Enhanced Badminton Stroke Recognition Using Hybrid RGB–Skeleton Features and Ensemble Learning

Farida Asriani
Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia | Electrical Engineering Department, Universitas Jenderal Soedirman, Indonesia
faridaasriani1975@mail.ugm.ac.id

Azhari Azhari
Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia
arisn@ugm.ac.id (corresponding author)

Wahyono Wahyono
Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia
wahyo@mail.ugm.ac.id

Received: 16 October 2025 | Revised: 12 November 2025 | Accepted: 29 November 2025

 DOI: https://doi.org/10.48084/etasr.15586

ABSTRACT

Previous studies on automatic recognition of badminton strokes have shown persistent gaps, particularly uneven classification accuracy across stroke categories and suboptimal overall performance. To overcome these limitations, this study proposes a hybrid framework that combines handcrafted spatiotemporal features with a two-stage feature selection and weighted ensemble learning. From representative video frames, RGB-based descriptors (Histogram of Oriented Gradients (HOG), Histogram of Optical Flow (HOF), Motion Boundary Histogram (MBH)) and skeleton-based features (Range of Motion Index (ROMI) for spatial and Dynamic Time Warping (DTW) for temporal) are extracted. Dimensionality reduction is applied through Autoencoder compression followed by SelectKBest to preserve the most informative features. The refined hybrid features are then classified using a weighted soft voting ensemble that integrates Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), and Adaptive Boosting (AdaBoost) to enhance both accuracy and class-wise balance. Experimental evaluation on a badminton stroke dataset demonstrates the effectiveness of this approach, achieving 98.21% accuracy. The results highlight that the best-performing feature combination is ROMI, DTW, and HOF, confirming that hybrid handcrafted features with ensemble Machine Learning (ML) significantly improve robustness and stability, offering strong potential for practical applications in performance analysis, training systems, and sport analytics.

Keywords-badminton stroke; hybrid; spatiotemporal; feature selection; voting classifier; ensemble learning

I. INTRODUCTION

Human Action Recognition (HAR) has become a central topic in computer vision with applications in surveillance [1], healthcare [2], and sports [3]. In the sports domain, badminton stroke recognition is particularly challenging due to fast-paced rallies, complex body movements, and subtle technical variations. Accurate classification of strokes such as drive, overhead, and underhand is essential for athlete assessment and intelligent coaching systems [4]. Previous approaches have relied on handcrafted RGB descriptors such as Histogram of Oriented Gradients (HOG), Histogram of Optical Flow (HOF), and Motion Boundary Histogram (MBH), which capture spatial and motion dynamics but generate high-dimensional and redundant features [5]. Skeleton-based methods using pose estimation frameworks like MediaPipe or AlphaPose effectively represent human motion through joint coordinates. Nevertheless, they cannot model racket–shuttlecock interactions, resulting in inconsistent recognition accuracy among stroke classes. Sensor-based methods using RGB-D cameras [6], accelerometers [7], gyroscopes, or inertial sensors [8] have further attempted badminton stroke recognition but have generally achieved lower performance (83%–89%).

More recently, Deep Learning (DL) and multimodal frameworks have been actively explored for badminton action recognition. Network architectures such as Long Short-Term

Memory (LSTM) [9], Convolutional Neural Network (CNN) [10], and Graph Convolutional Network (GCN) [11] have been applied. In addition, multimodal datasets like VideoBadminton [12] have been developed to support the integration of multiple data types, enabling richer feature representation. Intelligent coaching systems such as DeCoach [4] further demonstrate the potential of Artificial Intelligence (AI)-driven analysis for player assessment and feedback. Moreover, advanced models such as Quantum CNN have been explored for badminton stroke recognition [13]. While these studies improved robustness, they still face limitations in achieving consistently high and balanced accuracy across all stroke classes.

Recent studies have emphasized that integrating multiple complementary feature representations can significantly improve classification robustness. A comprehensive survey on multi-view classification highlights how combining heterogeneous feature perspectives can enhance model discrimination capabilities [14]. Similarly, the combination of Machine Learning (ML), DL, and feature selection has shown improved performance in complex classification tasks such as medical image analysis [15]. These findings support the motivation of this study to fuse skeleton-based and RGB motion features within a unified ensemble framework.

