International Journal of Artificial Intelligence
p-ISSN: 2407-7275, e-ISSN: 2686-3251

Original Research Paper

A Standard Deep Learning-Based Model That Integrates R3D-18 With RNN for Clip Classification and Transition Point Detection in Basketball Video Footages

Aliga Paul[1*], Joshua Ojo Nehinbe[2], Kingsley Eghonghon Ukhurebor[2]

[1] *Department of Computer Science, Faculty of Science, Ambrose Ali University. Ekpoma, Nigeria.*

[2] *Department of Computer Science, Faculty of Science, Edo State University. Nigeria.*

Article History
Received:
19.07.2025

Revised:
05.01.2026

Accepted:
28.01.2026

***Corresponding Author:**
Aliga Paul
Email:
aliga.paul@aauekpoma.edu.ng

Abstract: Research on basketball video clip analysis is rapidly emerging as a vital tool for coaching, player development, scouting, recruitment of coaches and players. Such analysis enhances post-game review, tactical planning, practice design, in-game adjustments, visual feedback and decision-making. However, many existing models lack robust architectures for transition detection, limiting their effectiveness. Additionally, high costs, limited customization, internet dependency and complex interfaces restrict widespread adoption of most models across professional and grassroots levels. To address these challenges, this paper implements a deep learning model with Python language. The model integrates Residual 3D Network with 18 layers (R3D-18) to capture spatial-temporal features, Convolutional neural networks (CNNs) to perform feature extraction and Recurrent Neural Network (RNN) to accomplish temporal modeling of offline basketball videos. Using entropy thresholding and top-k (i.e., top-10) accuracy metrics on four basketball datasets, the model performs clip-wise and video-level action classification with exceptional confidence, achieving 99.92% and 99.52% certainty that all the evaluative footages depict basketball videos. The model further detected no transitions in *basketball_video1* and *basketball_video3* while notable transitions appeared at clips 8 and 7 in *basketball_video2* and *basketball_video4*, respectively corresponding to spikes in uncertainty likely caused by activity changes like switching from basketball to running. Average entropy scores of 4.787, 3.5341, 3.7912 and 3.1976 across datasets were reported, reflecting potentially elusive variations and uncertainties within clip-level predictions due to nuanced activity shifts despite strong overall classification confidence.

Keywords: Basketball Video Clips Analysis, Convolutional Neural Networks (CNNs), Deep Learning, Entropy Thresholding, Recurrent Neural Network (RNN).

1. Introduction

Research on basketball video clip analysis is rapidly emerging as a potential supporting tool for coaching, player development, scouting, recruitment of coaches and players [1]. Studies have shown that analysis of past basketball videos can enhance post-game review, tactical planning, practice design, in-game adjustments, visual feedback and decision-making [2]. Nonetheless, many existing models lack robust architectures for transition detection, limiting their effectiveness. Additionally, high costs, limited customization, internet dependency and complex interfaces restrict extensive adoption of most models across professional and grassroots levels. To address these challenges, there are increasing demands to pragmatically identify insightful human actions that can meaningfully influence decision-making and this issue has spurred a surge in research focused on post-review analysis of basketball video footage over the years [3], [4]. This trend marks a shift from traditional machine learning approaches toward deep learning models, which are better suited for processing basketball video clips due to their ability to capture both spatial and temporal dimensions [5]. Basketball video clips are short segments extracted from full game footage that highlight specific actions, events or sequences that can elicit curiosity. These clips can capture scoring moments (e.g. dunks, three-pointers and layups), defensive plays (e.g. steals, blocks) and key transitions like sudden shifts between offense and defense, free throws, timeouts, fouls, and strategic team movements [6], [7]. They are widely used for performance analysis, action detection, broadcast highlights, replays and video summarization by coaches, analysts and media teams.

