing and feature extraction, thereby reducing data transmission to the cloud and minimizing latency, which improves system responsiveness and enables efficient real-time sports analytics.

Classification performance measures are used to assess the suggested system performance.

Accuracy: Correctly predicted instances out of total predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

Precision: Correct positive predictions out of all predicted positives.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

Recall: Correctly detected positives out of actual positives.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

F1-score: Harmonic balance between precision and recall.

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

Performance analysis evaluates predicted actions to generate insights on movements like forehand, backhand, serve, and volley. Results are stored on cloud platforms and displayed via dashboards. Security is ensured using encryption, secure communication, and role-based access control, while athlete data is anonymized for privacy. Coaches and players can access performance statistics, and Table 1 shows the BiLSTM pseudocode.

The model uses Adam optimizer with a 0.001 learning rate, batch size of 32 , and 50 epochs.
Learning rate is reduced on plateau, and categorical cross-entropy is used as the loss function.

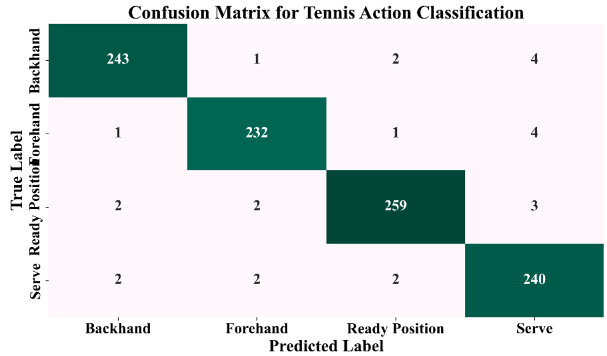


Fig. 5. Confusion Matrix

3. Results and discussion

The proposed BiLSTM system accurately recognizes tennis actions using spatial-pose fusion and preprocessing, ensuring efficient cloud-based processing. Tested on an i5 system, Fig. 5 shows high accuracy with minimal errors. Confusion matrix and ROC curve confirm strong performance.

Fig. 6 (a), (b) shows training and validation performance over 50 epochs. Accuracy improves from 0.25 to 0.99 (training) and 0.30 to 0.98 (validation).

Fig. 7 (a), (b) shows player movement signals. Single-player elbow motion follows a sinusoidal pattern (-1.0 to +1.0), while multi-player data shows overlapping joint variations for three participants (red, green, blue), indicating individual and group movement dynamics.

Fig. 8 compares cloud processing time with data size for AWS EC2 and Google Cloud. System latency includes transfer, inference, and response, enabling real-time feedback with an average inference time of less than 50 ms per frame and overall system latency maintained below 200 ms. Google Cloud performs faster than AWS EC2, which takes around 18 seconds at 5000 MB, showing better efficiency across all data sizes.

Fig. 9 shows ROC curves for backhand, forehand, ready position, and serve, all near the topleft indicating high accuracy. AUC values are 0.970, 0.975, 0.980, and 0.985 (best for serve), confirming strong classification performance above random guessing with minimal false positives.

Fig. 10 shows AUC comparison of models. CNN scores 0.820 , MobileNetV2 0.880 , ResNet50 0.910 , while the proposed BiLSTM + EfficientNet achieves 0.982 , showing superior classification performance over traditional DL models.

Fig. 11 compares evaluation metrics, showing the proposed model outperforms CNN and ResNet50 with lower inference time and higher accuracy. EfficientNet reduces computation and BiLSTM captures temporal dependencies

Table 1. Pseudocode for proposed BiLSTM

Algorithm: Proposed BiLSTM-Based Sports Action Recognition Model
Input: Sports Image Dataset D; sequence length T; learning rate η
Output: Predicted Sports Action Labels
Procedure:

Initialize dataset $D = \{(X_i, Y_i)\}_{i=1}^{N}$.

Perform preprocessing:

a. Resize images to 224×224.
b. Normalize pixel values.
c. Apply noise removal.
d. Perform data augmentation (rotation, flipping, cropping).

