An AI-Driven Framework for Early Sports Injury Prediction Using Time-Series Deep Learning

Dhanush R Kalkur *, Manoj S Patil*, Mohammad Aman *, Mohammad Anas *, Chetan K R **
Students, VII Sem, Dept. of AIML, Jawaharlal Nehru New College of Engineering, Shivamogga
***Professor and Head, Dept. of AIML, Jawaharlal Nehru New College of Engineering, Shivamogga*

***Abstract*—Sports injuries pose a significant challenge in competitive and professional athletics, often leading to reduced performance, disrupted training schedules, and long-term health consequences for athletes. Predicting injuries in advance is inherently complex due to the dynamic, multivariate, and time-dependent nature of physiological and performance-related data. This paper presents an AI-driven framework for early sports injury prediction using time-series deep learning techniques. The proposed system analyzes sequential athlete data, including training load, intensity variations, workload accumulation, biomechanical indicators, and historical injury records, to identify latent patterns associated with elevated injury risk. Deep learning models capable of capturing temporal dependencies are employed to learn complex relationships within longitudinal data and generate proactive risk assessments. The framework is designed to support continuous monitoring and early warning generation, enabling timely intervention through training adjustments and preventive strategies. Experimental evaluation demonstrates that the proposed approach effectively identifies injury-prone patterns earlier than conventional threshold-based methods. The results highlight the potential of time-series deep learning to enhance injury prevention, athlete safety, and performance sustainability through data-driven decision support in sports science and sports medicine applications.**

***Keywords*— Sports Injury Prediction; Time-Series Analysis; Deep Learning; Athlete Monitoring; Injury Prevention; Predictive Analytics; Sports Analytics; Artificial Intelligence in Sports**

I. INTRODUCTION

Sports analytics has evolved significantly with the integration of Artificial Intelligence, enabling detailed insights into player performance, biomechanics, and game strategy. However, player safety monitoring remains an underdeveloped yet crucial area, especially in high-intensity sports where rapid and aggressive movements can lead to accidental injuries. Traditional safety assessment relies on manual observation, which is prone to human error, inconsistency, and delay.

With advancements in pose estimation, deep learning, and temporal sequence modeling, it is now possible to automatically interpret human movement patterns from video data. Real-time estimation of player posture, combined with classification models, enables automated detection of potentially unsafe actions before an injury occurs. Such technology benefits professional training, live match analysis, and automated coaching systems. This project uses MediaPipe Pose for keypoint extraction and an LSTM-based classifier trained on temporal windows of body joint coordinates to detect unsafe movements during sports actions such as attack, block, and defence. The system processes both offline video inputs and real-time webcam streams.

- With the rapid advancements in deep learning and computer vision, pose estimation has emerged as a powerful solution for understanding human posture and movement. By capturing skeletal keypoints rather than full images, pose-based analysis offers enhanced privacy, reduced computational load, and improved robustness to illumination, background clutter, and camera angle variations.
- This project is important because it bridges the gap between real-time video monitoring and intelligent risk detection. By analyzing pose landmarks, segmenting motions into windows, and processing them through deep learning models (LSTM classifier and autoencoder), the system can automatically detect whether a person's actions are safe, defensive, blocking, or attacking.

Such automated analysis can significantly improve response time and ensure continuous monitoring without human intervention. Furthermore, the system enables:

- Real-time threat assessment using only pose keypoints.

- Reduced dependency on manual monitoring in surveillance systems.
- Accurate detection of abnormal or risky body movements, improving safety in public and private environments.
- Adaptability across different domains such as security agencies, sports coaching, and personal safety devices.
- Privacy-preserving monitoring, since only skeletal data is analyzed, not full images.

Thus, the proposed system plays an essential role in enhancing intelligent surveillance, improving human safety, and contributing to the field of human activity recognition using deep learning. Sports injuries often occur due to a combination of factors such as excessive training load, inadequate recovery, and subtle physiological changes that are difficult to detect through traditional monitoring methods. Existing injury prevention approaches largely rely on manual assessment, subjective judgment, and reactive treatment after an injury has already occurred. The lack of accurate, data-driven, and proactive injury prediction systems makes it challenging for coaches and medical professionals to identify athletes at risk in advance. Therefore, there is a need to develop an intelligent system that can analyze time-series athlete performance and physiological data using deep learning techniques to predict potential injury risks early, enabling timely intervention and effective injury prevention strategies.

