Karate Kata Scoring and Performance Evaluation Using Video Analysis and Deep Learning

H.L.S.S. Liyanage*, H.M.S.D. Deshpriya, M.T.Kavindya, R.M.L.D.Ranasinghe, H.M.D.P.Herath, A.K.N.L.Aththanagoda
Department of Computing, Rajarata University of Sri Lanka, Mihintale, Sri Lanka

ABSTRACT

Karate kata, a fundamental element of traditional martial arts, consists of choreographed sequences of stances, strikes, and blocks performed against imaginary opponents, requiring precise posture, timing, and technique. However, self-practice often lacks objective feedback, leading to unnoticed errors and uneven skill development, and even competition scoring can be influenced by subjective judgment. To address this, we present a vision-based scoring system that evaluates kata from structured video recordings without wearable sensors, providing accurate, consistent scores to help practitioners track progress. The system processes videos through a multi-stage pipeline: frames are extracted at 10 frames per second using OpenCV and segmented with YOLOv8-seg to remove background clutter. Pose estimation is performed with MediaPipe, generating 99 normalized 3D keypoints, while hip-centered normalization and pelvic-width scaling ensure consistency across different body sizes. A modified ResNet50 classifies frames into nine fundamental stances with high accuracy and near real-time processing (~150 ms per frame at 1080p), and a custom Convolutional Neural Network evaluates the angular accuracy of 14 critical joints using trigonometric analysis against reference poses. Each kata is decomposed into 19 key positions, with a stance marked correct only if both classification and joint angles ($\leq \pm 2.5°$ deviation) meet the threshold, and final scores are calculated on a 5.0–10.0 scale to align with traditional judging. Experimental results show that ResNet50 achieves 95.97% accuracy, while the CNN reaches 96.11%, demonstrating that this approach offers a low-cost, accessible, and consistent scoring tool. It supports remote training, reduces judging bias, and provides practitioners with reliable feedback to improve their kata performance.

KEYWORDS: *Karate Pose Classification, 3D Keypoints, CNN, ResNet50, Pose Scoring, Sports Video Analysis*

BACKGROUND AND OBJECTIVE

Karate kata encodes both the technical precision and strategic flow of martial arts practice. In competitions and training, judges typically assess kata based on stance stability, joint alignment, and smooth transitions. However, human evaluation can be inconsistent, and remote learners cannot receive an immediate, measurable score. While computer vision techniques offer scalability and non-intrusive data collection, they often suffer from limitations such as joint occlusions, depth ambiguity in monocular setups, and sensitivity to environmental variations like dynamic lighting and background clutter[1], [2]. The iKarate system uses Kinect IR and F-DTW, achieving 91.07% accuracy with 830 ms inference time in a structured lab environment[3]. Yoga Assist, designed specifically for yoga, recognizes 3 poses using OpenPose with 95% accuracy, though inference time is not reported[4]. Our methodology addresses these technical constraints

by leveraging a fully vision-based pipeline that combines MediaPipe's 3D keypoint extraction with YOLOv8-powered semantic segmentation to isolate the athlete from background noise. The objective of this research is to create an AI-driven, vision-only kata scoring system that produces consistent results without wearable devices, improving fairness and accessibility, mainly focusing on reducing depth ambiguity and occlusion problems.

https://doi.org/10.54389/YKLM7019

METHODOLOGY

Figure 1: We used a subset of the MADS dataset and custom-recorded 'Geisaki Dai Ichi' kata videos, extracting 3,600 frames (400 per class × 9 classes) at 10 FPS[5]. Backgrounds were removed using YOLOv8-seg, a deep learning segmentation model, and poses were tracked with MediaPipe, a real-time framework for detecting 2D/3D human keypoints. The client–server setup handles model inference on the server, with a mobile interface (Android Studio) supporting MP4 uploads up to 600 MB and secure authentication. Hip-centered coordinate normalization, scaled by a standard pelvic width, ensured size consistency, and data augmentation (flips, ±15° rotations) improved robustness. A fine-tuned ResNet50, a 50-layer image classification network, identified nine stance classes, while a custom Convolutional Neural Network (CNN) evaluated 14 joint angles using trigonometric analysis[6]. A stance was marked correct only if both classification and angular accuracy (≤ ±2.5°) met the threshold, with scores scaled to 5.0–10.0.