Standard deep learning approach is useful in basketball video analysis because it can combine 3D Convolutional Neural Networks (CNNs) (for feature extraction) with Recurrent Neural Network (RNN) (for temporal modeling of sequential data commonly found with video frames. Studies establish that the above combinations can improve architectures for detecting transitions in basketball videos over time [8],[9]. Basketball is played on a rectangular court, which is divided into two halves by a mid-court line, with a hoop positioned at each end. The game is typically played indoors on polished wooden surfaces and outdoor courts that is made of concrete or asphalt. Games can be played in various team formats, including the traditional 5-on-5 and the increasingly popular 3-on-3 formation. New formats such as 4-on-4 and 6-on-6 have also been gaining attention in recent years. These team configurations can feature single-gender participants. All the team combination can be male players or female players. Similarly, they can be a mix of genders involving male and female players simultaneously playing together [10]. However, playing or watching several basketball games with the view of achieving the above benefits can be daunting tasks. Huge resources and avoidable overheads must be incurred before semantic understanding of all past video actions and reasonable results can made [11]. For these reasons, clips analysis is an integral method to study basketball videos. The actions of some players may unintentionally trigger highly sensitive to noise, camera angles, lighting changes and occlusion.

Moreover, it might be difficult for some stakeholders such as sport analysts, crowd, club owners, players involved and couch to immediately or simultaneously document and create histograms representing each clip while playing or watching basketball games or while studying basketball videos. The premise is that numerous elusive episodes, tiny and indiscernible clips that might be indescribable to human knowledge [12]. The degree at which every human being can recognize subtle differences between similar actions in basketball videos (or games) vary from one person to another. Extracting local features such as Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) from basketball videos with hands might not be accurately achieved without incurring time constraints or inadvertently ignoring temporal order of some frames. SIFT is used to detect and describe distinctive important points in basketball images or frames such as ball position, hand gestures and players' jersey numbers while SURF is used to identify motion patterns, jersey textures and team logo appearances.

Furthermore, these two methods rely heavily on visual clarity, camera movements, occlusions (e.g., when one player blocks another) and resolution of basketball footages. Consequently, exploring clip classification in basketball video datasets is a valuable endeavor due to the growing demand for automated sports analytics, efficient video summarization and intelligent retrieval of sports content. However, this research area is facing series of challenges over the years. For instance, clip classification in basketball videos required structured discussions, outlining its significance, the limitations and the advantages of core techniques like entropy thresholding for transition detection. Traditional approaches for clip classification in basketball videos often rely on handcrafted features, shallow learning models and rule-based heuristics [13]. While these methods were foundational in early video analysis research, they exhibit severe limitations in handling complex and fast-paced content found in modern basketball games. Some features are very difficult for human beings to manually extract from frames or shots to elucidate visual appearance and motions of players in basketball games. Most human knowledge can be domain-specific and hard to capture different sports formats in the same way cameras would do [14].

The objective of this paper is to pragmatically propose a new model for detecting transitions within basketball video clips to assist human analysts to automatically detect, understand or discern the points of sudden changes in the prediction of uncertainty and contents of high and low entropy to lessen the above challenges. With the aid of Python language and input basketball video datasets, this paper proposes a model that integrates R3D-18 to capture spatial-temporal features, CNNs to perform feature extraction and RNN to accomplish temporal modeling of offline basketball videos. It further uses entropy thresholding, mean Average Precision (mAP) and Top-k accuracy as primary metrics for clips classification of actions within input basketball video datasets. One of the substantial contributions of the model reported in this paper is that it successfully detects transition variability in clip 8 of basketball_video3 and clips 4 and 6 of basketball_video4, highlighting sudden changes in prediction uncertainty and content across the input datasets. The model also provides interpretable confidence scores for the predicted activities, demonstrating high certainty (99.52%) that the analyzed videos depict basketball-related actions. The remainders of this paper are organized as follows: section 2 discusses related work while section 3 broadly explains the concept of basketball video analysis. Section 4 highlights the methodology for conducting this research, section 5 presents findings and discussion of key results obtained while section 6 concludes and summaries the paper. This section also suggests areas of future research to extend our findings on basketball video clip analysis [15].