Extract features:

$$F_{\text{spatial}} \leftarrow \text{FeatureExtraction}(X_i)$$
$$F_{\text{pose}} \leftarrow \text{PoseEstimation}(X_i)$$

Combine features:

$$F = \text{Concatenate}(F_{\text{spatial}}, F_{\text{pose}})$$

Convert into sequence format:

$$S = \{F_1, F_2, \ldots, F_T\}$$

Initialize BiLSTM model parameters.

a. Compute forward state: $\overrightarrow{h_t} = f(W_x x_t + W_h \overrightarrow{h_{t-1}} + b)$
b. Compute backward state: $\overleftarrow{h_t} = f(W_x x_t + W_h \overleftarrow{h_{t+1}} + b)$
c. Combine outputs: $h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$
d. Compute prediction: $\hat{y}_t = \text{Softmax}(h_t)$
e. Compute loss: $L = -\sum y_t \log(\hat{y}_t)$
f. Update weights using backpropagation.

For test input X:

a. Apply preprocessing.
b. Extract features $F_{\text{spatial}}, F_{\text{pose}}$.
c. Form sequence S.
d. Predict label $\hat{y}$ using trained BiLSTM.

Return predicted sports action (Forehand, Backhand, Serve, Volley).

End

Table 2. Method Comparison

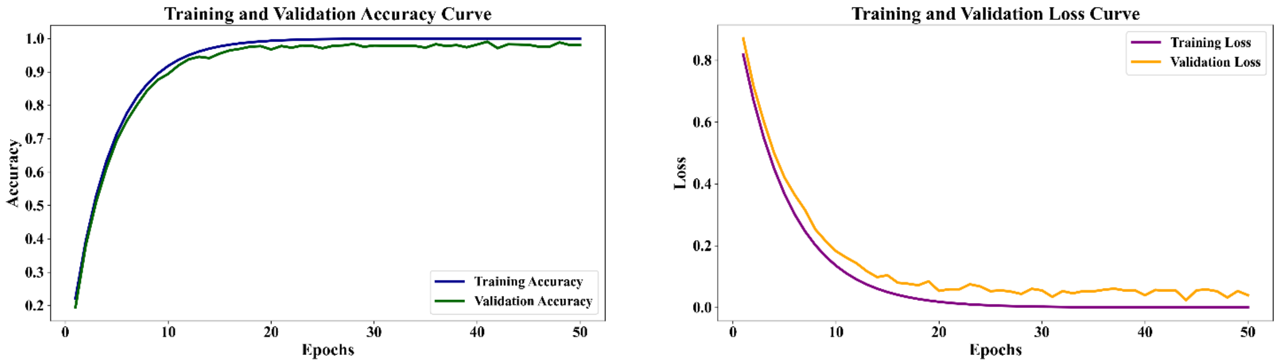
Method	Precision	Recall	F1-Score	Accuracy
CNN [12]	0.79	0.8	0.79	0.81
MobileNetV2 [8]	0.85	0.86	0.85	0.87
ResNet50 [6]	0.89	0.88	0.88	0.9
BiLSTM+EfficientNet (Proposed)	0.975	0.981	0.978	0.992

effectively. The model achieves 99.2% accuracy, with 97.5% precision, 98.1% recall, and 97.8% F1-score.

The ablation study shows EfficientNet alone gives moderate accuracy, while adding pose features significantly improves performance and yields better results. Fig. 12 further shows accuracy rising from 88.0 to 97.0 through BiLSTM, label smoothing, Mixup, and CutMix, with the full model achieving the best results.

