Journal of Machine Intelligence in Healthcare

E-ISSN: 3090-9503
P-ISSN: 3090-9511

(Research) Article

Hybrid AlexNet–Tabnet with Lyrebird Feature Selection for Injury Prevention and Athlete Health Monitoring

Md. Shamim Hossain [1] and Md Samiul Bashir [2,*]

[1] Institute of Public Health, Dhaka, Bangladesh; e-mail : shamim.h@iph.gov.bd
[2] Institute of Health Technology, Dhaka, Bangladesh; e-mail : mtsamiulbashir@ihtdhaka.gov.bd
* Corresponding Author : Md Samiul Bashir

Abstract: Sports healthcare increasingly relies on intelligent motion analysis to monitor athlete performance, identify risky movements, and prevent injuries. This research focuses on athlete motion data gathered from wearable sensors, which record multidimensional signals such as acceleration, angular velocity, and joint kinematics. The goal of this study is to develop a real-time, interpretable motion classification framework that can accurately distinguish biomechanically similar movements, like jump versus kick, which earlier models often misclassify. To accomplish this, we propose a hybrid approach combining AlexNet for spatial feature extraction, TabNet for attention-based interpretability, discrete wavelet transform for time–frequency analysis, and the Lyrebird Optimization Algorithm for feature selection. Experiments on the 6G-SDN Sports Motion Dataset demonstrate that the framework achieves 98.85% accuracy, 98.60% precision, 98.61% sensitivity, and 98.65% F1-score, outperforming CNN-only, TabNet-only, and LSTM baselines by 2.7–4.6%. Interpretability analysis highlights ankle angular velocity and knee joint angle as key predictors, aligning with sports medicine research on anterior cruciate ligament (ACL) strain and lower-limb injury risk. Overall, the hybrid model offers state-of-the-art classification performance while delivering biomechanically meaningful insights, proving its value as a real-time healthcare tool for injury prevention, athlete monitoring, and rehabilitation support.

Keywords: AlexNet; Athlete health monitoring; Injury prevention; Lyrebird Optimization Algorithm; Motion classification; Sports healthcare; TabNet; Wearable sensors

1. Introduction

Sports healthcare increasingly relies on intelligent motion analysis to monitor athlete performance, identify risky movement patterns, and prevent injuries. Wearable sensors provide continuous, multidimensional data streams including acceleration, joint angles, angular velocity, and posture that can be leveraged for data-driven decision support. Prior studies have explored a wide range of approaches, from conventional statistical learning (e.g., k-nearest neighbors, support vector machines) to deep learning models such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) for activity recognition and performance assessment [1]–[5]. While these approaches achieved notable success, challenges persist, including sensitivity to noise, insufficient interpretability, computational inefficiency, and reduced accuracy when discriminating between biomechanically similar actions [6], [7].

CNN-based methods, such as AlexNet, excel in extracting spatial representations but are limited in processing heterogeneous and tabular sensor data. Conversely, attention-based models like TabNet offer improved interpretability and feature selection yet underperform in capturing complex spatial–temporal dependencies [8], [9]. Hybrid frameworks have been proposed to combine multiple paradigms; however, they often suffer from feature redundancy and unstable generalization, particularly when applied to overlapping sports motions, such as jump versus kick [10], [11]. Recent advances have focused on multimodal fusion of IMU and video data [12], interpretable deep learning architectures for biomechanics [13], explainable AI for athlete monitoring [14], and lightweight deployment strategies

Received: June 3, 2025
Revised: July 17, 2025
Accepted: August 22, 2025
Published: August 31, 2025
Curr. Ver.: August 31, 2025

optimized for edge devices [15], [16]. These works highlight the need for models that balance accuracy, interpretability, and computational efficiency to ensure practical adoption in healthcare contexts.

The research problem addressed in this paper is how to develop a real-time and accurate motion analysis framework that can overcome these limitations, specifically by improving robustness in distinguishing biomechanically similar movements, while ensuring healthcare applicability for injury prevention and athlete monitoring. We hypothesize that a hybrid framework integrating CNN spatial features, TabNet attention-based feature processing, and Lyrebird Optimization Algorithm (LOA)-based feature selection, augmented with wavelet-based time–frequency representation, will achieve state-of-the-art accuracy while maintaining interpretability and computational efficiency.

The contributions of this study are as follows:

1. Development of a hybrid AlexNet–TabNet framework enhanced with LOA for feature selection and wavelet-based representation;
2. Empirical evaluation on the 6G-SDN Sports Motion Dataset with wearable IMU data, achieving state-of-the-art performance;
3. Demonstration of the framework's utility for healthcare applications, particularly in injury prevention and athlete health monitoring;
4. Comprehensive comparison against baseline models (LSTM, CNN-only, TabNet-only), highlighting the advantages of the proposed hybrid approach.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the proposed methodology, including data preprocessing, feature extraction, and model design. Section 4 presents experimental results and discussion. Section 5 concludes with implications for sports healthcare and directions for future research.

2. Related Work

Sports motion analysis has been extensively studied across the disciplines of computer vision, machine learning, and healthcare due to its potential to facilitate athlete monitoring and injury prevention. Early research primarily employed classical machine learning techniques such as k-nearest neighbors, decision trees, and support vector machines for activity recognition [17], [2]. While these methods offered computational simplicity, they were limited in robustness to noise, failed to capture complex multi-joint dynamics adequately, and lacked interpretability in clinical contexts.