Basketball game is complex and dynamically evolving like other sports. Studies have shown that video clips analysis are emerging areas in coaching, player development, recruitment and scouting aspects of basketball games. Shallow Machine Learning Models (SMLMs) are approaches commonly used to extract features of basketball video clips [16], [17]. Once these models extract features either by handcrafted or aggregated technique, they further use shallow models (such as Support Vector Machines (SVM), Random Forests, K-Nearest Neighbors (KNN), Hidden Markov Models (HMM)) for clip classification [18], [19]. However, limitations of these models have raised serious concerns in recent time. There are numerous tools, algorithms and technologies developed to support basketball video footage analysis [20], [21]. Limited basketball knowledge and the lack of accessible to information required to encourage fans to wash basketball videos have been explored [22]. Synergy Sports are increasingly been used to analyze basketball video platforms. The strength of these methods include high-level of pro-level statistical tagging and scouting. However, one of their weaknesses is that game-watching behaviors can change over time. Proprietary tools can very expensive and primarily designed for famous professional teams such as National Collegiate Athletic Association (NCAA) and National Basketball Association (NBA). NacSport is a highly customizable and best platform for performance analysts that are willing to invest time. However, it requires manual tagging required, time-consuming and not friendly to new users. Since it has learning curve, it often requires training and previous knowledge to efficiently use it [23]. Dartfish is excellent for technical and motion analysis but more helpful to individual player development than team strategy.

Wearable devices that come with smart sensors have been integrated with basketball video to achieve advanced analytics [24]. The skepticism, key weaknesses and limitations of each of the existing platforms and lack of universal acceptance of wearable devices by basketball players and referees during training settings, contact tracing and fitness assessments have necessitated the development of Artificial Intelligence (AI), deep learning and Machine Learning (ML) models to complement the above basketball video analysis platforms [25]. However, most of the existing methods heavily dependent on high technical understanding before they can be incorporated in models designed to achieve clips classification in the basketball video analytics.

2. Literature Revies

2.1. The Concept of Basketball video analysis

Basketball video analysis refers to the process of studying and interpreting video footage of basketball games [26]. Such basketball footages are indications of many recorded clips that can include viewers, supporters' interference and basketball players during competition or practice sessions of some basketball players. Basketball video analysis is used to gain insights into performance, team strategies and game dynamics. It is widely used by players, coaches, analysts and scouts to improve individual and team performance. Basketball video analysis is a core part of basketball data science of players' performance and evaluation. It is a form of statistics tracking of points, rebounds, assists, turnovers, dunk, dribbling, shooting basketball, basketball layup, basketball block, etc. [27]. This domain can focus on clip movement analysis to understand the speed, agility, spacing and off-the-ball movement of individual player and the entire members of basketball team. Some basketball analysis can estimate the shooting form (e.g. shot segmentation), scoring classification and how shot selection and accuracy by location of the players aggregate to form efficiency and success of the team during practice sessions or competition.

Tactical and strategic insights that indicate team offense (or defense) within the duration of basketball game. Apart from helping couches to understand play execution, rotations, zone versus man-to-man defenses, it can show set plays, breakdowns and identify successful and failed plays. Scouting opponents is an integral part of analyzing tendencies and weaknesses basketball video footage at training sessions. There are several features used in basketball video analysis to detect and describe key points or regions of interest within individual video frames. These features are fundamental for tasks such as player tracking, ball detection, action recognition and event classification. A wide range of feature descriptors and representations are used in basketball video analysis [28]. These features can be grouped into traditional features and deep learning-based features. Traditional Features of basketball video clips include Histogram of Oriented Gradients (HOG), Optical Flow, Oriented FAST and Rotated BRIEF (ORB), Color Histograms and Motion History Images (MHI). Histogram of Oriented Gradients (HOG) is used to capture edge directions such as shape and silhouette of basketball clips. It can recognize upright players when contrasted with crouched postures during defense. Analysis of HOG can be used in player segmentation and human detection such as player pose. Image segmentation and Optical Flow have been used to measure pixel-level motion between frames in basketball video footage. Analysts use optical flow to track player movement, ball motion and team formation dynamics, especially if they want to distinguish fast breaks from defensive shifts [29]. Oriented FAST and Rotated BRIEF (ORB) can reveal fast binary feature descriptor and rotation-invariant in basketball video footage (or clips). It is used for real-time tracking of basketball during passes and fast key-point matching. Color Histograms illustrate the distribution of colors in specific regions of basketball video footage. It can be used to discriminate Jersey color identification of players and team differentiation. This feature is specifically useful in exploring the segmentation of basketball players from opposing teams. Motion History Images (MHI) represents recent movement in basketball footage. However, they only focus on action recognition and clustering of actions like shooting, dribbling, dunk movement and so on in basketball video clips.

