International Journal of Engineering Technologies and Management Research
April 2026 13(4)
www.ijetmr.com

Original Article

ML AND RAG-BASED INTELLIGENT SYSTEM FOR YOGA POSE RECOGNITION AND CORRECTIVE GUIDANCE

Dr. Harish Barapatre [1*], Pratik Malgunde [2], Atharva Pratap [2], Rayan Shaikh [2]
[1] Associate Professor, Department of Computer Engineering, Yadavrao Tasgaonkar Institute of Engineering and Technology, Bhivpuri Road Karjat, Maharashtra, 410201, India
[2] Student, Department of Computer Engineering, Yadavrao Tasgaonkar Institute of Engineering and Technology, Bhivpuri Road Karjat, Maharashtra, 410201 India

ABSTRACT

Yoga pose recognition has gained significant importance in digital health and fitness systems, where accurate posture assessment and corrective feedback are critical for safe practice. Traditional computer vision–based approaches rely on pose estimation models but often lack contextual understanding and personalized guidance. To address this limitation, this paper proposes a hybrid framework that integrates Machine Learning (ML)–based pose recognition with Retrieval-Augmented Generation (RAG) for intelligent feedback generation. The system utilizes human pose estimation techniques to extract skeletal keypoints and classify yoga poses using supervised learning models. Subsequently, a RAG module retrieves relevant expert knowledge from a curated yoga knowledge base and generates context-aware corrective suggestions. This dual-layer architecture ensures both high recognition accuracy and meaningful interpretability of results. The proposed approach aims to bridge the gap between static classification systems and interactive AI-driven coaching by enabling real-time feedback and adaptive recommendations. The framework is designed as a conceptual model with potential applicability in mobile health applications, smart fitness systems, and remote yoga training platforms. By combining data-driven learning with knowledge retrieval mechanisms, the system enhances both usability and reliability in real-world scenarios.

Keywords: Yoga Pose Recognition, Machine Learning, Retrieval-Augmented Generation, Computer Vision, Human Pose Estimation, Digital Health, AI-Based Fitness Systems

INTRODUCTION

Yoga has emerged as a widely adopted practice for improving physical health, mental well-being, and overall lifestyle balance. With the increasing demand for digital fitness solutions, automated yoga training systems have gained attention as a scalable alternative to in-person instruction. However, accurate pose execution is critical in yoga, as improper posture can reduce effectiveness and may even lead to injuries. This creates a strong need for intelligent systems capable of recognizing poses and providing corrective guidance in real time.

Recent advancements in Computer Vision and Machine Learning have enabled the development of human pose estimation models that can detect skeletal keypoints from images and videos. Techniques such as OpenPose and MediaPipe have significantly improved the accuracy of body landmark detection. These approaches allow systems to classify yoga poses based on spatial

*Corresponding Author:
Email address: Dr. Harish Barapatre (harishkbarapatre@gmail.com), Vishal Kumar Goar (pratikpm850@gmail.com)
Received: 28 February 2026; **Accepted:** 12 March 2026; **Published** 30 April 2026
DOI: 10.29121/ijetmr.v13.i4.2026.1768
Page Number: 79-90
Journal Title: International Journal of Engineering Technologies and Management Research
Journal Abbreviation: Int. J. Eng. Tech. Mgmt. Res.
Online ISSN: 2454-1907
Publisher: Granthaalayah Publications and Printers, India
Conflict of Interests: The authors declare that they have no competing interests.
Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
Authors' Contributions: Each author made an equal contribution to the conception and design of the study. All authors have reviewed and approved the final version of the manuscript for publication.
Transparency: The authors affirm that this manuscript presents an honest, accurate, and transparent account of the study. All essential aspects have been included, and any deviations from the original study plan have been clearly explained. The writing process strictly adhered to established ethical standards.

relationships between joints. Despite these advancements, most existing systems focus only on classification and lack the ability to provide meaningful feedback or corrections tailored to the user.