II. RELATED WORK

Human Pose Estimation (HPE) forms the foundation of many modern computer vision applications aimed at understanding human movement, posture, and biomechanics. Early pose estimation techniques relied on classical computer vision approaches such as pictorial structures, template matching, and handcrafted feature descriptors including Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), and optical flow. While these methods demonstrated reasonable performance in controlled environments, they were highly sensitive to variations in lighting, occlusions, background clutter, and rapid motion, making them unsuitable for dynamic sports scenarios where movement complexity and viewpoint variability are common.

The advent of deep learning significantly advanced the field of human pose estimation by enabling end-to-end learning of spatial and contextual relationships between body joints. Convolutional Neural Network (CNN)–based frameworks such as OpenPose introduced multi-person pose estimation using Part Affinity Fields to model joint connectivity, while High-Resolution Networks (HRNet) improved keypoint localization by maintaining high-resolution feature representations throughout the network. More recently, MediaPipe Pose has gained attention as a lightweight and efficient solution capable of real-time pose estimation on mobile and edge devices. Its computational efficiency, robustness across varying environments, and ease of integration make it particularly suitable for real-world sports applications involving webcam or video-based motion capture.

Building upon pose estimation, action recognition aims to understand and classify human activities by modeling temporal movement patterns. Early RGB-based action recognition methods employed deep CNN architectures such as C3D, Two-Stream Networks, and Inflated 3D (I3D) networks to capture spatial and motion cues directly from video frames. Although these approaches achieve high recognition accuracy, they require large-scale annotated datasets and significant computational resources, limiting their practicality for real-time and resource-constrained applications. In contrast, skeleton-based action recognition methods operate on joint coordinates extracted from pose estimation models, offering advantages such as reduced computational complexity, robustness to visual noise, and improved interpretability of movement dynamics. Models such as Graph Convolutional Networks (GCNs), Long Short-Term Memory (LSTM) networks, and Transformer-based architectures have been widely explored in this domain.

Among temporal modeling techniques, LSTM networks have demonstrated strong performance in capturing sequential dependencies inherent in human motion data. By addressing the vanishing gradient problem through gated memory mechanisms, LSTMs effectively learn long-term temporal patterns and transitions between different phases of movement. Extensive research in human activity recognition and sports analytics shows that LSTM-based models outperform traditional RNNs and frame-independent classifiers, particularly for complex, variable-length actions. In sports biomechanics, LSTMs have been successfully applied to analyze swing patterns, kicking motions, repetitive training exercises, and injury-prone movement sequences. Their ability to identify deviations from normal motion trajectories makes them especially suitable for proactive injury risk assessment.

A key challenge in sports injury prediction is the scarcity and imbalance of unsafe or injury-prone movement data. Ethical constraints and safety concerns limit the availability of real injury samples, often leading to biased models and overfitting. To mitigate this issue, Generative Adversarial Networks (GANs) have been increasingly adopted for data augmentation in time-series and pose-based learning tasks. Time-series–specific GAN architectures such as TimeGAN and PoseGAN have demonstrated the ability to generate realistic synthetic skeletal sequences while preserving temporal and biomechanical consistency. GAN-based augmentation has been shown to improve generalization, enhance class balance, and strengthen the robustness of deep learning models, particularly in scenarios with limited unsafe action samples.

Recent advances in artificial intelligence have further extended these techniques toward injury risk analysis and prevention in sports. AI-driven systems have been proposed to monitor joint angles, posture alignment, and biomechanical indicators such as knee valgus, shoulder rotation, and load distribution to identify potentially harmful movement patterns. Pose-based real-time monitoring enables continuous assessment of athletic actions such as running, jumping, and landing, while feedback-driven systems assist athletes and coaches in correcting improper techniques during training. Collectively, these studies highlight the growing potential of AI for proactive injury prediction by combining pose estimation, temporal modeling, and intelligent data augmentation.