Our previous study utilized skeleton-based features (Range Of Motion Index (ROMI) and Fast Dynamic Time Warping (FDTW)) with an ensemble of Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), and Adaptive Boosting (AdaBoost), achieving good accuracy but with unbalanced class-wise results due to limited visual information [16]. Skeleton and RGB modalities exhibit complementary strengths. While RGB data provide rich appearance and contextual cues that compensate for the abstract nature of skeleton representations, skeleton data, in turn, reduce the influence of illumination, background, and camera-view variations inherent in RGB-based approaches [17]. In this work, we introduce a novel hybrid framework that combines handcrafted RGB descriptors (HOG, HOF, MBH) with skeleton-based features (ROMI, Dynamic Time Warping (DTW)) and applies a two-stage feature selection (Autoencoder followed by SelectKBest) before fusion. Furthermore, an optimized weighted soft voting ensemble is used to enhance classification balance and robustness. To the best of our knowledge, this is the first badminton action recognition study that systematically integrates hybrid handcrafted features, two-stage feature selection, and optimized ensemble learning.

The main contributions of this work are:

- Hybrid RGB–skeleton feature integration: Development of a multimodal feature fusion approach that combines handcrafted RGB motion descriptors (HOG, HOF, MBH) with skeleton-based spatiotemporal features (ROMI, DTW).
- Two-stage feature selection: Application of Autoencoder compression and SelectKBest on individual feature groups before fusion, ensuring compact and highly discriminative feature representations.
- Weighted ensemble learning: Implementation of a soft voting ensemble combining SVM, LR, RF, and AdaBoost to enhance accuracy and improve prediction stability across stroke classes.
- Improved class-wise performance: The proposed framework effectively mitigates uneven accuracy among different badminton stroke categories, achieving balanced and reliable recognition.

II. RESEARCH METHODOLOGY

Figure 1 shows the proposed method, which consists of four main stages. First, the preprocessing stage is applied to RGB video data by extracting sequences of frames and selecting representative frames. Second, the feature extraction stage is conducted on RGB data using HOG, HOF, and MBH. In contrast, skeleton data are processed through MediaPipe to obtain 3D pose landmarks, which are then transformed into ROMI and DTW features. Third, a feature selection and fusion stage is performed to optimize the data representation. Finally, the classification stage employs an ensemble learning approach using RF, LR, SVM, and AdaBoost, combined through a weighted soft voting classifier to recognize badminton stroke techniques.

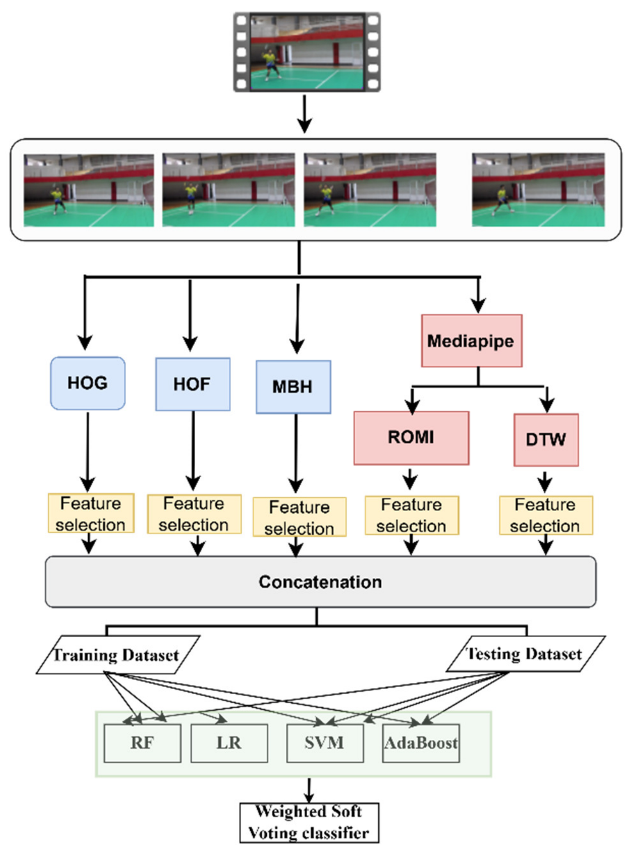

Fig. 1. Overview of the proposed method.

A. Dataset

The dataset consists of video recordings of six fundamental badminton stroke types: overhead forehand, overhead backhand, drive forehand, drive backhand, underhand forehand, and underhand backhand. A total of 18 trained badminton athletes registered under the Indonesian Badminton Association (PBSI) participated in this study, comprising both male and female players aged 13–22 years, each having a minimum of five years of structured training experience to ensure correct stroke execution.

The dataset contains 1,333 samples for training/testing and 433 samples for independent validation. The distribution of samples across the six stroke classes is balanced, with each class consisting of approximately 275–320 samples, supporting fair learning and robust generalization. To maintain consistency in stroke mechanics, only right-handed players were included in this study. The data collection was conducted in a standard indoor badminton court using the court's existing ceiling lighting without additional lighting equipment, ensuring a uniform illumination condition across sessions. A single RGB camera mounted on a tripod was placed at a height of 170 cm and positioned 70 cm to the right side of the court's center line, capturing a lateral view of the player to ensure the visibility of full-body motion and racket movement. This setup allowed clear observation of both spatial posture and temporal dynamics of each stroke.