Another limitation of current approaches is the absence of contextual intelligence. While deep learning models can achieve high classification accuracy, they operate as black-box systems and do not explain why a pose is incorrect or how to improve it. This reduces user trust and limits practical usability, especially for beginners who require guided instruction.

To overcome these challenges, this work introduces a hybrid framework that combines ML-based pose recognition with Retrieval-Augmented Generation. The ML component handles pose detection and classification using extracted skeletal features, while the RAG module retrieves relevant yoga knowledge and generates human-like corrective suggestions. This integration enables the system to move beyond simple classification and provide interpretable, context-aware feedback.

The key contributions of this work are:

1) A unified ML + RAG architecture for yoga pose recognition and intelligent feedback generation.
2) Integration of pose estimation with knowledge retrieval to enhance explainability.
3) A conceptual design for real-time corrective guidance in digital fitness systems.
4) A scalable framework adaptable to mobile and edge-based deployment environments.

This approach aligns with the growing trend of combining data-driven models with knowledge-based systems to create more robust and user-centric AI solutions. By incorporating both recognition and reasoning capabilities, the proposed system aims to improve the effectiveness of automated yoga training platforms.

LITERATURE REVIEW

Yoga pose recognition has been extensively explored using computer vision and machine learning techniques. Early approaches relied on handcrafted feature extraction from images, followed by traditional classifiers such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN). These methods were limited in handling complex body variations and lacked robustness under real-world conditions.

With the advancement of deep learning, pose estimation frameworks such as OpenPose and MediaPipe enabled accurate extraction of human skeletal keypoints. Several studies utilized these frameworks to classify yoga poses using Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). While these models achieved high classification accuracy, they primarily focused on pose identification rather than correction or feedback.

Recent research has also explored temporal modeling techniques, where sequential pose data is processed using Long Short-Term Memory (LSTM) networks to capture motion patterns. This improves recognition in dynamic yoga sequences. However, such models still lack interpretability and fail to provide actionable insights to users.

Another line of work focuses on real-time fitness coaching systems using wearable sensors and vision-based tracking. These systems attempt to provide feedback by comparing user poses with predefined templates. Although effective to some extent, template-based approaches are rigid and do not adapt well to individual body differences or varying skill levels.

In parallel, the emergence of Retrieval-Augmented Generation has transformed how AI systems generate context-aware responses. RAG combines information retrieval with generative models to produce more accurate and grounded outputs. It has been widely applied in domains such as question answering, healthcare, and education, where domain knowledge is essential. However, its application in fitness and yoga guidance remains largely unexplored.

Most existing yoga recognition systems suffer from three major limitations:

1) Lack of interpretability and explainable feedback
2) Absence of domain knowledge integration
3) Limited personalization for users

To better understand the current landscape, Table 1 presents a comparative analysis of key approaches.

Table 1

Table 1 Comparative Analysis of Existing Methods

Paper	Method	Limitation
Pose Estimation + CNN	Keypoint extraction + classification	No feedback mechanism
LSTM-based Sequence Model	Temporal pose analysis	High complexity, no interpretability
Template Matching Systems	Predefined pose comparison	Rigid, not adaptive
Sensor-based Systems	Wearable data + ML	Requires hardware, costly

RAG-based Systems (general domain)	Retrieval + generation	Not applied to yoga domain

From the analysis, it is evident that while significant progress has been made in pose detection and classification, there is a clear gap in combining recognition with intelligent, knowledge-driven feedback systems. This motivates the need for a hybrid ML and RAG-based framework for yoga pose recognition and guidance.

RESEARCH GAP AND PROBLEM STATEMENT

Despite significant advancements in yoga pose recognition using Machine Learning and Computer Vision, existing systems remain largely limited to pose classification without delivering meaningful corrective guidance. Most approaches rely on frameworks such as OpenPose and MediaPipe to extract skeletal keypoints and then apply classification models. While these techniques achieve reasonable accuracy, they fail to address practical usability in real-world yoga training scenarios.

The critical research gaps identified from the literature are as follows:

1) Lack of Intelligent Feedback

Current systems can detect whether a pose is correct or incorrect but do not explain why the pose is incorrect or how to correct it. This limits user learning and reduces system effectiveness.