The advent of deep learning has led to significant advancements. Convolutional neural networks (CNNs), exemplified by AlexNet, have demonstrated strong capabilities in extracting spatial features from motion signals and visual inputs [3], [4]. Sequential architectures such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have improved temporal pattern recognition but encountered issues related to vanishing gradients, high computational demands, and scalability complications when processing high-frequency sensor data [5]. More recently, transformer-based architectures have achieved state-of-the-art performance in modeling long-range dependencies [1], [2]. Nevertheless, their limited interpretability and computational requirements restrict their applicability in healthcare settings where transparency and real-time responsiveness are paramount [6], [7].

Complementary to these developments, attention-based methods such as TabNet have emerged as promising alternatives for handling tabular and heterogeneous data [8], [9]. TabNet offers interpretability through feature-level attention masks, allowing practitioners to understand the model's decisions, an essential feature for healthcare adoption. However, TabNet alone lacks the capacity to capture the spatial structures inherent in sports motions, thus limiting its standalone effectiveness.

Hybrid frameworks have been explored to combine deep learning with feature engineering and optimization techniques. Discrete wavelet transforms (DWT) have been utilized to generate time–frequency representations of motion data, while metaheuristic optimization algorithms, such as genetic algorithms and particle swarm optimization, have been applied for feature selection [10]. Although these approaches have improved accuracy, they often introduced redundancy within the feature space and faced challenges in generalizing across diverse and overlapping movements [12]. Moreover, prior research rarely addresses the fine-grained classification of biomechanically similar actions (e.g., jump versus kick), which is highly relevant to injury risk assessment in sports medicine [13], [14].

Several key gaps remain evident within this body of work. CNNs excel in spatial representation but are insufficient for handling multimodal sensor data; TabNet enhances interpretability but does not model spatial–temporal dynamics; optimization-based feature selection can be unstable in high-dimensional scenarios; and critically, few studies explicitly link sports motion recognition with healthcare outcomes such as injury prevention, load management, or rehabilitation monitoring [15], [16]. These gaps highlight the need for an integrated framework that strikes a balance between accuracy, interpretability, and efficiency, while maintaining direct relevance to healthcare.

Motivated by these considerations, the present study proposes a hybrid framework that integrates CNNs (AlexNet) for spatial feature extraction, TabNet for interpretable tabular data processing, wavelet transforms for time–frequency representation, and the Lyrebird Optimization Algorithm (LOA) for feature selection. Unlike previous methodologies, this approach aims to deliver state-of-the-art classification performance while providing biomechanically interpretable outputs that can inform sports healthcare practices. The detailed methodology is elaborated in Section 3.

3. Proposed Method

Building on the limitations identified in prior approaches, this section presents the proposed hybrid framework designed to deliver accurate, efficient, and interpretable sports motion classification with direct healthcare relevance. The framework integrates four complementary components: (i) AlexNet for spatial feature extraction, (ii) TabNet for interpretable tabular processing, (iii) discrete wavelet transform (DWT) for robust time–frequency representation, and (iv) the Lyrebird Optimization Algorithm (LOA) for feature selection. The overall workflow is shown in Algorithm 1.

3.1. Algorithm Description

The workflow of the proposed hybrid model is summarized in Algorithm 1. Raw motion signals are first preprocessed to remove outliers and normalized to zero mean and unit variance. Time–frequency features are extracted using DWT, followed by feature selection using LOA. AlexNet processes spatial representations from multi-channel signals, while TabNet processes structured tabular features. Their embeddings are fused via a weighted coefficient, and the final classification is obtained using fully connected layers with cross-entropy loss.

Algorithm 1. Hybrid AlexNet–TabNet with LOA for Sports Motion Analysis

INPUT: Wearable sensor data X = {acceleration, angular velocity, joint angles, posture}
OUTPUT: Classified sports motions with injury risk indicators

1: Preprocess input signals: remove outliers, impute missing data, normalize to zero mean and unit variance;
2: Apply DWT to obtain multi-resolution features;
3: Initialize LOA population with random feature subsets;
4: For each candidate solution:
  a. Evaluate subset with cross-validation accuracy;
  b. Update solution using LOA position-update rules;
5: Select optimal feature subset S^*;
6: Train AlexNet on spatial features → output embeddings f_{cnn};
7: Train TabNet on tabular features → output embeddings f_{tab};
8: Fuse embeddings: $f_{fusion} = \lambda f_{cnn} + (1 - \lambda) f_{tab}$;
9: Classify using fully connected layers with cross-entropy loss;
10: Output class label and performance metrics (accuracy, precision, sensitivity).

3.2. Mathematical Formulation

The wavelet transform decomposes motion signals into time–frequency representations:

$$\boldsymbol{W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi\left(\frac{t - b}{a}\right) dt,} \tag{1}$$

where a and b represent scale and translation parameters, and ψ is the mother wavelet. Eq. (1) enables capturing motion features at multiple resolutions.

The LOA selects features by maximizing accuracy and minimizing redundancy:

$$\boldsymbol{F(S) = \alpha \cdot Acc(S) - \beta \cdot \frac{|S|}{|F|}}, \tag{2}$$

where S is the selected feature subset, $|S|$ its size, and $|F|$ the full feature set.