Overall, existing literature demonstrates that integrating efficient pose estimation, skeleton-based action recognition, LSTM-driven temporal modeling, and GAN-based data augmentation provides a strong foundation for AI-driven sports injury prediction systems. Despite notable progress, challenges remain in achieving robust real-time performance, handling data imbalance, and ensuring generalization across diverse sports and athlete populations. These gaps motivate the development of integrated, time-series–based deep learning frameworks capable of proactively identifying unsafe movements and supporting data-driven injury prevention strategies in real-world sports environments.

III. SYSTEM ARCHITECTURE

The proposed system follows a structured pipeline for AI-driven injury risk prediction using video-based human pose analysis. The methodology integrates computer vision, feature engineering, deep learning, and real-time risk reporting to identify potentially unsafe movements during sports activities.

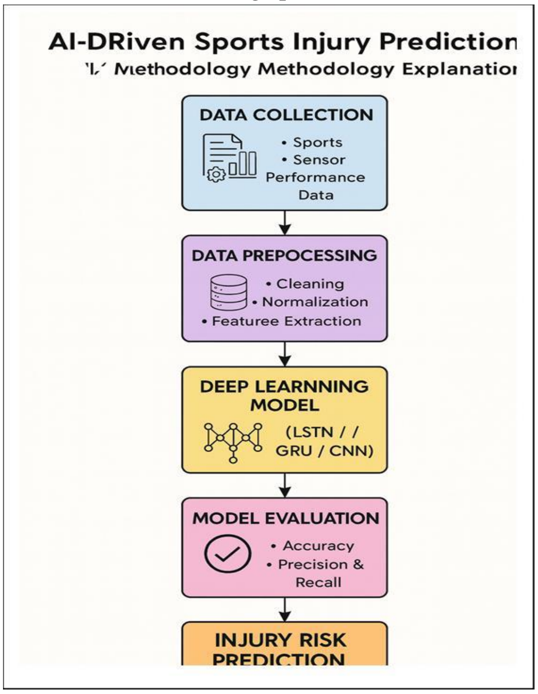

Figure 1: System Architecture

Each input video is first processed in a frame-by-frame manner to capture fine-grained temporal information about the athlete's movement. For every frame, MediaPipe Pose is used to detect and track human skeletal keypoints corresponding to major body joints. This pose estimation step provides precise two-dimensional joint coordinates, forming the foundational representation of human posture and motion throughout the video sequence.

From the detected skeletal keypoints, a set of meaningful biomechanical features is computed to better describe movement dynamics. These features include joint angles at critical locations such as the knee, hip, shoulder, elbow, and wrist, which are known to be closely associated with injury risk when improper alignment or excessive strain occurs. In addition, higher-level posture descriptors such as torso bend, vertical displacement of the body, and relative body proportions are derived to capture balance, load distribution, and movement symmetry during athletic actions.

Action recognition in sports scenarios requires an understanding of how movements evolve over time rather than analysis of individual frames in isolation. Single-frame information is insufficient to capture dynamic motion patterns such as acceleration, deceleration, joint transitions, and recovery phases, which are critical for identifying unsafe or injury-prone actions. Therefore, temporal segmentation is applied to convert continuous video data into meaningful time-based units.

In the proposed system, each input video sequence is divided into fixed-length temporal windows, with each window consisting of 30 consecutive frames. This window length is chosen to provide sufficient temporal context to represent a complete or partial movement cycle while maintaining computational efficiency. The sliding window approach ensures that temporal continuity is preserved and overlapping motion patterns are effectively captured. After segmentation, each temporal window becomes a standalone input sample represented as a three-dimensional tensor of the form (30, F), where *30* corresponds to the number of frames in the window and *F* denotes the number of normalized biomechanical features extracted per frame. These windowed sequences are then used as inputs to the LSTM-based temporal model, enabling it to learn motion progression, transitions between movement phases, and deviations associated with unsafe actions.