B. Feature Extraction

Videos were decomposed into frame sequences, with 15 representative frames selected for analysis. RGB frames were used to extract HOG, HOF, and MBH features, whereas 3D skeleton coordinates obtained via MediaPipe were transformed into ROMI and DTW features.

HOG was applied to the 7th frame of each video, selected as the most representative stroke moment, after converting it to grayscale and resizing to 64 × 64 pixels. Using parameters of a 16 × 16 block size, an 8 × 8 block stride, an 8 × 8 cell size, and nine orientation bins, the method produced a 1,764-dimensional feature vector that compactly captured structural stroke patterns through gradient orientations.

HOF extracted motion features using Farneback optical flow after background subtraction. Each of 14 frame pairs from 15 sampled frames produced 48 features (6 blocks × 8 bins), resulting in 672 features per video to represent motion dynamics.

MBH utilized Farneback optical flow to compute the gradients of the horizontal (MBHx) and vertical (MBHy) components. Each of 14 frame pairs yielded 96 features (2 × 3 grid × 8 bins), producing 1,344 features per video that highlight stroke-specific motion while reducing camera effects.

ROMI, derived from MediaPipe 3D pose landmarks, uses the right hip as a reference to compute distances to 33 joints in the 7th frame, yielding 33 values that represent the body posture during a stroke.

DTW aligns skeleton joint motions by comparing ROMI sequences over 15 frames. Each of the 33 joints produced a DTW distance to a reference, resulting in a 33-dimensional vector that captures temporal movement variations for each stroke.

The 7th frame in HOG and ROMI was selected because it consistently represents the peak stroke moment, where body posture and racket orientation are most distinguishable. HOG and ROMI are posture-based spatial descriptors, and extracting them from the peak frame provides the most discriminative structural information. Meanwhile, temporal motion patterns are captured separately by HOF, MBH, and DTW, resulting in a complementary hybrid representation that balances spatial detail and temporal dynamics without redundancy.

C. Feature Selection

Feature selection was conducted in two stages: non-linear compression using an Autoencoder [18], followed by supervised selection using SelectKBest [19]. The Autoencoder generated compact latent features by minimizing the Mean Squared Error (MSE) reconstruction loss, which were then refined using SelectKBest to retain the most relevant ones. This resulted in an efficient and informative subset, thereby improving classification performance. Algorithm 1 shows the feature selection process. Autoencoder and SelectKBest are applied to each feature group independently, and the reduced vectors are concatenated afterward.

Algorithm 1. Feature selection with Autoencoder-SelectKBest

Input: Matrix $X_{scaled} \in \mathbb{R}^{n \times d}$, target $Y \in \mathbb{R}^{n \times 1}$

Output: $X_{selected} \in \mathbb{R}^{n \times k}$, trained selector

$H^{(1)} = \sigma\left(W^{(1)}X + b^{(1)}\right)$

$W^{(1)} \in \mathbb{R}^{256 \times d}$

$H^{(2)} = \sigma\left(W^{(2)}H^{(1)} + b^{(2)}\right)$

$W^{(2)} \in \mathbb{R}^{128 \times 256}$

$Z = \sigma\left(W^{(3)}H^{(2)} + b^{(3)}\right)$

$W^{(3)} \in \mathbb{R}^{\text{output\_dim} \times 128}$

$H^{(4)} = \sigma\left(W^{(4)}Z + b^{(4)}\right)$

$W^{(4)} \in \mathbb{R}^{128 \times \text{output\_dim}}$

$H^{(5)} = \sigma\left(W^{(5)}H^{(4)} + b^{(5)}\right)$

$W^{(5)} \in \mathbb{R}^{d \times 256}$

$\hat{X} = W^{(6)}H^{(5)} + b^{(6)}$

$W^{(6)} \in \mathbb{R}^{d \times 256}$

$\mathcal{L}\left(X, \hat{X}\right) = \frac{1}{n}\sum_{i=1}^{n}\left\|X_i - \hat{X}_i\right\|^2$ (MSE loss)

Use Adam(learning rate = 0.001)

$Z = f_{\text{enc}}(X)$

$X_{\text{selected}} = \text{Top-}k(Z)$

Return $X_{selected}$, selector

D. Classification Model

The classification stage used an ensemble model combining SVM [20], LR [21], RF [22], and AdaBoost through a weighted soft voting classifier. Predicted probabilities from each base model were multiplied by optimized grid-searched weights and summed for the final decision [23]. This approach integrates the strengths of linear, margin-based, tree-based, and boosting methods, yielding more robust and accurate stroke recognition than single classifiers. Figure 2 shows the ensemble model architecture, and Algorithm 2 shows the optimal weight search in the weighted soft voting classifier.