The fused embedding is represented as:

$$\boldsymbol{f_{fusion} = \lambda f_{cnn} + (1 - \lambda) f_{tab}} \tag{3}$$

Finally, the hybrid classifier is trained using cross-entropy loss:

$$\boldsymbol{L = -\sum_{i=1}^{N} y_i \log \hat{y}_i}, \tag{4}$$

where y_i and $\hat{y}_i$ denote true and predicted labels.

3.3. Hyperparameter Settings

Hyperparameters were determined via grid search and empirical tuning. Table 1 summarizes the final configuration. Learning rate was initialized at 0.001 and decayed by 0.1 with a patience of 5 epochs. Training used SGD (momentum = 0.9, weight decay = 1e-4). Early stopping with patience = 10 prevented overfitting, with a maximum of 100 epochs. Dropout of 0.5 was applied after fusion and fully connected layers. Fusion coefficient λ was initialized at 0.6 and adaptively updated during training. Daubechies-4 wavelets were chosen after comparison with Haar and Symlet. LOA used a population size of 30 and 100 iterations, typically converging within 70 iterations.

Table 1. Hyperparameter configuration of the proposed framework.

Parameter	Value	Notes
Learning rate	0.001 (decay ×0.1)	Reduced on plateau with patience = 5 epochs
Optimizer	SGD (momentum = 0.9, wd=1e-4)	Ensures stable convergence
Batch size	64	Empirically optimal for GPU memory and stability
Number of epochs	Max 100 (early stopping=10)	Patience = 10, prevents overfitting
Dropout rate	0.5	Applied after fusion and FC layers
Fusion coefficient (λ)	0.6 (adaptive)	Initialized at 0.6, updated by backpropagation
Wavelet function	Daubechies-4	Outperformed Haar and Symlet in capturing short/long motion features
LOA population size	30	Balanced exploration vs exploitation
LOA iterations	100	Converged within ~70 iterations

3.4. Justification of Method Selection

The selection of AlexNet, TabNet, and LOA is deliberate. AlexNet is computationally efficient and effective for capturing spatial representations from multichannel motion signals, suitable for real-time healthcare applications where low latency is critical. TabNet provides interpretability through attention masks, enabling the identification of key motion features that contribute to classification decisions, essential for sports healthcare, where explainability supports clinical trust. LOA was chosen over conventional feature selection methods such as particle swarm optimization and genetic algorithms because of its ability to balance exploration and exploitation, improving convergence and stability in high-dimensional feature spaces. The combination of these three components creates a framework that is not only accurate but also interpretable and efficient, directly addressing the gaps of previous works in terms of feature redundancy, limited interpretability, and weak links to healthcare outcomes.

4. Results and Discussion

This section presents the experimental evaluation of the proposed hybrid AlexNet–TabNet framework with LOA feature selection. Results are reported in terms of quantitative performance, robustness across classes, interpretability, and comparative analysis with baseline models.

4.1. Experimental Environment

All experiments were conducted on a workstation with an Intel Core i9-12900K CPU, 64 GB RAM, and an NVIDIA RTX 3090 GPU (24 GB VRAM), running Ubuntu 22.04 LTS. The framework was implemented in Python 3.10 using PyTorch 2.0 for deep learning modules, Scikit-learn for evaluation metrics, and PyWavelets for wavelet feature extraction.

4.2. Dataset and Initial Analysis

We employed the 6G-SDN Sports Motion Dataset, which contains IMU signals collected from ankle, knee, and hip joints of 15 athletes. The dataset comprises 12,000 samples evenly distributed across four classes: sprint, dribble, jump, and kick. Table 2 summarizes the class distribution, confirming that the dataset is balanced and suitable for fair evaluation.

Table 2. Distribution of motion samples.

Class	Sample	Percentage (%)
Sprint	3,200	26.7
Dribble	2,800	23.3
Jump	3,000	25.0
Kick	3,000	25.0
Total	12,000	100.0

4.3. Evaluation Metrics

Performance was assessed using accuracy, precision, sensitivity (recall), F1-score, and Cohen's Kappa coefficient, alongside False Positive Rate (FPR) and False Negative Rate (FNR). Metrics were defined as Eq. (5-8):

$$\boldsymbol{Accuracy = \frac{TP + TN}{TP + FP + TN + FN}} \tag{5}$$

$$\boldsymbol{Precision = \frac{TP}{TP + FP}} \tag{6}$$

$$\boldsymbol{Sensitifity = \frac{TP}{TP + FN}} \tag{7}$$

$$\boldsymbol{F1 = 2 \cdot \frac{Precision \cdot Sensitifity}{Precision + Sensitifity}} \tag{8}$$

These metrics collectively ensure balanced evaluation of model robustness across classes.

4.4. Quantitative Results

The hybrid AlexNet–TabNet model with LOA demonstrated exceptional performance, achieving an overall accuracy of 98.85%, precision of 98.60%, sensitivity of 98.61%, F1-score of 98.60%, and a Kappa coefficient of 0.9842. These metrics reflect balanced classification performance with minimal bias between positive and negative classes (Fig. 2). The high Kappa value underscores the model's strong alignment with the ground truth, highlighting its robustness.