2) Absence of Knowledge Integration

Most ML-based models operate purely on data-driven learning and do not incorporate structured yoga knowledge such as posture alignment rules, breathing techniques, or expert recommendations.

3) Limited Personalization

Existing approaches treat all users uniformly and do not adapt feedback based on individual differences such as flexibility, posture variation, or skill level.

4) Poor Interpretability

Deep learning models function as black-box systems, making it difficult to generate explainable outputs that users can trust and follow.

5) Underutilization of RAG in Fitness Domain

Although Retrieval-Augmented Generation has shown strong performance in knowledge-driven applications, its integration with pose recognition systems for generating contextual feedback has not been adequately explored.

BASED ON THESE GAPS, THE PROBLEM CAN BE FORMALLY DEFINED AS:

There is a need to design an intelligent yoga pose recognition system that not only classifies poses accurately but also provides real-time, personalized, and context-aware corrective feedback using integrated knowledge sources.

The challenge lies in combining visual pose understanding with knowledge-based reasoning in a unified framework. The system must be capable of extracting meaningful features from human posture, evaluating correctness, retrieving relevant domain knowledge, and generating human-like guidance that is both accurate and interpretable.

To address this problem, this paper proposes a hybrid ML and RAG-based framework that bridges the gap between pose recognition and intelligent feedback generation. The proposed system aims to enhance user engagement, improve learning outcomes, and provide a scalable solution for digital yoga training platforms.

PROPOSED FRAMEWORK AND SYSTEM ARCHITECTURE

The proposed system follows a hybrid architecture that integrates Machine Learning–based pose recognition with Retrieval-Augmented Generation to provide intelligent and context-aware feedback. The framework is designed as a modular pipeline to ensure scalability, interpretability, and real-time performance.

Figure 1

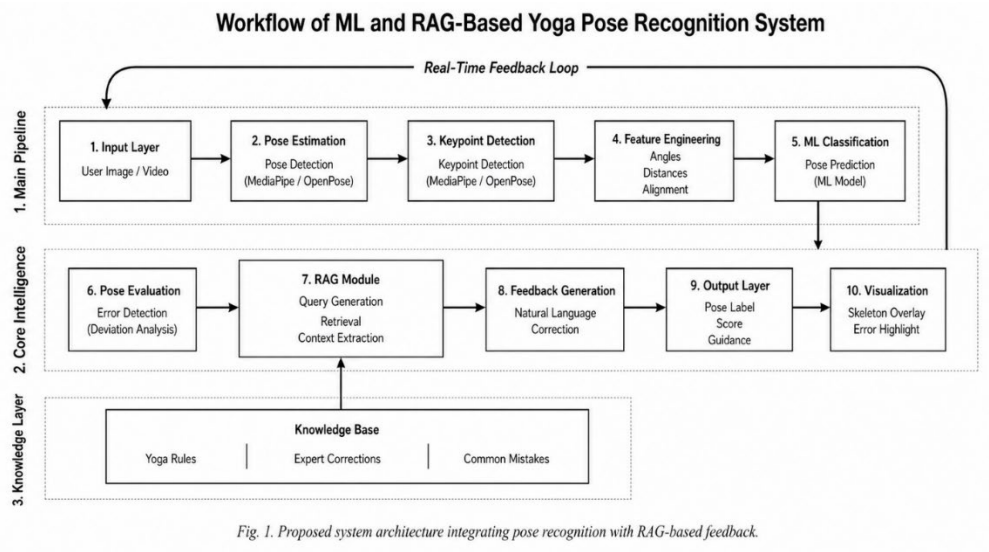

Fig. 1. Proposed system architecture integrating pose recognition with RAG-based feedback.

Figure 1 Shows the Proposed System Architecture

Overall System Flow

Input (User Video/Image)

- Pose Detection and Keypoint Extraction
- Feature Engineering
- Pose Classification (ML Model)
- Pose Evaluation Module
- RAG-Based Knowledge Retrieval
- Feedback Generation Module
- Final Output (Pose Label + Corrective Guidance)

Component Description

1) Input Layer

The system accepts user input in the form of images or real-time video streams. This allows both offline analysis and live yoga session monitoring.