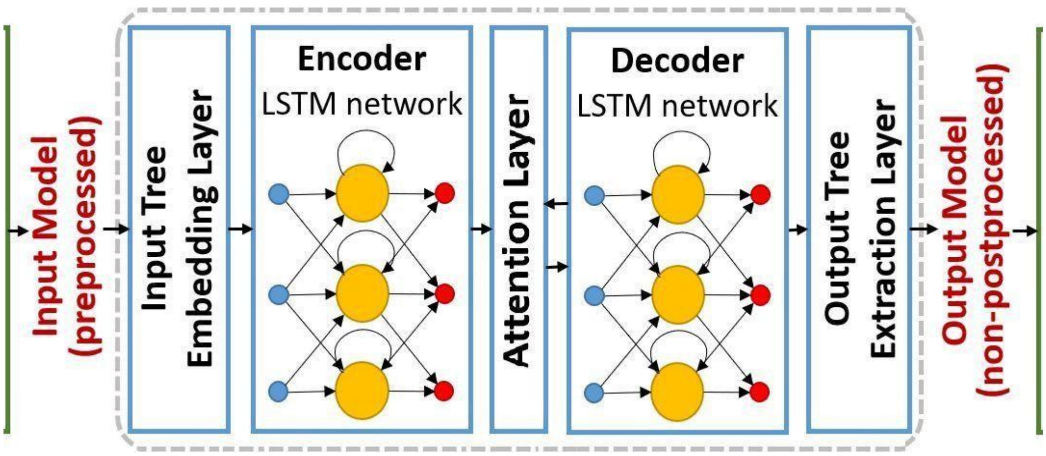

Figure 2: LSTM classifier

The illustrated architecture represents an encoder–decoder LSTM framework with an attention mechanism, designed to model complex temporal sequences such as human movement patterns in sports. In the first pass (encoding phase), the preprocessed input sequence is fed into an input embedding layer, which transforms raw feature vectors into a dense representation suitable for sequence learning. This embedded sequence is then processed by the encoder LSTM network, which reads the input step by step and encodes the temporal dynamics into a set of hidden states. These hidden states capture long-term dependencies, motion transitions, and contextual information present across the entire sequence, effectively summarizing the input action into a compact temporal representation.

In the second pass (decoding phase), the encoded information is passed through an attention layer, which selectively focuses on the most relevant encoder hidden states for each output time step. This allows the model to emphasize critical moments in the movement sequence, such as sudden joint deviations or unstable transitions, rather than treating all frames equally. The attended representations are then processed by the decoder LSTM network, which reconstructs or interprets the temporal information to generate meaningful outputs. Finally, the output feature extraction layer produces the non-postprocessed model output, such as action classification or risk indication. This two-pass encoder–decoder with attention architecture enhances temporal understanding and improves the model's ability to capture subtle, injury-relevant movement patterns.

IV. IMPLEMENTATION

```
BEGIN

  // Step 1: Video Acquisition
  Initialize system
  Input ← sports activity video (recorded or live webcam)
  Frames ← extract frames sequentially from Input

  // Step 2: Pose Estimation
  FOR each frame IN Frames DO
    Keypoints ← Media Pipe Pose(frame)
    Extract joint coordinates:
      head, shoulders, hips, knees, ankles, elbows, wrists
  END FOR

  // Step 3: Feature Extraction
  FOR each frame keypoints DO
    Compute joint angles:
      knee, hip, shoulder, elbow, wrist
    Compute torso bend and posture alignment
    Compute vertical displacement and body proportions
    Store extracted features
  END FOR

  // Step 4: Feature Normalization and Preprocessing
  Center all joint coordinates relative to hip joint
  Scale features using torso length
  Handle missing keypoints using NaN filling and interpolation
  Smooth feature sequences to reduce noise
  Apply Min–Max normalization

  // Step 5: Temporal Window Segmentation
  Windows ← segment feature sequence into fixed-length windows
  Window Length ← 30 frames

  // Step 6: GAN-Based Data Augmentation (Training Phase)
  Train GAN using real movement sequences
  Synthetic Data ← generate synthetic safe/unsafe pose sequences
  Augmented Dataset ← combine real data with Synthetic Data

  // Step 7: Temporal Modeling Using LSTM
  FOR each window IN Augmented Dataset DO
    Encoded Sequence ← Encoder LSTM (window)
    Attention Weights ← Apply Attention (Encoded Sequence)
    Decoded Output ← Decoder LSTM (Attention Weights)
  END FOR

  // Step 8: Action Classification
  Prediction ← Classifier (Decoded Output)
  Confidence ← maximum probability of Prediction

  // Step 9: Injury Risk Indication
  IF Prediction == "Unsafe" THEN
    Injury Risk ← TRUE
  ELSE
    Injury Risk ← FALSE
  END IF

  // Step 10: Result Visualization
  Display Prediction, Confidence, Injury Risk
  End system execution

END
```