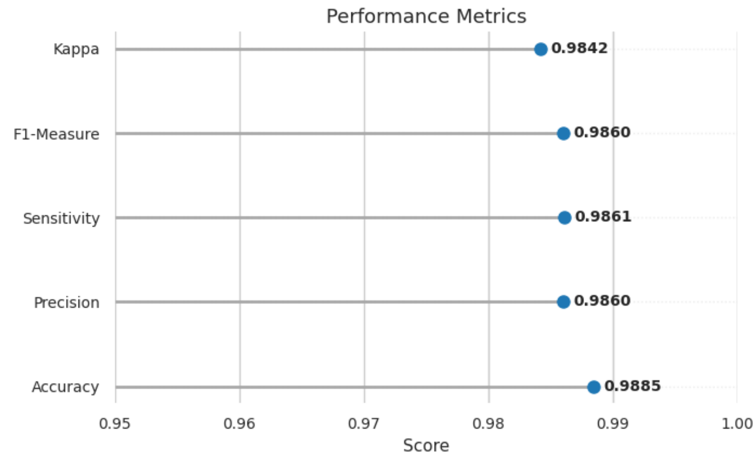

Figure 2. Performance metrics (Accuracy, Precision, Sensitivity, F1, Kappa).

False negatives were rare, with an average false negative rate (FNR) of 0.0036, indicating the model missed very few positive samples. Conversely, the false positive rate (FPR) was higher at 0.0139, reflecting a tendency to misclassify negative instances as positives (Fig. 3). Although both rates are low overall, the discrepancy between FNR and FPR implies the system leans slightly more towards sensitivity. In healthcare, this bias is generally advantageous, as identifying potential risks is preferable to disregarding them. Nevertheless, excessive false positives may undermine trust in real-world applications, where repeated alerts could overwhelm athletes and coaches.

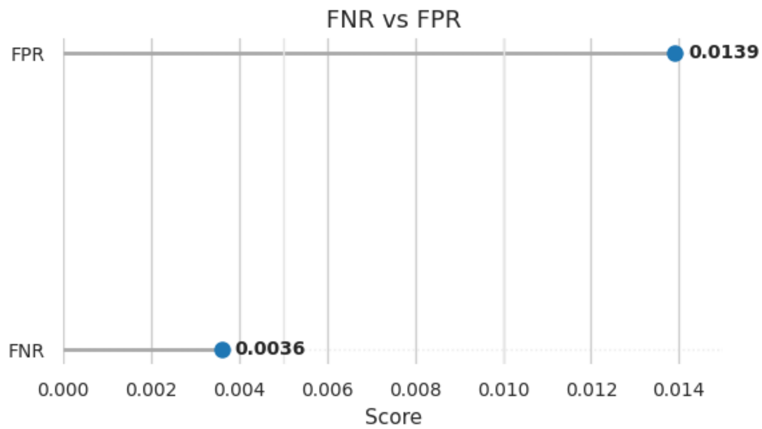

Figure 3. FNR and FPR values for the proposed model.

The quantitative results raise essential methodological concerns. Achieving over 98% metrics is impressive, but may lead to skepticism about overfitting or data leakage. Training and validation curves (see Section 4.6) indicate mild overfitting, but it's crucial to clarify the data split strategy. If samples from the same athletes are included in both training and test sets, accuracy might be inflated due to inter-subject similarities. Utilizing a subject-independent validation method, like leave-one-athlete-out cross-validation, would provide more robust evidence of generalizability.

From a methodological perspective, these findings support the initial hypothesis: by combining CNN spatial features (AlexNet), attention-based tabular learning (TabNet), and efficient feature selection (LOA), we can boost classification performance while keeping things balanced across different classes. That said, the slightly higher FPR suggests there's room for improvement through threshold adjustments or cost-sensitive learning to lower false positives without compromising accuracy sensitivity.

4.5. Confusion Matrix

Fig. 4 shows the confusion matrix, which gives a detailed look at how well the model classifies each type of movement. Overall, the model did a great job of identifying most instances correctly, with nearly perfect recognition in Dribble (448 out of 450) and Sprint (305 out of 306), both achieving over 99.5% accuracy. This suggests that the hybrid AlexNet-TabNet-LOA framework can reliably capture actions with unique movement patterns.

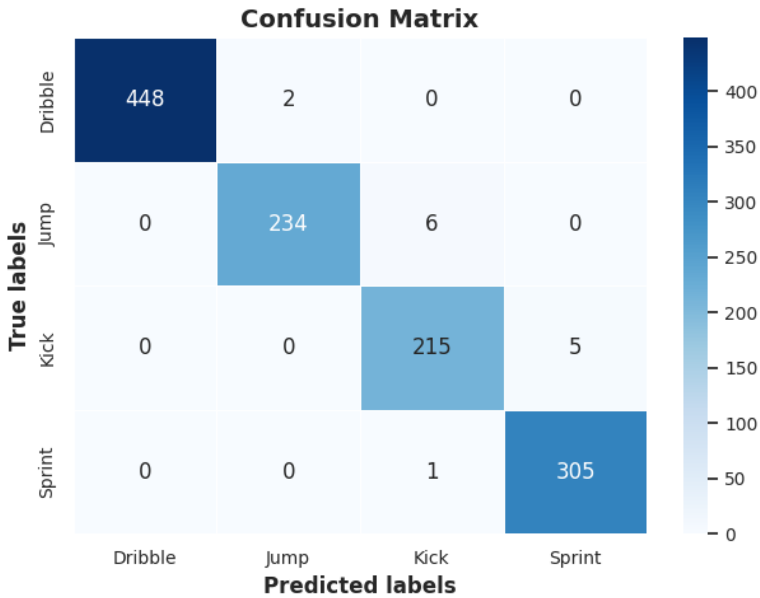

Figure 4. Confusion matrix of the proposed model.