2) Pose Detection and Key point Extraction

Human body landmarks are extracted using pose estimation frameworks such as Media Pipe or Open Pose. These frameworks identify key joints such as shoulders, elbows, hips, knees, and ankles, forming a skeletal representation of the human body.

3) Feature Engineering

The extracted key points are transformed into meaningful features such as:

- Joint angles (e.g., elbow angle, knee angle)
- Distance between joints
- Body alignment metrics

These features provide a structured representation for classification.

4) Pose Classification Module

A supervised ML model (e.g., Random Forest, SVM, or Neural Network) is used to classify the yoga pose based on engineered features. The output is a predicted pose label (e.g., Tree Pose, Warrior Pose).

5) Pose Evaluation Module

This module compares the user's pose with an ideal reference pose. It identifies deviations such as incorrect angles, misalignment, or imbalance. The output is an error vector representing posture deviations.

6) RAG-Based Knowledge Retrieval

The system uses a RAG pipeline to retrieve relevant information from a curated yoga knowledge base. This includes:

- Correct posture descriptions
- Common mistakes
- Expert correction guidelines

The retrieval step ensures that feedback is grounded in domain knowledge rather than generated arbitrarily.

7) Feedback Generation Module

Using the retrieved knowledge and detected pose errors, a generative model produces personalized corrective feedback. For example:

"Straighten your back and raise your left arm slightly higher to maintain balance."

8) Output Layer

The final output includes:

- Detected pose label
- Confidence score
- Detailed corrective suggestions
- Optional visual overlay for guidance

SYSTEM CHARACTERISTICS

- Real-time capable for live yoga sessions
- Explainable output through knowledge integration
- Scalable for mobile and web-based deployment
- Adaptable for different user skill levels

This architecture effectively combines visual intelligence with contextual reasoning, enabling the system to move beyond traditional classification models and deliver meaningful user-centric guidance.

MATHEMATICAL MODEL

The proposed system integrates pose estimation, classification, deviation analysis, and RAG-based feedback generation into a unified mathematical framework. The model is designed to quantify pose correctness and generate interpretable guidance.

1) Pose Feature Representation

The human pose is represented as a set of keypoints extracted from the image:

Display Format:

$P = \{p_1, p_2, p_3, \ldots, p_n\}$ … (Eq. 1)

Word Equation Format:

P = {p_1, p_2, p_3, \ldots, p_n}

Where:

P = Set of body keypoints

$p_i = (x_i, y_i)$ coordinates of the i-th joint

n = Total number of detected joints

These keypoints are further transformed into feature vectors based on joint angles and distances.

Pose Classification Function

The classification model maps feature vector F to a pose label Y:

Display Format:

$Y = f(F)$ … (Eq. 2)

Word Equation Format:

$Y = f(F)$

Where:

F = Feature vector derived from keypoints

f(.) = Trained ML classifier

Y = Predicted yoga pose class

2) Pose Deviation (Error) Calculation

The deviation between user pose and ideal pose is computed using feature difference:

Display Format:

$E = ||F_{user} - F_{ref}||$ … (Eq. 3)

Word Equation Format:

E = |F_{user} - F_{ref}|

Where:

E = Pose error magnitude

F_user = Feature vector of user pose

F_ref = Feature vector of reference (ideal) pose

This error helps identify incorrect posture elements.

3) Weighted Pose Score

To evaluate pose quality, a weighted scoring function is used:

Display Format:

$Score = \alpha_1 E_1 + \alpha_2 E_2 + \alpha_3 E_3 + \ldots + \alpha_k E_k$ … (Eq. 4)

Word Equation Format:

Score = \alpha_1 E_1 + \alpha_2 E_2 + \alpha_3 E_3 + \ldots + \alpha_k E_k

Where:

E_k = Individual error components (e.g., arm angle, leg alignment)

α_k = Weight assigned to each component

Score = Overall pose correctness score

Lower score indicates better alignment with the ideal pose.