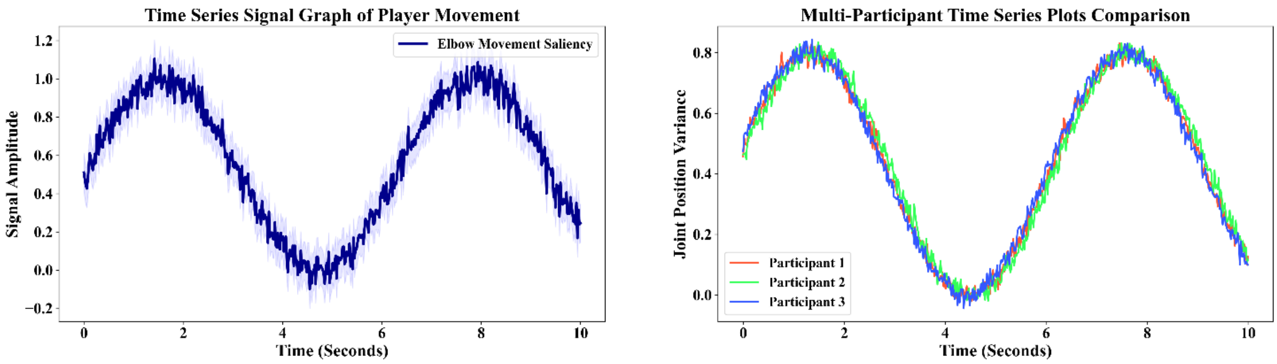
Table 2 compares models for sports action recognition using precision, recall, F1-score, and accuracy. The proposed method outperforms previous works by effectively combining spatial and temporal features. CNN shows moderate performance, MobileNetV2 and ResNet50 improve gradually, while BiLSTM + EfficientNet achieves the best results with 0.975 precision, 0.981 recall, 0.978 F1-score, and 0.992 accuracy.

Experimental results show that the proposed BiLSTM with EfficientNet model outperforms baseline methods across all evaluation metrics. Compared to CNN, it achieves a notable improvement in accuracy along with



(a) Accuracy (b) Loss

Fig. 6



(a) Sigle player (b) Multi player

Fig. 7

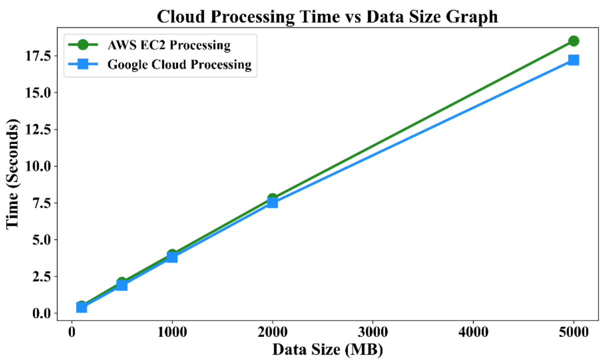


Fig. 8. Cloud Processing time vs data size

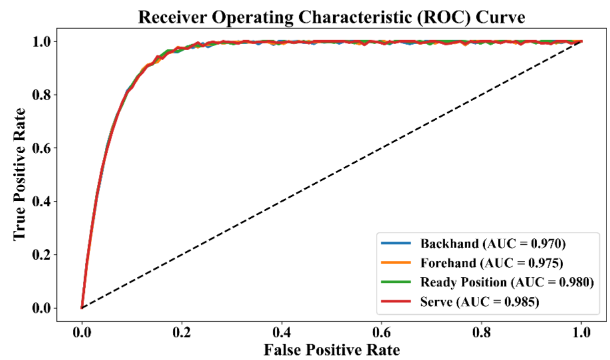


Fig. 9. ROC curve

higher precision, recall, and F1-score. This performance gain is attributed to the effective integration of spatial, temporal, and pose-based features, supported by preprocessing and augmentation strategies, demonstrating the robustness of the proposed approach for sports action recognition.

4. Conclusion

The study proposes a real-time sports feedback system using BiLSTM, EfficientNet, and cloud integration, enabling accurate tennis action recognition with efficient storage and real-time access. The proposed AI-based sports perfor-