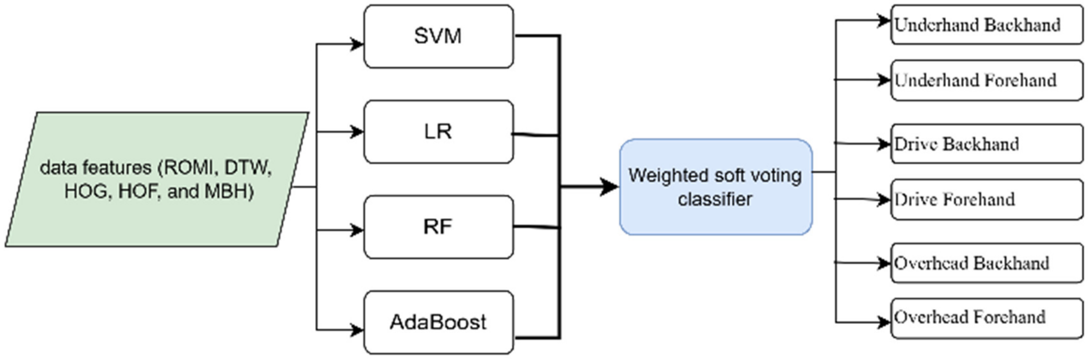

Fig. 2. Ensemble model architecture.

Algorithm 2. Optimal weight search in weighted soft voting classifier

Input: Predicted probabilities: $pred_{SVM}$, $pred_{LR}$, $pred_{RF}$, $pred_{AdaBoost}$; ground truth labels: y_{true}; weight range list: $W = \{0.0, 0.1, \dots, 1.0\}$

Output: Best weights, best accuracy, best final prediction

// Initialize
$best\_acc \leftarrow 0$
$best\_weights \leftarrow (0, 0, 0, 0)$
$best\_pred \leftarrow None$
// Weight search loop
for $w_{SVM} \in W$ do
for $w_{LR} \in W$ do
for $w_{RF} \in W$ do
for $w_{AdaBoost} \in W$ do
$total_w \leftarrow w_{SVM} + w_{LR} + w_{RF} + w_{AdaBoost}$
if $total_w = 0$ then
Proceed to the next iteration
end if
// Compute combined prediction
$combined_{pred} \leftarrow \frac{w_{SVM} \cdot pred_{SVM} + w_{LR} \cdot pred_{LR} + w_{RF} \cdot pred_{RF} + w_{AdaBoost} \cdot pred_{AdaBoost}}{total_w}$
$final\_pred \leftarrow \text{argmax}(combined\_pred, axis = 1)$
$acc \leftarrow Accuracy(final\_pred, y_{true})$
if $acc > best\_acc$ then
$best\_acc \leftarrow acc$
$best\_weights \leftarrow (w_{SVM}, w_{LR}, w_{RF}, w_{AdaBoost})$
$best\_pred \leftarrow final\_pred$
end if
end for
end for
end for
end for
return $best\_weights, best\_acc, best\_pred$

III. EXPERIMENTAL RESULTS AND DISCUSSION

Hyperparameter optimization, including the Autoencoder configuration, the value of k in SelectKBest, and the ensemble voting weights, was performed solely on the training set using internal cross-validation. The independent validation set of 433 samples was not used for any tuning process and was reserved exclusively for final performance evaluation to ensure unbiased generalization assessment.

A. Hybrid Feature Analysis

The feature selection process successfully reduced feature dimensionality, as presented in Table I, whereas the corresponding classification performance using SVM is summarized in Table II. Specifically, the dimensionality of HOG decreased from 1,764 to 64, MBH from 1,344 to 32, and ROMI and DTW from 33 to 16. This reduction not only improved computational efficiency but also lowered the risk of overfitting. In terms of classification accuracy, some features experienced a slight decline, such as ROMI (0.8392 → 0.8000) and DTW (0.8353 → 0.7560), whereas others showed improvements, such as HOG (0.6553 → 0.7217), with MBH remaining relatively stable (0.9210 → 0.8833). Meanwhile, HOF exhibited almost no change (0.9390 → 0.9372). Overall, these results demonstrate that the Autoencoder–SelectKBest strategy effectively reduced dimensionality while maintaining, and in certain cases enhancing, the discriminative capability of features for classification.