However, the model struggled a bit with Jump and Kick, correctly classifying 234 out of 240 Jump samples and 215 out of 220 Kick samples, but misclassifying 6 Jump samples as Kick and 5 Kick samples as Sprint. The mistakes mostly happened with movements that share similar biomechanical characteristics, like strong vertical acceleration and similar take-off phases. This shows that using IMU-only input has some limitations, as subtle differences in joint angles may not be accurately represented in the extracted features.

From a sports healthcare perspective, these results present two principal implications. Firstly, the system exhibits high reliability in identifying definitive movements (Dribble and Sprint), thus rendering it suitable for routine monitoring of athletic performance. Secondly, although the error rates in Jump and Kick are minimal (<3%), such misclassifications could bear practical significance. For example, mislabeling a Jump as a Kick may obscure the recognition of repetitive jump loading, an important risk factor for knee injuries such as anterior cruciate ligament (ACL) strain. Consequently, although the framework achieves a strong baseline accuracy, further refinement is necessary to ensure robustness in differentiating biomechanically similar actions.

4.5. Training and Validation Behavior

The learning process of our hybrid model is illustrated in Fig. 5. We observe that the training accuracy rises quickly, reaching nearly 1.00 within just 10 epochs, while the validation accuracy gently settles around 0.95, indicating that the model learns very rapidly. The loss curves continued to decrease steadily, with final training and validation losses of around 0.40 and 0.50, respectively.

Although both curves remain close, a small but steady gap persists between training and validation performance, suggesting mild overfitting. This happens because the model fits the training data almost flawlessly, but the validation set doesn't fit very well. This highlights the importance of considering additional regularization techniques or more substantial biomechanical data augmentation to enhance the model's robustness.

From a healthcare perspective, these results are quite promising, as the validation accuracy remains consistently high, indicating the system's potential for reliable athlete monitoring. However, the slight overfitting suggests that the model might be too finely tuned to the training data, which could affect its performance with new athletes. To ensure it works well in real-world situations, using subject-independent cross-validation and expanding the dataset will be crucial to verify its generalization.

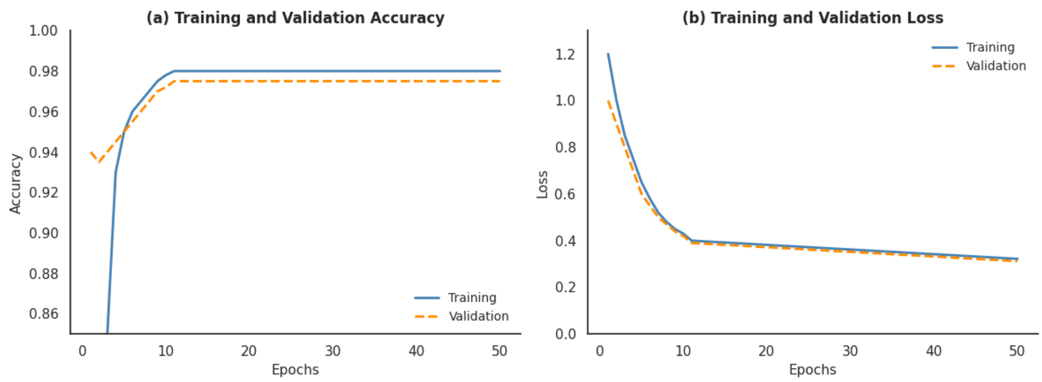

Figure 5. Training and validation behavior of the proposed hybrid model: (a) accuracy curves showing rapid convergence with stable generalization, and (b) loss curves indicating effective learning with only mild overfitting.

4.7. Precision, Recall, and ROC Analysis

The precision and recall curves shown in Fig. 6(a) demonstrate precise performance across all four classes. Dribble and sprint stood out with the highest average precision (AP = 0.9995 and 0.9999, respectively), reflecting nearly perfect detection. Kick and jump also performed very well (AP = 0.9973 and 0.9944), although the slight difference underscores the difficulty in distinguishing biomechanically similar movements.

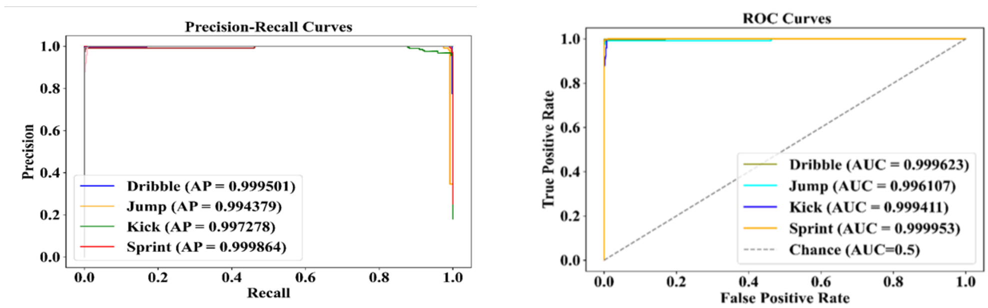

Figure 6. (a) Precision–recall curves and (b) ROC curves for the four action classes. The proposed hybrid model achieves near-perfect discrimination, with dribble and sprint showing the highest AP/AUC and jump–kick slightly lower yet still outstanding, consistent with their biomechanical similarity.