4) RAG-Based Feedback Generation

The final feedback is generated by combining pose error and retrieved knowledge:

Display Format:

$Feedback = G(R(K, Q), E)$ … (Eq. 5)

Word Equation Format:

Feedback = G(R(K, Q), E)

Where:

K = Knowledge base (yoga instructions, corrections)

Q = Query generated from detected pose and errors

R(K, Q) = Retrieved relevant knowledge

G(.) = Generative model

E = Pose error vector

This equation ensures that feedback is both data-driven (from E) and knowledge-grounded (from K).

OVERALL INTERPRETATION

The mathematical model defines a pipeline where:

- Keypoints → Features → Classification
- Features → Error → Score
- Error + Knowledge → Feedback

This structured formulation ensures that the system is not only predictive but also explainable and interpretable.

Algorithm AND Pseudocode

The proposed algorithm integrates pose detection, classification, deviation analysis, and Retrieval-Augmented Generation-based feedback generation into a unified workflow.

Algorithm: ML + RAG-Based Yoga Pose Recognition and Feedback System

Input:

User image/video frame I

Knowledge base K (yoga rules, corrections)

Reference pose features F_ref

Output:

Predicted pose label Y

Pose score Score

Corrective feedback Feedback

Steps:

1) Acquire Input
   Capture image/frame I from camera or dataset
2) Pose Detection
   Extract keypoints P from I using pose estimation model
   $P = \{p_1, p_2, \ldots, p_n\}$
3) Feature Extraction
   Compute feature vector F from keypoints P
   (joint angles, distances, alignment metrics)
4) Pose Classification
   Apply trained ML model

$Y = f(F)$

5) Pose Deviation Calculation

Compute error vector

$E = F - F_{ref}$

6) Pose Scoring

Calculate weighted score

$Score = \Sigma (\alpha_k \times E_k)$

7) Query Formation for RAG
   - Generate query Q based on:
   - Predicted pose Y

Error vector E

Example:

"Errors in Warrior Pose: bent back, low arm height"

8) Knowledge Retrieval

Retrieve relevant documents

$D = R(K, Q)$

Feedback Generation

9) Generate corrective feedback

$Feedback = G(D, E)$

10) Output Results

Display:

- Pose label Y

- Score
- Corrective feedback

11) (Optional) Real-Time Loop

Repeat steps 1–10 for continuous video input

Algorithm Characteristics

- Modular design (each component independent)
- Supports real-time processing
- Combines deterministic (ML) and generative (RAG) logic
- Enables explainable AI through structured feedback

This algorithm ensures a seamless flow from pose detection to intelligent guidance, making the system suitable for real-world yoga training applications.

METHODOLOGY AND WORKING

The proposed system follows a structured pipeline that integrates visual pose understanding with knowledge-driven feedback generation. Since this is a conceptual framework, the methodology focuses on the logical working of each component and how they interact to produce accurate and explainable outputs.

1) Data Acquisition

The system operates on two types of inputs:

- Static images (for offline analysis)
- Real-time video streams (for live yoga sessions)

A dataset of labeled yoga poses is used to train the pose classification model. The dataset may include multiple pose categories with variations in body alignment, lighting conditions, and camera angles to improve robustness.

2) Preprocessing

The input image/frame is processed to enhance quality and consistency:

- Resize and normalize input frames
- Remove noise and background disturbances
- Standardize input format for pose estimation

This step ensures stable keypoint extraction regardless of environmental variations.

3) Pose Estimation

Human body landmarks are extracted using pose estimation models such as MediaPipe. The model identifies key joints and generates a skeletal structure representing the human posture.

4) Feature Engineering

From the extracted keypoints, meaningful features are computed:

- Joint angles (e.g., elbow, knee, hip angles)
- Relative distances between joints
- Symmetry and alignment measures

These features are normalized and converted into a structured feature vector suitable for machine learning models.