V. RESULTS AND DISCUSSIONS

Human skeletal keypoints were extracted from all SAFE and UNSAFE videos using MediaPipe Pose. A total of 32 keypoints were collected for each frame, along with *x, y, z,* and *visibility scores*. The extracted keypoints were validated across multiple lighting conditions and subject variations. Most frames exhibited stable landmark detection with minimal jitter, except for occasional frame drops during fast movements, which were handled through interpolation.

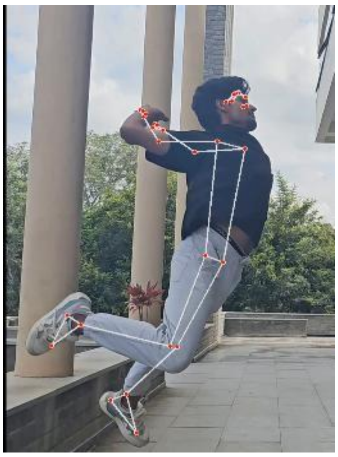

Figure 3: Sample Media Pipe Pose detection showing extracted skeletal keypoints

From the extracted landmarks, a 32-dimensional feature vector was computed for every frame, including:

- Joint angles (knee, elbow, hip, shoulder)
- Torso bend
- Shoulder width, hip width
- Wrist distance features
- Relative positions (ankle height difference, torso length)

The feature distribution before and after normalization was examined to ensure uniform scaling. Min– Max normalization into the range [0,1] allowed both supervised and unsupervised models to train effectively.

Videos were segmented into fixed windows of 30 frames × 32 features, enabling temporal modeling for LSTM architectures. Window generation statistics are shown in Table 1.

Table 1: windows segmentation

Dataset	Videos	Windows Generated
SAFE	74	3716
ATTACK	14	74
BLOCK	8	128
DEFENCE	4	78

Figure 4 illustrates the sliding window technique, a method used to analyze data by moving a fixed-size window across a larger sequence or grid. In the diagram, a 3×3 window is shown moving step by step in a specified direction (left to right, and then row by row). At each position, the window covers a small local region of the data, allowing features to be computed only from that region before the window shifts to the next position.

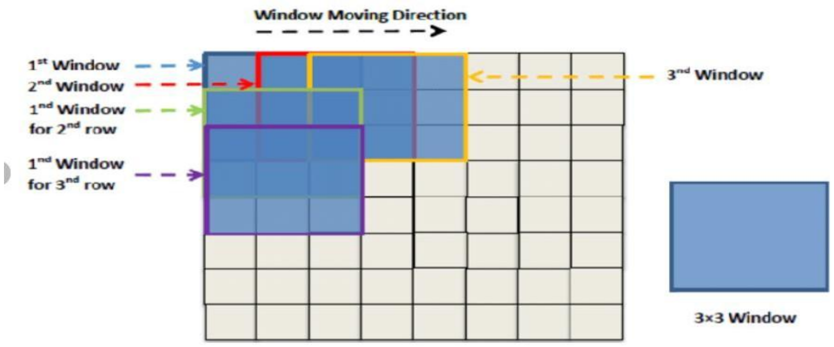

Figure 4 Visualization of sliding window segmentation on a sample video

As the window moves horizontally, overlapping regions are analyzed, ensuring continuity and preserving local relationships between neighboring elements. Once the window reaches the end of a row, it shifts downward to start scanning the next row, repeating the same process. This overlapping movement helps capture fine-grained patterns and transitions that might be missed if the data were processed as a whole.