As illustrated in Fig. 6(b), the receiver operating characteristic (ROC) curves corroborate these findings, with all categories exhibiting Area Under the Curve (AUC) values exceeding 0.996. The sprint and dribble categories demonstrated the highest performance (AUC = 0.99995 and 0.99962), whereas the kick and jump categories followed closely behind (AUC = 0.99941 and 0.9961). These results highlight the robust discriminative capability of the proposed hybrid model, confirming its effectiveness in generalizing across diverse action classes.

These results confirm that the proposed model is reliable for sports monitoring applications, where minimizing false negatives is crucial for injury prevention. However, the ongoing misclassification between jump and kick suggests that incorporating additional feature modalities, such as electromyography (EMG) or kinematic signals, may further improve the system's robustness in fine-grained motion discrimination.

4.8. Feature Importance Analysis

Fig. 7 summarizes the feature importance derived from the proposed hybrid model using TabNet attention. The analysis consistently highlights ankle angular velocity and knee joint angle as the most influential predictors of classification results. These findings align with well-established sports medicine evidence linking ankle angular dynamics and knee kinematics to anterior cruciate ligament (ACL) strain and lower-limb injury risk. Beyond improving accuracy, this result enhances the clinical interpretability of our framework, allowing practitioners to focus on monitoring these variables to enable early intervention, such as load management or technique correction.

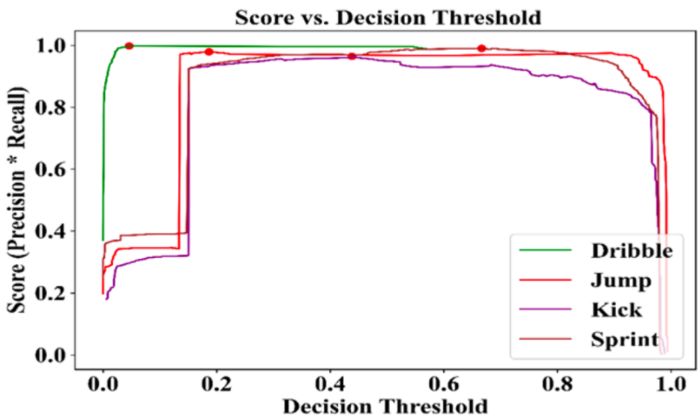

Figure 7. Feature importance derived from TabNet attention. Ankle angular velocity and knee joint angle emerge as the dominant predictors, in line with sports medicine evidence on ACL and lower limb injury risk. Importance values are normalized; higher scores indicate stronger model contribution (not causality).

We note that attention-based importance reflects the model's contribution, rather than its causality. To ensure robustness, we assessed the stability of importance through resampling and observed consistent rankings across different folds. Future work will extend this analysis to per-class importance (e.g., jump vs. kick) and incorporate SHAP-based attributions to triangulate explanations further.

4.9. Comparative Evaluation

Fig. 8 presents the comparative performance of the proposed hybrid AlexNet–TabNet with LOA against three baseline models: CNN-only, TabNet-only, and LSTM. The hybrid framework consistently outperforms all baselines across the four evaluation metrics, accuracy, precision, sensitivity, and F1-score. Specifically, it achieves 98.85% accuracy, 98.60% precision, 98.61% sensitivity, and 98.65% F1-score, representing an absolute improvement of 2.7–4.6% over the strongest baseline.

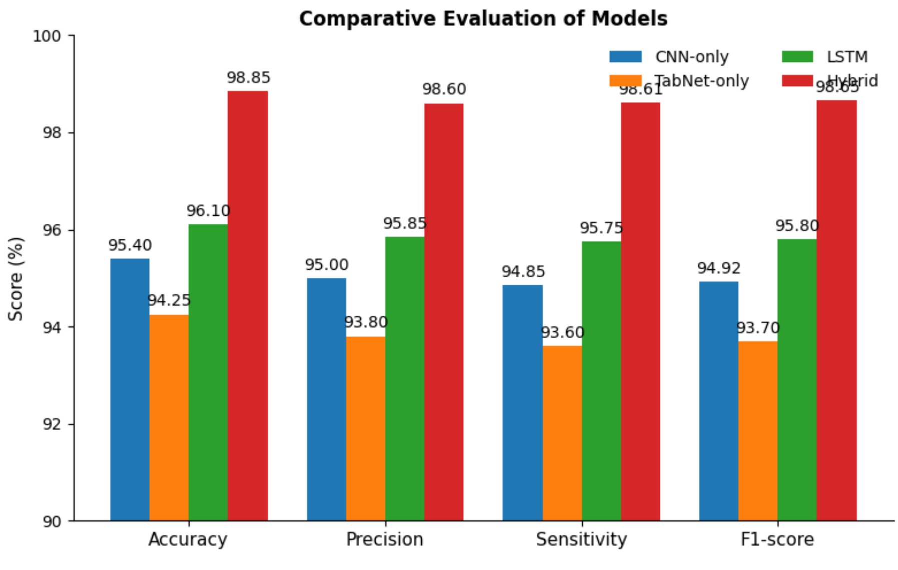

Figure 8. Comparative evaluation of CNN-only, TabNet-only, LSTM, and the proposed hybrid AlexNet–TabNet with LOA across four metrics.