5) Model Training and Classification

A supervised ML model is trained using labeled pose data. During inference:

- The feature vector is passed to the trained model
- The system predicts the yoga pose label
- A confidence score is generated

This step forms the core recognition component of the system.

6) Pose Evaluation

The predicted pose is compared with an ideal reference pose stored in the system:

- Compute deviation for each feature

- Identify incorrect body parts (e.g., bent knee, tilted spine)
- Generate an error vector representing posture differences

This evaluation step enables the system to move beyond classification and quantify correctness.

7) RAG-Based Feedback Generation

The system integrates Retrieval-Augmented Generation to provide intelligent feedback:

- Query Formation: Convert pose errors into a structured query
- Retrieval: Fetch relevant yoga instructions and correction rules from knowledge base
- Generation: Produce natural language feedback using a generative model

This ensures that feedback is:

- Context-aware
- Knowledge-grounded
- Human-readable

8) Output Generation

The final output includes:

- Detected pose label
- Pose correctness score
- Detailed corrective suggestions

Optional features may include visual overlays highlighting incorrect joints.

9) Evaluation Strategy (Conceptual)

Since this is a conceptual framework, evaluation is defined as:

- Classification accuracy for pose recognition
- Error detection consistency
- Quality of generated feedback (relevance and clarity)

10) System Loop for Real-Time Application

For live applications:

- The system continuously processes video frames
- Feedback is updated dynamically
- Users can adjust posture in real time

Overall Working Summary

Input → Preprocessing → Pose Estimation → Feature Extraction → Classification → Error Detection → RAG Retrieval → Feedback Generation → Output

This methodology ensures that the system not only detects yoga poses but also provides meaningful and actionable guidance, making it suitable for intelligent fitness and digital health applications.

EXPECTED RESULTS AND DISCUSSION

The proposed ML and Retrieval-Augmented Generation-based yoga pose recognition system is expected to demonstrate improved functionality compared to traditional pose classification systems. Since the framework is conceptual, the results are discussed in terms of logical outcomes and system capabilities rather than numerical performance metrics.

1) Improved Pose Recognition Accuracy

By leveraging structured feature extraction from pose keypoints and supervised machine learning models, the system is expected to achieve reliable classification across multiple yoga poses. The use of normalized joint angles and alignment features enhances robustness against variations in body shape, camera angle, and lighting conditions.

2) Effective Pose Error Detection

The deviation-based evaluation mechanism enables the system to identify specific posture errors rather than providing binary correct/incorrect outputs. This fine-grained analysis allows detection of:

- Incorrect limb angles

- Body misalignment
- Imbalance in posture

Such detailed error identification is critical for real-world usability.

3) Context-Aware Feedback Generation

The integration of RAG is expected to significantly improve the quality of feedback. Unlike traditional systems that rely on predefined rules, the proposed framework retrieves domain-specific knowledge and generates human-like corrective suggestions. This leads to:

- More natural and understandable feedback
- Reduction in generic or repetitive responses
- Improved user engagement

4) Enhanced Interpretability

The system provides transparent outputs by linking detected errors with knowledge-based explanations. This improves trust and usability, especially for beginners who require guided instruction.

5) Real-Time Guidance Capability

The modular architecture supports real-time processing, allowing users to receive continuous feedback during live yoga sessions. This creates an interactive training experience similar to a human instructor.

6) Comparative Advantage Over Existing Systems

Compared to traditional ML-only approaches:

- ML-only systems → Detect pose but do not explain errors
- Proposed system → Detect pose + explain + guide correction

Compared to rule-based systems:

- Rule-based systems → Static and non-adaptive
- Proposed system → Dynamic, knowledge-driven, and adaptive

7) Practical Applicability

The system is expected to be applicable in:

- Mobile fitness applications
- Smart home workout systems
- Online yoga training platforms
- Rehabilitation and physiotherapy support

LIMITATIONS (EXPECTED)

- Performance may depend on quality of pose estimation
- RAG output quality depends on knowledge base design
- Real-time performance may require optimization for edge devices

OVERALL DISCUSSION

The proposed system shifts the paradigm from simple pose recognition to intelligent guidance systems. By combining visual intelligence with knowledge retrieval and generation, it creates a more holistic and user-centric solution. The framework is expected to enhance both learning outcomes and user experience in digital yoga platforms.