TABLE I. FEATURE DIMENSIONS BEFORE AND AFTER SELECTION

No	Feature	Feature dimension (before selection)	Feature dimension (after selection)
1	ROMI	33	16
2	DTW	33	16
3	HOG	1,764	64
4	HOF	672	64
5	MBH	1,344	32

TABLE II. ACCURACY BEFORE AND AFTER FEATURE SELECTION

No	Feature	Accuracy (before selection)	Accuracy (after selection)
1	ROMI	0.8392	0.8
2	DTW	0.8353	0.756
3	HOG	0.6553	0.7217
4	HOF	0.939	0.9372
5	MBH	0.921	0.8833

In this experiment, hybrid features were constructed through a concatenation strategy, combining different sets of RGB-based descriptors (HOG, HOF, MBH) and skeleton-based features (ROMI, DTW). Several variations of hybrid combinations were evaluated using an SVM classifier to investigate the most discriminative representation for badminton stroke recognition. As shown in Table II, the performance varied across feature combinations, with accuracy ranging from 0.7520 to 0.9663. Notably, as shown in Table III, the concatenation of ROMI, DTW, and HOF achieved the best result, reaching an accuracy of 0.9663 with a relatively low total loss of 0.3871 and only 96 feature dimensions. This indicates that integrating skeleton-based temporal information with motion-based RGB descriptors provides a more complementary and compact feature representation. In contrast, larger concatenations, such as ROMI, DTW, HOG, HOF, and MBH (with 192 dimensions), did not yield improvements and even suffered from higher loss, suggesting redundancy and reduced discriminative power. Therefore, the ROMI–DTW–HOF combination can be considered the most effective hybrid configuration, as it balances dimensionality efficiency, loss reduction, and recognition accuracy.

TABLE III. PERFORMANCE OF HYBRID RGB–SKELETON FEATURES WITH SVM

No	Hybrid feature	Feature dimension	Total loss	Accuracy
1	ROMI, DTW	32	0.0189	0.9405
2	HOG, HOF	128	0.4804	0.8887
3	HOG, MBH	96	0.301	0.8312
4	HOF, MBH	96	0.557	0.9336
5	HOG, HOF, MBH	160	0.6692	0.9102
6	ROMI, DTW, HOG	96	0.1311	0.752
7	**ROMI, DTW, HOF**	**96**	**0.3871**	**0.9663**
8	ROMI, DTW, MBH	64	0.2077	0.907
9	ROMI, DTW, HOG, HOF	160	0.4993	0.8849
10	ROMI, DTW, HOG, MBH	128	0.3199	0.8591
11	ROMI, DTW, HOG, HOF, MBH	192	0.6881	0.9167

B. *Ensemble Model Analysis*

As shown in Table IV, each model demonstrates distinct characteristics. SVM achieved consistently high accuracy (0.9603) with stable and well-distributed predictions, reflecting its strength in optimal margin separation but sensitivity to kernel parameters. LR achieved the highest individual accuracy (0.9663), yet exhibited uneven error distribution, particularly in misclassifying the drive forehand class. RF maintained stable performance (0.9583) and robustness to non-linear data, though its predictions were less smooth for complex patterns. AdaBoost performed the weakest (0.8591) due to its sensitivity to outliers and noise.

To verify that the performance improvement of the Ensemble ML model over the best single classifier (LR) was not due to random variation, the McNemar's test was conducted using the class-wise predictions on the held-out validation set. The test indicated a statistically significant difference ($p < 0.05$), confirming that the performance gain of the proposed ensemble approach is meaningful and not attributable to chance.

TABLE IV. EVALUATION OF ML AND ENSEMBLE ML MODELS

Model	Accuracy	Precision	Recall	F1-score
SVM	0.9603	0.9621	0.9603	0.9599
LR	0.9663	0.9684	0.9663	0.9663
RF	0.9583	0.9599	0.9583	0.958
AdaBoost	0.8591	0.8643	0.8591	0.8581
Ensemble ML	**0.9821**	**0.9825**	**0.9833**	**0.9828**

According to Table V, the ensemble model assigned the highest weights to SVM (0.48) and RF (0.44), highlighting their strong contribution to the final prediction. In contrast, LR received a smaller weight (0.08) due to its higher error variance, and AdaBoost was assigned a weight of 0.0 because its predictions exhibited higher variance and reduced ensemble stability. As a result, the weighted optimization excluded AdaBoost to maximize the overall accuracy. This balanced weighting explains the superior performance of the Ensemble ML model (0.9821), which effectively integrates the strengths of the individual classifiers.

TABLE V. WEIGHT DISTRIBUTION IN THE ENSEMBLE MODEL

Model	Weight
SVM	0.4800
LR	0.0800
RF	0.4400
AdaBoost	0.0000

A brief examination of the ensemble model's weighted computational complexity indicates it is mostly influenced by the quadratic SVM and log-linear RF components. An Intel i7 CPU operating at 3.0 GHz with 16 GB of RAM requires around 0.55 s to process a 3-s video (30 fps, 15 frames) in its entirety. The proposed system remains computationally efficient and capable of near real-time feedback, with an accuracy rate of 0.9821 (98.21%).