3.2. Architecture for AI Models in Basketball Video Clip Classification

Standard Deep Learning-Based Features (in Figure 1) can be used to design high-performance basketball video systems. These models integrate features learned by Neural Networks (NNs) such as Convolutional Neural Networks (CNNs) features with 3-dimensional Convolutional Neural Networks (3D CNNs), pose estimation and transformers [30].

A Convolutional Neural Network (CNN) is a category of deep learning model precisely designed for processing visual video footages and images. CNN learns to extract features such as shapes, edges and patterns from basketball videos and automatically use the extracted feature to detect and perform ball classification, court segmentation and so on the input basketball videos based on those features. The 3-dimensional Convolutional Neural Networks (3D CNNs) can capture spatio-temporal patterns over multiple basketball video frames. It is specifically effective in action recognition such as shot attempt and event detection even from a short clip. Pose estimation is basically a form of computer vision technique used to detect and track the positions of key body joints and body motion (e.g. shooting postures and defensive block), player stance, gesture, elbows, knees and wrists from basketball images or videos. It's commonly used in applications like human-computer interaction, augmented reality, sports analysis and surveillance.

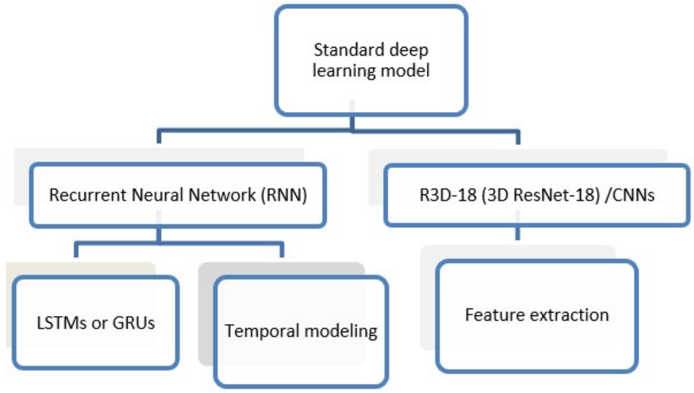

Figure 1. Architecture of Deep Learning Approach in Basketball Video Analysis

Visual Geometry Group (VGG) and Residual Network (ResNet) are common examples of CNN features. VGG was led by Professor Andrew Zisserman and developed by the Oxford's Visual Geometry Group (VGG) in the early 2000s at the Department of Engineering Science at the University of Oxford. Its main characteristics subsume very deep network up to 19 layers. It also uses 3x3 convolutional filters consistently across layers and it is easy to understand and implement. However, its main weakness is that it requires large model size and many of computational power and memory. Hence, these drawbacks make it slow to train and run. Though, VGG is useful in Basketball video analysis to extract spatial features from frames such as player positions and court layout, but then, it is less efficient than newer models. Residual Network (ResNet) was developed by Microsoft Research [31]. ResNet introduces residual connections to allow very deep networks (even 100+ layers) to train effectively. ResNet was a major breakthrough in deep learning because it addresses the degradation problem in Very Deep Neural Networks (VDNNs) by introducing residual (or skip) connections). The original ResNet architecture included versions with 18, 34, 50, 101 and 152 layers. It solves the vanishing gradient problem that older deep networks such as VGG cannot solve. ResNet is very popular for extracting features like identifying players, detecting actions (e.g. passes, shots, etc.) and tracking movement across frames from video frames. However, ResNet has complex architecture and it is harder to understand for beginners. This is usually computationally expensive for real-time applications if it is not properly optimized.