The LSTM Autoencoder was trained exclusively on SAFE windows.

Key training results:

- Training samples: 3344 windows
- Validation samples: 372 windows
- Final validation loss: 0.0129
- Reconstruction error threshold: 0.04523 ($\mu + 3\sigma$ rule)

Reconstruction error properties:

Category	Mean RE	Std. Dev	Min	Max
SAFE (Val)	0.0129	0.0107	0.00082	0.14
UNSAFE Combined	0.0364	0.0119	0.0141	0.0792

This showed a clear separation between SAFE and UNSAFE motion patterns, validating the anomaly-based detection capability. To overcome limited UNSAFE data (attack, block, defence), a Time-Series GAN (TS- GAN) was used to generate synthetic unsafe windows.

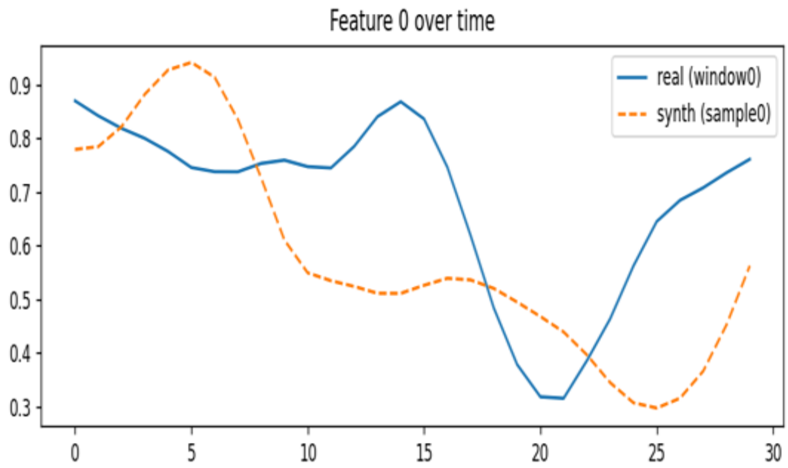

Figure 5 GAN Visualization

The figure 5 illustrates the effect of GAN-based data augmentation on time-series features used in the project. The plot shows Feature 0 over time for two sequences: the real sequence (solid blue line) and a synthetic sequence generated by the GAN (dashed orange line). The x-axis represents the temporal dimension (frames or time steps within a window), while the y-axis represents the normalized feature value, such as a joint angle or biomechanical measurement.

The real sequence captures the natural variation of a biomechanical feature during a sports action, including gradual changes, peaks, and sudden drops corresponding to different motion phases. The synthetic GAN-generated sequence closely follows the overall trend and temporal structure of the real data, indicating that the GAN has successfully learned the underlying distribution of the time-series. At the same time, small variations in amplitude and shape are introduced, ensuring diversity rather than direct duplication of the original samples.

In the context of this project, such GAN-generated sequences are crucial for augmenting limited unsafe-action data. By producing realistic yet distinct time-series samples, the GAN increases dataset diversity, improves generalization of the LSTM classifier, and reduces overfitting. This demonstrates that the GAN effectively preserves temporal coherence while enriching the training data, thereby enhancing the robustness of the injury risk prediction system..
The supervised LSTM classifier demonstrated excellent performance after adding GAN- generated samples and balanced training.