The Ensemble ML model, as illustrated in Figure 3, demonstrates the most robust and balanced performance across all stroke classes, achieving near-perfect accuracies ranging from 95.1% to 100%. Unlike the individual models, which exhibit weaknesses such as uneven error distribution in LR or reduced accuracy in RF for specific classes, the Ensemble ML effectively integrates their complementary strengths. By combining the margin-based stability of SVM with the non-linear adaptability of RF, while minimizing the weaknesses of LR and eliminating the negative contribution of AdaBoost, the ensemble achieves superior predictive accuracy and consistency. This confirms that ensemble learning not only enhances overall accuracy but also ensures fairer error distribution, making it the most reliable approach for badminton stroke classification.

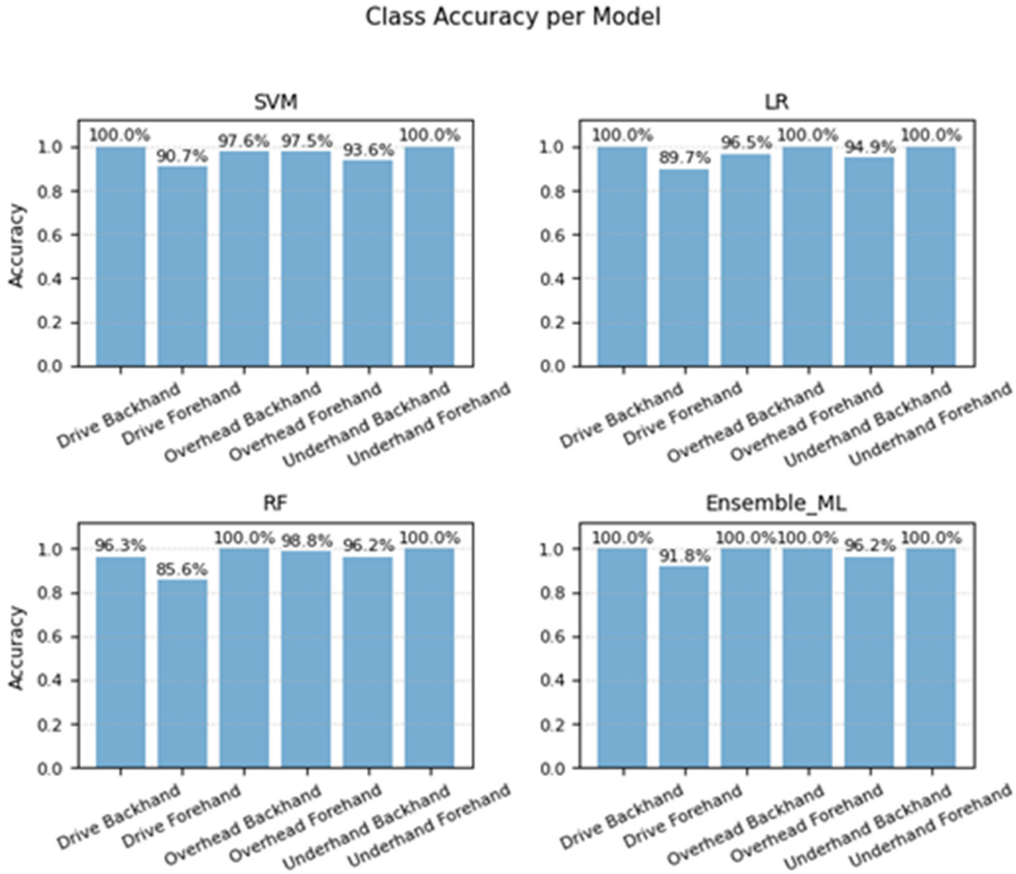

Fig. 3. Class-wise accuracy of SVM, LR, RF, and Ensemble ML.

IV. COMPARATIVE ANALYSIS OF RESULTS

Table VI summarizes the performance of the proposed method compared to previous approaches.

TABLE VI. PERFORMANCE COMPARISON OF BADMINTON STROKE RECOGNITION METHODS

No	Method	Feature	Accuracy
1	**Ensemble ML**	**Spatiotemporal 3D skeleton + RGB**	**0.9821**
2	E2 (weighted ensemble SVM–LR–AdaBoost) [16]	Spatiotemporal 3D skeleton	0.9538
3	LSTM [24]	Skeleton data extracted by AlphaPose	0.8000
4	CNN [24]	Skeleton data extracted by AlphaPose	0.6000
5	GCN [11]	Human skeleton sequence	0.9200
6	SVM [25]	3D skeleton (RGB-D sensors)	0.9200
7	SVM [26]	Accelerometer and gyroscope	0.8889
8	SVM [8]	Inertial sensor	0.8340