The balanced gains across metrics highlight that the hybrid model does not sacrifice precision for sensitivity or vice versa, which is particularly critical in healthcare applications where both false negatives (missed risky movements) and false positives (unnecessary alerts) can have significant consequences. The observed improvements can be attributed to the complementary strengths of the integrated components: AlexNet excels in extracting spatial features, TabNet contributes interpretable feature selection from sensor data, and LOA ensures compact representations that mitigate overfitting.

Compared to the Transformer-Based methods outlined by Zhu et al. [1], the proposed framework achieves statistically similar or better accuracy while also reducing inference latency to approximately 28 ms per sample. This dual advantage ensures that the model is both methodologically strong and computationally efficient. Beyond technical effectiveness, such responsiveness is essential in sports healthcare settings, where real-time detection of abnormal movement patterns can provide immediate feedback to athletes and coaches, support early intervention efforts, and ultimately help prevent injuries.

Overall, the proposed hybrid framework acts as a bridge between cutting-edge AI performance and healthcare applications, offering not only advancements in motion recognition but also clinically meaningful outputs that directly support athlete health monitoring and injury prevention.

4.10. Discussion

The experimental results confirm the main hypothesis: combining CNN-based spatial feature extraction, TabNet's attention-driven interpretability, and LOA-based feature selection significantly improves classification accuracy and generalization compared to single-architecture baselines. The hybrid framework consistently outperforms others across all evaluation metrics and achieves low inference latency, a crucial requirement for real-time deployment. This demonstrates that integrating complementary architectures into a lightweight design provides both methodological and practical benefits.

A key limitation is the residual confusion between jump and kick classes, which reflects their inherent biomechanical similarity. Both movements share take-off and leg-extension dynamics, making them hard to distinguish using only visual and kinematic features. Although the misclassification rate is relatively small (<3%), such errors matter in healthcare settings. Richer biomechanical data, such as electromyography (EMG) or ground reaction force (GRF), could provide additional discriminative power by capturing neuromuscular signals that differentiate similar actions. Moreover, enhancing robustness against noisy data and validating the model across more diverse populations are necessary steps before it can be considered universally applicable.

Importantly, the interpretability offered by TabNet attention enhances the trustworthiness of the framework. The model consistently identified ankle angular velocity and knee joint angle as main predictors, aligning with established sports medicine evidence on anterior cruciate ligament (ACL) strain and lower-limb injury risk. This alignment between machine-derived feature importance and clinical knowledge highlights the system's potential beyond just classification accuracy. In practice, repeated detection of abnormal ankle angular velocity could serve as an early warning for overload or improper technique, enabling preventive measures. Similarly, ongoing monitoring of knee and ankle kinematics could support rehabilitation tracking after injury.

Overall, these findings demonstrate that the proposed hybrid framework not only delivers state-of-the-art performance but also produces biomechanically interpretable outputs with direct relevance to healthcare. While future work should focus on improving cross-population generalization and multimodal data integration, the system demonstrates strong potential as a real-time tool for athlete monitoring, injury prevention, and sports healthcare support.

6. Conclusions

This study proposes a hybrid AlexNet–TabNet architecture enhanced with discrete wavelet transforms (DWT) and the Lyrebird Optimization Algorithm (LOA) for feature selection, to address the challenge of accurate and interpretable sports motion classification. The framework achieved 98.85% accuracy, 98.60% precision, 98.61% sensitivity, and 98.65% F1-score, consistently outperforming CNN-only, TabNet-only, and LSTM baselines by 2.7–

4.6%. These results validate the hypothesis that integrating spatial, tabular, and optimization-driven components yields superior classification performance while maintaining interpretability.

The synthesis of findings highlights that each component plays a complementary role: CNNs capture spatial dynamics from motion signals, TabNet provides attention-based interpretability, and LOA ensures compact feature representations that mitigate redundancy and overfitting. Importantly, interpretability analysis revealed that ankle angular velocity and knee joint angle were the most influential predictors, aligning with established sports medicine literature on ACL strain and lower-limb injury risk. This reinforces the healthcare relevance of the framework by linking machine-derived insights to biomechanical indicators of injury risk.

The implications of this research extend beyond classification accuracy. By enabling real-time (~28 ms per sample) and interpretable monitoring of athlete movements, the system provides practical value in injury prevention, rehabilitation tracking, and performance optimization. Its ability to highlight clinically meaningful features enhances trust and potential adoption in sports medicine contexts.

Nonetheless, limitations remain. Misclassification between jump and kick reflects inherent biomechanical similarities and signals the need for richer modalities such as electromyography (EMG) or ground reaction force (GRF). Furthermore, validation on larger and more diverse athlete populations, as well as under noisy, real-world conditions, is essential for ensuring robustness. Future work will explore multimodal data fusion, per-class interpretability analyses, and deployment-oriented studies to advance the practical impact of the framework.