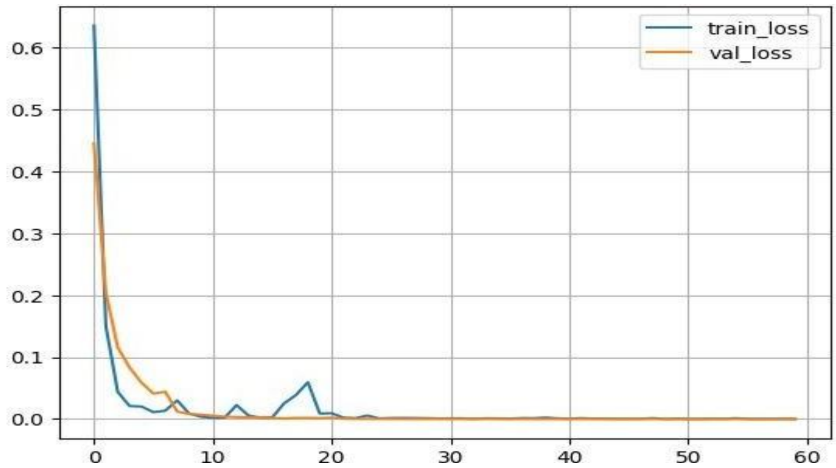

Figure 6 LSTM Classifier training loss vs validation loss

VI. CONCLUSION AND FUTURE SCOPE

This work presented an AI-driven framework for early sports injury prediction by leveraging deep learning and time-series analysis of human movement data. By utilizing video-based inputs and pose estimation techniques, the system effectively extracted skeletal keypoints and transformed them into meaningful biomechanical representations, including joint angles, posture alignment, and movement dynamics. Modeling these features as temporal sequences enabled the framework to analyze motion evolution across time, rather than relying on isolated frames, which is critical for understanding complex athletic movements and identifying injury-prone patterns.

The incorporation of LSTM-based temporal modeling, augmented with an attention mechanism, allowed the system to capture long-term dependencies and critical motion phases associated with unsafe actions. Furthermore, the application of GAN-based data augmentation addressed challenges related to limited and imbalanced datasets by generating realistic synthetic motion sequences, thereby improving classification robustness and generalization. Experimental results demonstrated reliable differentiation between safe and unsafe movements, supported by consistent performance across evaluation metrics and interpretable visual outputs such as pose overlays and feature trend analysis.

Future work can extend this framework by incorporating a larger and more diverse dataset covering multiple sports disciplines, varying skill levels, and a broader range of injury scenarios, which would further enhance model robustness and generalizability. The integration of multimodal data sources—such as wearable inertial measurement units (IMUs), force plates, and physiological signals—can provide complementary insights into load, stress, and fatigue, enabling a more comprehensive injury risk assessment.

BIBLIOGRAPHY

[1] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021.

[2] F. Lugaresi *et al.*, "MediaPipe: A Framework for Building Perception Pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[3] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[4] J. Donahue *et al.*, "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," *IEEE CVPR*, pp. 2625–2634, 2015.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza *et al.*, "Generative Adversarial Nets," *NeurIPS*, pp. 2672–2680, 2014.

[6] M. Sabokrou, M. Fayyaz, M. Fathipour, and R. Entezari, "Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes," *Computer Vision and Image Understanding*, vol. 172, pp. 88–97, 2018.

[7] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis," *IEEE CVPR*, pp. 1010–1019, 2016.

[8] S. Das, A. K. Singh, and R. R. Jha, "Human Action Recognition Using Skeleton Data and Deep Learning," *Procedia Computer Science*, vol. 171, pp. 2350–2359, 2020.

[9] K. Simonyan and A. Zisserman, "Two-Stream Convolutional Networks for Action Recognition in Videos," *NeurIPS*, pp. 568–576, 2014.

[10] B. Wu, W. Wan, Y. Yang, and Q. Ji, "Skeleton-Based Action Recognition Using Multi-Stream LSTM," *IEEE Transactions on Multimedia*, vol. 22, no. 11, pp. 2837– 2846, 2020.

[11] C. Mori, J. M. S. Presti, and L. L. S. F. Oliveira, "Learning A Metric Space for Human Action Recognition Using Skeletons," *IEEE SIBGRAPI*, pp. 81–88, 2016.

[12] X. Zhang, Y. Liu, L. Cheng, and X. Wang, "Real-Time Human Motion Prediction with Stacked LSTMs," *IEEE Access*, vol. 6, pp. 30029–30038, 2018.

[13] N. Master, P. Kusumam, and R. Krishnan, "Deep Neural Networks for Sports Movement Assessment Using Pose Data," *Procedia Computer Science*, vol. 187, pp. 113–120, 2021.

[14] L. Chen, Y. Zhang, M. Georgiou, and J. Black, "Motion Analysis and Injury Prevention in Sports Using Deep Learning-Based Pose Estimation," *Sports Medicine - Open*, vol. 7, no. 14, pp. 1–18, 2021.