Here, E2 refers to our earlier weighted ensemble model (SVM–LR–AdaBoost) [16], which was designed using only spatiotemporal skeleton features. Methods relying solely on skeleton data or AlphaPose-based DL models generally achieve lower accuracies (60%–80%), whereas sensor-based systems show moderate performance (83%–88%) but are less practical because they require multiple wearable devices and restrict natural player movement. In contrast, the proposed hybrid representation, which fuses RGB motion descriptors (HOG, HOF, MBH) with spatiotemporal skeleton features (ROMI, DTW) integrates them through a weighted soft voting ensemble, achieves 98.21% accuracy. This not only surpasses previous methods but also provides more stable performance across stroke classes, indicating stronger robustness and generalizability compared to single-modality or sensor-dependent methods.

V. CONCLUSION

This study successfully developed a badminton stroke recognition model based on visual data using a hybrid spatiotemporal feature approach combined with ensemble learning. The main findings are:

A hybrid representation was established by integrating handcrafted RGB features (Histogram of Oriented Gradients (HOG), Histogram of Optical Flow (HOF), Motion Boundary Histogram (MBH)) with visual–skeleton features (Range of Motion Index (ROMI), Dynamic Time Warping (DTW)). This integration enriched spatial and temporal information, effectively addressing uneven class accuracies across stroke types.

An Ensemble Machine Learning (Ensemble ML) approach combining Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), and Adaptive Boosting

(AdaBoost) through a weighted soft voting classifier was implemented, achieving a notable accuracy of 98.21%.

This study was conducted in a controlled indoor court environment with fixed lighting and a single lateral camera view. While this setup ensured consistent motion capture quality, it may limit applicability in more dynamic settings such as multi-camera match broadcasts or outdoor environments. In addition, only right-handed athletes were included to maintain stroke consistency, which may reduce generalizability to left-handed players. Future work will address this by applying mirrored skeleton augmentation and expanding data collection to include left-handed athletes.

ACKNOWLEDGMENT

This work is supported by Indonesian Education Scholarship (BPI), Center for Higher Education Funding and Assessment, Ministry of Higher Education, Science, and Technology of Republic Indonesia (PPPAPT), and Indonesia Endowment Fund for Education (LPDP), as acknowledged in decree number 03664/BPPT/BPI.06/9/2024.

REFERENCES

[1] N. Gupta and B. B. Agarwal, "Recognition of Suspicious Human Activity in Video Surveillance: A Review," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10529–10534, Apr. 2023, https://doi.org/10.48084/etasr.5739.

[2] S. A. Alshammari and N. S. Albalawi, "Enhancing Healthcare Monitoring: A Deep Learning Approach to Human Activity Recognition using Wearable Sensors," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18843–18848, Dec. 2024, https://doi.org/10.48084/etasr.9255.

[3] C. Ma, D. Yu, and H. Feng, "Recognition of Badminton Shot Action Based on the Improved Hidden Markov Model," *Journal of Healthcare Engineering*, vol. 2021, no. 1, Oct. 2021, Art. no. 7892902, https://doi.org/10.1155/2021/7892902.

[4] I. Ghosh, S. Ramasamy Ramamurthy, A. Chakma, and N. Roy, "DeCoach: Deep Learning-based Coaching for Badminton Player Assessment," *Pervasive and Mobile Computing*, vol. 83, July 2022, Art. no. 101608, https://doi.org/10.1016/j.pmcj.2022.101608.

[5] J. Yang, Z. Shi, and Z. Wu, "Vision-based action recognition of construction workers using dense trajectories," *Advanced Engineering Informatics*, vol. 30, no. 3, pp. 327–336, Aug. 2016, https://doi.org/10.1016/j.aei.2016.04.009.

[6] H. Y. Ting, K. S. Sim, and F. S. Abas, "Automatic Badminton Action Recognition Using RGB-D Sensor," in *3rd International Conference on Key Engineering Materials and Computer Science*, Singapore, 2014, vol. 1042, pp. 89–93, https://doi.org/10.4028/www.scientific.net/AMR.1042.89.

[7] T. Steels, B. Van Herbruggen, J. Fontaine, T. De Pessemier, D. Plets, and E. De Poorter, "Badminton Activity Recognition Using Accelerometer Data," *Sensors*, vol. 20, no. 17, Sept. 2020, Art. no. 4685, https://doi.org/10.3390/s20174685.

[8] N. F. Ghazali, N. Shahar, and M. A. As'Ari, "Badminton Strokes Recognition using Inertial Sensor and Machine Learning Approach," in *2022 2nd International Conference on Intelligent Cybernetics Technology & Applications*, Bandung, Indonesia, 2022, pp. 1–5, https://doi.org/10.1109/ICICyTA57421.2022.10037897.