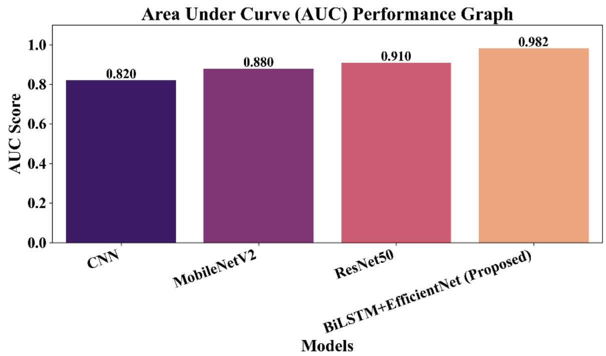


Fig. 10. AUC curve

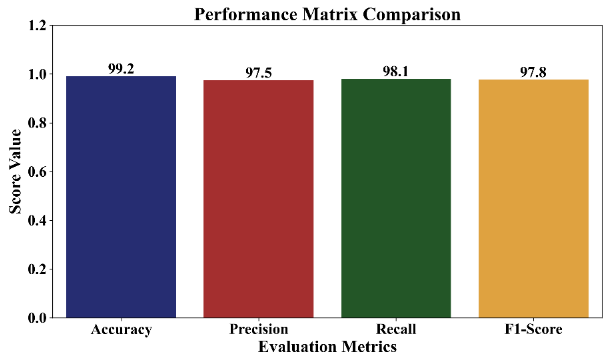


Fig. 11. Matrix comparison

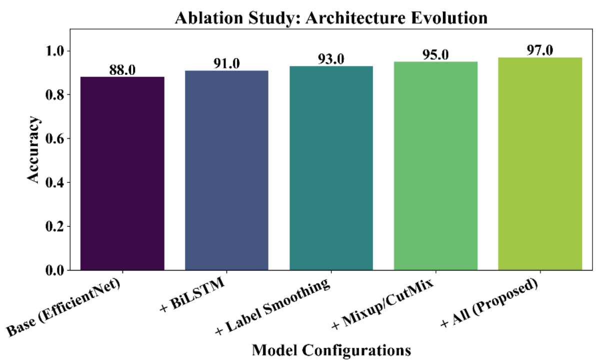


Fig. 12. Ablation study

mance analysis framework can be extended to other sports domains and multi-action recognition scenarios. With suitable dataset adaptation and retraining, the model can be applied to activities beyond tennis, enabling broader applicability in sports analytics. Future work includes video-based recognition, wearable sensors, attention mechanisms, and mobile application development. The framework extends to multiple sports via retraining and adapts to varied environments through its modular design.

5. Acknowledgement

Declarations

Data availability:

https://www.kaggle.com/datasets/orvile/tennis-player-actions-dataset

Conflicts of interest:

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Funding statement:

This study was supported by : Henan Provincial Young Backbone Teachers Training Program for Higher Education Institutions (No.2023GGJS167)

Author contributions:

Chenxi Qiao: Conceptualization, Methodology, Data Curation, Software, Writing - Original Draft Preparation.

Qiang Zhang: Supervision, Validation, Formal Analysis, Writing - Review & Editing, Project Administration.

Ethical approval:

This study does not involve human participants or animals, and therefore ethical approval was not required.

Consent to participate:

Not applicable.

Consent to publication:

Not applicable.

Competing interests:

The authors declare that they have no competing interests.

References

[1] X. Lei and P.-L. P. Rau, (2023) *"Emotional Responses to Performance Feedback in an Educational Game During Cooperation and Competition with a Robot: Evidence from fNIRS"* **Computers in Human Behavior** *138*: 107496. DOI: 10.1016/j.chb.2022.107496. URL: https://doi.org/10.1016/j.chb.2022.107496.

[2] A. Alshardan, H. Mahgoub, S. Alahmari, M. Alonazi, R. Marzouk, and A. Mohamed, (2025) *"Cloud-to-Thing Continuum-Based Sports Monitoring System Using Machine Learning and Deep Learning Model"* **PeerJ Computer Science** *11*: e2539. DOI: 10.7717/peerj-cs.2539. URL: https://doi.org/10.7717/peerj-cs.2539.

[3] L. Zhao, (2023) *"A Hybrid Deep Learning-Based Intelligent System for Sports Action Recognition via Visual Knowledge Discovery"* **IEEE Access *11***: 46541–46549. DOI: 10.1109/ACCESS.2023.3275012. URL: https://doi.org/10.1109/ACCESS.2023.3275012.

[4] H. U. R. Siddiqui et al., (2023) *"Enhancing Cricket Performance Analysis with Human Pose Estimation and Machine Learning"* **Sensors *23***(15): 6839. DOI: 10.3390/s23156839. URL: https://doi.org/10.3390/s23156839.

[5] K. Wang, L. Wang, and J. Sun, (2025) *"The Data Analysis of Sports Training by ID3 Decision Tree Algorithm and Deep Learning"* **Scientific Reports *15***(1): 15060. DOI: 10.1038/s41598-025-99996-5. URL: https://doi.org/10.1038/s41598-025-99996-5.