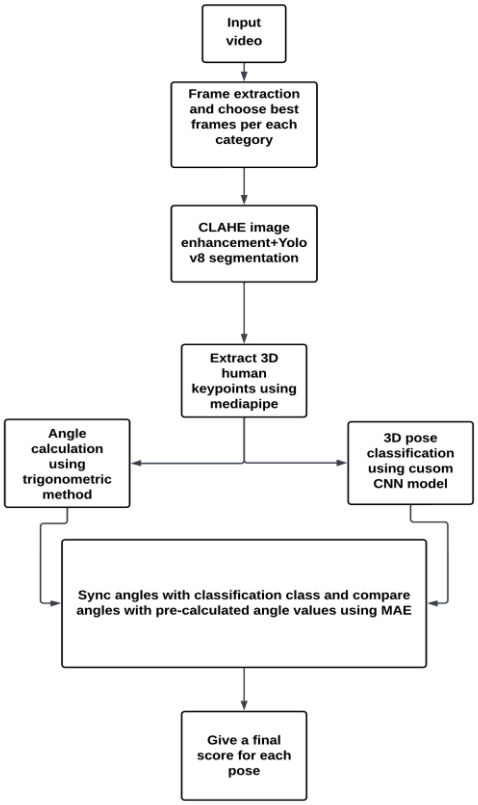

Figure 1 - Methodology high-level architecture - how models are integrated

RESULTS

Figure 2, the custom 11-layer CNN, trained on normalized 3D keypoints with regularization and preprocessing (CLAHE + YOLOv8-seg), achieved 98.16% training accuracy and 96.11% on unseen data, with a mean angular error of ±2.5°. Though it's ~578 milliseconds, inference time limits real-time use. Figure 3, the fine-tuned ResNet50, a 50-layer neural network that can classify images, trained on 3,600 frames, resized inputs (384×512), achieved 98.28% training and 95.97% validation accuracy, with ~150 ms inference per 1080p frame. Together, the models provide consistent kata scores across sessions, with hip-centered normalization and pelvic scaling reducing positional bias.

	precision	recall	f1-score	support
hachijiDachi_jodanYoko	0.90	0.99	0.94	70
sanchinDachi_ageUke	0.99	0.90	0.94	82
sanchinDachi_jodanTsuki	0.97	0.97	0.97	91
sanchinDachi_sotoUke	1.00	0.95	0.97	75
shikoDachi_gedanBarai	1.00	1.00	1.00	73
sotoUke_maeGeri	0.95	1.00	0.98	79
zenkutsuDachi_awaseTsuki	0.98	0.95	0.97	87
zenkutsuDachi_chudanTsuki	0.95	0.97	0.96	87
zenkutsuDachi_empiUke	0.99	1.00	0.99	76
accuracy			0.97	720
macro avg	0.97	0.97	0.97	720
weighted avg	0.97	0.97	0.97	720

Figure 2 - Classification report of the custom CNN used for 3D classification

Classification Report:

	precision	recall	f1-score	support
0	0.98	0.99	0.98	80
1	0.94	1.00	0.97	79
2	0.82	1.00	0.90	74
3	0.99	0.76	0.86	88
4	1.00	0.99	0.99	73
5	0.99	0.96	0.97	71
6	0.99	0.99	0.99	95
7	0.98	0.98	0.98	86
8	0.99	1.00	0.99	74
accuracy			0.96	720
macro avg	0.96	0.96	0.96	720
weighted avg	0.96	0.96	0.96	720

Figure 3 - Classification report of fine-tuned ResNet50

CONCLUSION

This research introduces a practical, vision-only kata scoring framework that bridges the gap between computer vision research and martial arts training. For academia, it contributes a reproducible pipeline that combines deep learning and biomechanical analysis in a novel application domain. Karate offers a standardized, objective scoring method that supports fairer judging, consistent athlete development, and broader access to performance evaluation tools. This work also presents a robust approach to karate pose classification and evaluation using 3D human pose detection, keypoint extraction, and CNN-based classification. By integrating MediaPipe for extracting body angles and using a comprehensive dataset, the

model achieves high accuracy in identifying and evaluating karate katas. The system facilitates feedback generation by comparing extracted angles with predefined pose standards, enabling learners to refine their techniques effectively. The inclusion of a gamified training environment enhances user engagement, making this tool valuable for martial arts training and assessment. The results demonstrate the system's potential to support athletes and instructors in achieving precision and consistency in karate practices. Future improvements will target faster on-device processing and expanded kata coverage.

REFERENCES

[1] M. Fernando, K. D. Sandaruwan, and A. M. K. B. Athapaththu, "Evaluation of Taekwondo Poomsae movements using skeleton points," *J Natl Sci Found*, vol. 52, no. 1, pp. 143–156, 2024, doi: 10.4038/jnsfsr.v52i1.11986.
[2] J. Echeverria and O. C. Santos, "Toward modeling psychomotor performance in karate combats using computer vision pose estimation," *Sensors*, vol. 21, no. 24, Dec. 2021, doi: 10.3390/s21248378.
[3] B. Emad *et al.*, "IKarate: Improving Karate Kata," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 466–473. doi: 10.1016/j.procs.2020.03.090.
[4] R. Huang, J. Wang, H. Lou, H. Lu, and B. Wang, "Miss Yoga: A Yoga Assistant Mobile Application Based on Keypoint Detection." [Online]. Available: https://github.com/gym-tutor/Miss-Yoga-App
[5] Zhang, W., Liu, Z., Zhou, L., Leung, H., & Chan, A. B. (2017). Martial arts, dancing and sports dataset: A challenging stereo and multi-view dataset for 3D human pose estimation. *Image and Vision Computing, 61*, 22–39. https://doi.org/10.1016/j.imavis.2017.02.002.
[6] J. Y. Choi, E. Ha, M. Son, J. H. Jeon, and J. W. Kim, "Human Joint Angle Estimation Using Deep Learning-Based Three-Dimensional Human Pose Estimation for Application in a Real Environment," *Sensors*, vol. 24, no. 12, Jun. 2024, doi: 10.3390/s24123823.