[9] B. Purnama, B. Erfianto, and I. R. Wirawan, "Time Series Classification of Badminton Pose using LSTM with Landmark Tracking," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 1, pp. 27–37, Nov. 2024, https://doi.org/10.35882/jeeemi.v7i1.488.

[10] Y. Wang, W. Fang, J. Ma, X. Li, and A. Zhong, "Automatic Badminton Action Recognition Using CNN with Adaptive Feature Extraction on Sensor Data," in *15th International Conference on Intelligent Computing*, Nanchang, China, 2019, pp. 131–143, https://doi.org/10.1007/978-3-030-26763-6_13.

[11] J. Liu and B. Liang, "An Action Recognition Technology for Badminton Players Using Deep Learning," *Mobile Information Systems*, vol. 2022, no. 1, May 2022, Art. no. 3413584, https://doi.org/10.1155/2022/3413584.

[12] Q. Li, T.-C. Chiu, H.-W. Huang, M.-T. Sun, and W.-S. Ku, "VideoBadminton: A Video Dataset for Badminton Action Recognition," in *2024 IEEE International Conference on Big Data*, Washington, DC, USA, 2024, pp. 1387–1392, https://doi.org/10.1109/BigData62323.2024.10825009.

[13] X. Zhu, L. Liu, J. Huang, G. Chen, X. Ling, and Y. Chen, "The analysis of motion recognition model for badminton player movements using machine learning," *Scientific Reports*, vol. 15, no. 1, May 2025, Art. no. 19030, https://doi.org/10.1038/s41598-025-02771-9.

[14] K. Berahmand, F. Daneshfar, M. Rahmaninia, M. Haghighat, and M. Jalili, "A Comprehensive Survey on Multi-View Classification: Methods, Applications, and Challenges," *ACM Transactions on Intelligent Systems and Technology*, vol. 16, no. 6, Nov. 2025, Art. no. 144, https://doi.org/10.1145/3767728.

[15] H. H. S. Junaid, F. Daneshfar, and M. A. Mohammad, "Automatic colorectal cancer detection using machine learning and deep learning based on feature selection in histopathological images," *Biomedical Signal Processing and Control*, vol. 107, Sept. 2025, Art. no. 107866, https://doi.org/10.1016/j.bspc.2025.107866.

[16] F. Asriani, A. Azhari, and W. Wahyono, "Improving Badminton Action Recognition Using Spatio-Temporal Analysis and a Weighted Ensemble Learning Model," *Computers, Materials & Continua*, vol. 81, no. 2, pp. 3079–3096, Nov. 2024, https://doi.org/10.32604/cmc.2024.058193.

[17] R. Yue, Z. Tian, and S. Du, "Action recognition based on RGB and skeleton data sets: A survey," *Neurocomputing*, vol. 512, pp. 287–306, Nov. 2022, https://doi.org/10.1016/j.neucom.2022.09.071.

[18] K. Berahmand, F. Daneshfar, E. S. Salehi, Y. Li, and Y. Xu, "Autoencoders and their applications in machine learning: a survey," *Artificial Intelligence Review*, vol. 57, no. 2, Feb. 2024, Art. no. 28, https://doi.org/10.1007/s10462-023-10662-6.

[19] R. Nair and A. Bhagat, "Feature Selection Method To Improve The Accuracy of Classification Algorithm," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 6, pp. 124–127, Apr. 2019.

[20] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sept. 1995, https://doi.org/10.1007/BF00994018.

[21] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An Introduction to Logistic Regression Analysis and Reporting," *The Journal of Educational Research*, vol. 96, no. 1, pp. 3–14, Sept. 2002, https://doi.org/10.1080/00220670209598786.

[22] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, https://doi.org/10.1023/A:1010933404324.

[23] M. S. Toor *et al.*, "An Optimized Weighted-Voting-Based Ensemble Learning Approach for Fake News Classification," *Mathematics*, vol. 13, no. 3, Feb. 2025, Art. no. 449, https://doi.org/10.3390/math13030449.

[24] Z. Liang and T. E. Nyamasvisva, "Badminton Action Classification Based on Human Skeleton Data Extracted by AlphaPose," in *2023 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence*, Xi'an, China, 2023, pp. 1–4, https://doi.org/10.1109/ICSMD60522.2023.10490491.

[25] Md. A. I. Anik, M. Hassan, H. Mahmud, and Md. K. Hasan, "Activity recognition of a badminton game through accelerometer and gyroscope," in *2016 19th International Conference on Computer and Information Technology*, Dhaka, Bangladesh, 2016, pp. 213–217, https://doi.org/10.1109/ICCITECHN.2016.7860197.