[6] Y. Hu and D. Liu, (2025) *"Design of Sports Action Recognition and Evaluation Based on Improved DTW Algorithm"* **ICIC International**: DOI: 10.24507/ijicic.21.01.37. URL: https://doi.org/10.24507/ijicic.21.01.37.

[7] C. Mou, (2024) *"The Attention Mechanism Performance Analysis for Football Players Using the Internet of Things and Deep Learning"* **IEEE Access *12***: 4948–4957. DOI: 10.1109/ACCESS.2024.3350036. URL: https://doi.org/10.1109/ACCESS.2024.3350036.

[8] Y. Zheng and L. Cai, (2025) *"Artificial Intelligence-Based Automatic Identification and Classification of Diverse Sports Using Advanced Deep Learning Models"* **International Journal of Information and Communication Technology *26***(23): 91–113. DOI: 10.1504/IJICT.2025.147123. URL: https://doi.org/10.1504/IJICT.2025.147123.

[9] D. S. Jain, (2025) *"AI in Sports: Deep Learning Models for Player Performance Analysis and Injury Prediction"* **International Journal of Research Technology *13***(1): 11–17. URL: https://ijrt.org/j/article/view/130.

[10] H. Jiang, (2024) *"Analysis of Youth Sports Physical Health Data Based on Cloud Computing and Gait Awareness"* **Journal of Intelligent Systems *33***(1): DOI: 10.1515/jisys-2023-0155. URL: https://doi.org/10.1515/jisys-2023-0155.

[11] Y. B. S. Hussain, V. Gunasekhar, D. Reshma, K. S. Kumar, S. V. Reddy, and K. Bhargavi, (2024) *"Revolutionizing Sports Information Systems for Real-Time Analytics for Better Decision-Making"* **Journal of Computer Education and Sports Health *1***(1): 27–40. DOI: 10.69996/jcesh.2024003. URL: https://doi.org/10.69996/jcesh.2024003.

[12] K. Chang, P. Sun, and M. U. Ali, (2024) *"Retracted Article: A Cloud-Assisted Smart Monitoring System for Sports Activities Using SVM and CNN"* **Soft Computing *28***(1): 339–362. DOI: 10.1007/s00500-023-09404-1. URL: https://doi.org/10.1007/s00500-023-09404-1.

[13] C. Yan, (2025) *"TinyML-Enhanced Cloud-Edge Collaborative Framework for Real-Time Sport Action Recognition"* **Internet Technology Letters *8***(5): e70100. DOI: 10.1002/itl2.70100. URL: https://doi.org/10.1002/itl2.70100.

[14] Y. Jia and J. Liang, (2025) *"Real-Time Performance Prediction in Sports Using Machine Learning Algorithms"* **International Journal of Data Science *10***(7): 285–297. DOI: 10.1504/IJDS.2025.151196. URL: https://doi.org/10.1504/IJDS.2025.151196.

[15] S. H. Grandhi, B. R. Gudivaka, R. L. Gudivaka, R. K. Gudivaka, D. K. R. Basani, and M. M. Kamruzzaman, (2025) *"Detection and Diagnosis of ECH Signal Wearable System for Sportsperson Using Improved Monkey-Based Search Support Vector Machine"* **International Journal of High Speed Electronics and Systems**: 2540149. DOI: 10.1142/S0129156425401494. URL: https://doi.org/10.1142/S0129156425401494.

[16] 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 Science *69***(1): 45. DOI: 10.1186/s44147-022-00098-0. URL: https://doi.org/10.1186/s44147-022-00098-0.

[17] V. Vec, S. Tomažič, A. Kos, and A. Umek, (2024) *"Trends in Real-Time Artificial Intelligence Methods in Sports: A Systematic Review"* **Journal of Big Data *11***(1): 148. DOI: 10.1186/s40537-024-01026-0. URL: https://doi.org/10.1186/s40537-024-01026-0.

[18] *Tennis Player Actions Dataset*. Accessed March 11, 2026. 2026. URL: https://www.kaggle.com/datasets/orvile/tennis-player-actions-dataset.