CONCLUSION AND FUTURE SCOPE

This paper presented a hybrid ML and Retrieval-Augmented Generation-based framework for yoga pose recognition and intelligent feedback generation. The proposed system addresses key limitations of existing approaches by integrating pose estimation, classification, deviation analysis, and knowledge-driven feedback into a unified architecture. Unlike traditional systems that focus only on pose detection, the proposed framework enhances interpretability by providing context-aware and personalized corrective guidance.

The integration of machine learning with retrieval-augmented generation enables the system to move beyond static classification and deliver meaningful insights to users. By leveraging structured pose features and domain-specific knowledge, the

system ensures that feedback is both accurate and human-readable. This makes the framework suitable for real-time applications in digital fitness, remote yoga training, and health monitoring systems.

From a design perspective, the modular architecture allows flexibility in model selection, knowledge base expansion, and deployment across different platforms such as mobile devices and web-based applications. The conceptual nature of the framework ensures that it can be extended and adapted to various real-world scenarios without significant structural changes.

FUTURE SCOPE

1) Several directions can be explored to enhance the proposed system:

Integration with Advanced Deep Learning Models

Future work can incorporate transformer-based architectures and advanced pose estimation techniques to improve recognition accuracy and robustness.

2) Personalized Feedback Mechanisms

User-specific parameters such as flexibility, age, and experience level can be incorporated to generate more customized guidance.

3) Multimodal Learning

Combining visual data with audio or wearable sensor data can improve system reliability and enable more comprehensive analysis.

4) Real-Time Edge Deployment

Optimizing the system for edge devices can enable low-latency processing for mobile and wearable applications.

5) Expansion of Knowledge Base

Enhancing the RAG knowledge repository with expert-curated yoga datasets can improve feedback quality and diversity.

6) Clinical and Rehabilitation Applications

The framework can be extended for physiotherapy and rehabilitation scenarios where posture correction is critical.

In summary, the proposed ML and RAG-based system provides a strong foundation for next-generation intelligent fitness applications by combining recognition, reasoning, and guidance into a single cohesive framework.

ACKNOWLEDGMENTS

None.

REFERENCES

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.

Cao, Z., Hidalgo, G., Simon, T., Wei, S., and Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(1), 172–186.

Carreira, J., and Zisserman, A. (2017). Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (6299–6308).

Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (785–794).

Cortes, C., and Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273–297.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (4171–4186).

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The PASCAL visual object classes (VOC) challenge. International Journal of Computer Vision, 88(2), 303–338.

Girshick, R. (2015). Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (1440–1448).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2014). Generative Adversarial nets. In Advances in Neural Information Processing Systems (2672–2680).

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (770–778).

He, P., Liu, W., Gao, J., and Chen, W. (2020). DeBERTa: Decoding-enhanced BERT with Disentangled Attention. arXiv preprint arXiv:2006.03654.

Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.

Kingma, D. P., and Ba, J. (2015). Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems (1097–1105).

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature, 521, 436–444.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., and Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In Advances in Neural Information Processing Systems.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. Arxiv Preprint arXiv:1301.3781.

Pavlakos, G., Zhou, X., Derpanis, K. G., and Daniilidis, K. (2017). Coarse-To-Fine Volumetric Prediction for Single-Image 3d Human Pose. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (7025–7034).

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You Only Look once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (779–788).

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149.

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. (2011). Real-Time Human Pose Recognition in Parts from Single Depth Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (1297–1304).

Simonyan, K., and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.

Toshev, A., and Szegedy, C. (2014). DeepPose: Human Pose Estimation via deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (1653–1660).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (5998–6008).

Zhang, F., Bazarevsky, V., Vakunov, A., Sung, G., Chang, C.-L., and Grundmann, M. (2019). MediaPipe: A Framework for Building Perception pipelines. arXiv preprint arXiv:1906.08172.