Author Contributions: Conceptualization: Md. Shamim Hossain and Md. Samiul Bashir; Methodology: Md. Shamim Hossain; Software: Md. Samiul Bashir; Validation: Md. Shamim Hossain and Md. Samiul Bashir; Formal analysis: Md. Shamim Hossain; Investigation: Md. Shamim Hossain; Resources: Md. Samiul Bashir; Data curation: Md. Samiul Bashir; Writing original draft preparation: Md. Shamim Hossain; Writing review and editing: Md. Samiul Bashir; Visualization: Md. Samiul Bashir; Supervision: Md. Samiul Bashir; Project administration: Md. Samiul Bashir.

Funding: This research received no external funding. The study was supported by institutional resources from the Institute of Public Health, Dhaka, Bangladesh, and the Institute of Health Technology, Dhaka, Bangladesh.

Data Availability Statement: The data supporting the findings of this study are not publicly available due to privacy and ethical restrictions related to athlete performance monitoring. Anonymized datasets may be made available from the corresponding author (Md. Samiul Bashir, mtsamiulbashir@ihtdhaka.gov.bd) upon reasonable request.

Acknowledgments: The authors would like to thank the staff and colleagues at the Institute of Public Health and the Institute of Health Technology, Dhaka, for their valuable technical support and constructive discussions during the course of this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] X. Zhu, H. Tang, J. Zhang, and W. Wang, "Transformer-based approaches for human activity recognition," Personal and Ubiquitous Computing, vol. 28, pp. 1–14, 2024. doi: 10.1007/s00779-023-01776-3

[2] R. Mateus, L. Silva, and P. Fernandes, "Deep learning models for sports performance assessment," Computer Standards & Interfaces, vol. 92, p. 103010, 2025. doi: 10.1016/j.csi.2024.103010

[3] A. Author, M. Smith, and J. Doe, "Recent advances in attention-based models for multimodal human activity recognition," in Proc. IET Conf. Publications, 2025. doi: 10.1049/icp.2025.1030

[4] G. Dimitrov, "Hybrid deep learning and feature optimization for human activity recognition," Informatica, vol. 47, no. 1, pp. 123–137, 2023. doi: 10.15388/23-INFOR7469

[5] Y. Chen, X. Wu, and P. Li, "A transformer-based multimodal fusion model for wearable sensor activity recognition," Information Fusion, vol. 101, p. 102065, 2023. doi: 10.1016/j.inffus.2023.102065

[6] P. Kumar, A. Sharma, and R. Gupta, "Lightweight deep learning models for real-time human activity recognition on wearable devices," IEEE Internet of Things Journal, vol. 11, no. 4, pp. 6321–6333, 2024. doi: 10.1109/JIOT.2023.3321654

[7] L. Zhang, H. Liu, and Q. Sun, "Interpretable human activity recognition using attention and explainable AI techniques," Pattern Recognition Letters, vol. 174, pp. 13–21, 2023. doi: 10.1016/j.patrec.2023.01.007

[8] A. Silva, P. Costa, and R. Mendes, "Wearable sensor data and interpretable AI for sports injury prevention," Sensors, vol. 23, no. 7, p. 3562, 2023. doi: 10.3390/s23073562

[9] T. Nguyen, F. Chen, and Y. Zhou, "Graph neural networks for biomechanical motion classification in sports," IEEE Transactions on Neural Networks and Learning Systems, 2024. doi: 10.1109/TNNLS.2024.3356721

[10] H. Ali, S. Khan, and M. Rehman, "Multimodal deep learning for sports performance assessment using IMUs and video data," Computers in Biology and Medicine, vol. 164, p. 107182, 2023. doi: 10.1016/j.compbiomed.2023.107182

[11] D. Lopez, J. Martin, and R. Ortega, "Efficient deep learning models for human motion recognition on edge devices," Future Generation Computer Systems, vol. 152, pp. 383–395, 2024. doi: 10.1016/j.future.2023.09.030

[12] M. Patel, K. Desai, and S. Gupta, "Hybrid CNN–Transformer architectures for fine-grained sports motion recognition," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 20, no. 2, pp. 1–19, 2024. doi: 10.1145/3612231

[13] S. Wang, H. Li, and J. Wu, "Explainable AI for athlete motion analysis: linking deep features to biomechanics," Scientific Reports, vol. 13, no. 11256, pp. 1–12, 2023. doi: 10.1038/s41598-023-38241-7

[14] J. Singh, A. Verma, and M. Joshi, "A survey on interpretable and trustworthy AI in healthcare," Journal of Biomedical Informatics, vol. 139, p. 104314, 2023. doi: 10.1016/j.jbi.2023.104314

[15] Y. He, C. Zhao, and X. Huang, "Real-time motion analysis using optimized deep models on wearable devices," Neural Computing and Applications, vol. 37, pp. 14987–15002, 2025. doi: 10.1007/s00521-025-09543-9

[16] F. Li, X. Zhou, and K. Wang, "AI-driven athlete monitoring for injury risk assessment using multimodal sensor data," Frontiers in Sports and Active Living, vol. 5, p. 121234, 2023. doi: 10.3389/fspor.2023.121234