Essentially, features the models commonly extract from basketball videos with the above models can include player coordinates (x, y on court) position, player trajectories (movement and pose), speed and acceleration, ball location and trajectory indicating ball possession, shot type (layup, 3-pointer, dunk, etc.), steals, blocks, pass sequences (or patterns), turnovers, fouls, violations, etc. These features can be extracted using techniques such as pose estimation, object tracking, optical flow, convolutional neural networks (CNNs) and deep learning-based action recognition as well as pre-trained models like ResNet or EfficientNet. Research reveals that Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), or 3D CNNs are employed cooperatively to classify motion and sequence information in basketball videos [32]. Some architectures may not necessarily incorporate attention mechanisms and transformers to enhance temporal understanding and improve classification accuracy. Recurrent Neural Network (RNN) is a type of Artificial Neural Network (ANN) designed for processing sequential data especially where the sequence of actions matters. RNNs keep track of activities or events that have been calculated so far. RNNs use the same weights across to generalize across sequence lengths in all time steps. RNNs prove effective in processing input data in the form of step-wise clip but not good for extracting step-wise clip until all the sequences are completely processed.

3.3. Evaluation of Models for Classifying Basketball Video Clips

Exploring clip classification in basketball video datasets is a valuable endeavor due to the growing demand for automated basketball analytics, efficient basketball video summarization and intelligent retrieval of basketball's content [33]. The algorithm used in this performs basketball video clip classification using R3D-18 and Recurrent Neural Network (RNN) as sketched below.

Algorithm 1. Basketball Video Clip Classification Using R3D-18 + RNN.

```
BEGIN
# Step 1: Initialize Models
Load R3D18_Model with pretrained weights
Initialize RNN_Classifier
#Step 2: Process Videos
For each video in [basketball_Video1, basketball_Video2, basketball_Video3, basketball_Video4] do:
   ## Step 2.1: Clip Segmentation
   Clips ← DivideVideoIntoClips(video, clip_length)
   ##  Step 2.2: Feature Extraction
   Clip_Features ← []
   For each clip in Clips do:
      features ← R3D18_Model.extract_features(clip)
      Clip_Features.append(features)
   End For
   ## Step 2.3: Classification
   Video_Prediction ← RNN_Classifier.predict(Clip_Features)
   ## Step 2.4: Top-5 Predictions
   Top5_Predictions ← GetTopKPredictions(Video_Prediction, k=5)
   ## Step 2.5: Entropy Analysis
   Entropy_Value ← ComputeEntropy(Video_Prediction)
   Is_Repetitive ← (Entropy_Value < entropy_threshold)
   ## Step 2.6: Save Results
   SaveToResultsFile(video_name=video.name,
                predictions=TopK_Predictions, K= 1, 3, 5 & 10
                entropy=Entropy_Value,
                repetitive=Is_Repetitive)
End For
# Step 3: Repeat or Terminate
Repeat the above processes for each basketball_Video
     While no basketball_Video to process
     Terminate
END
```

This research uses entropy thresholding, mean Average Precision (mAP) and Top-k accuracy as primary metrics for clips classification of four basketball video datasets. Entropy thresholding is a data-driven statistical performance metric used to evaluate the uncertainty (information content) in basketball video frames. Entropy gives higher values for diverse clip predictions and lower values for repetitive actions in basketball video clips [2]. Mathematically, entropy is computed as Equation 1.

$$Entropy(p) = -\sum_{j=1}^{n} p(x_j)\log p(x_j) \tag{1}$$

This metric is useful in detecting clip boundaries and transitions for the purpose of discovering the point at which team actions suddenly change in every basketball video. Since entropy measures the amount of randomness (or disorderliness) in data, a sudden change in entropy values of basketball player's team can indicate segment borders (or limits), replay and cut where significant transition of actions in the video has occurred. Suppose the entropy values for 10 clips are 0.5, 0.6, 1.8, 0.7, 0.6, 0.4, 0.3, 1.9, 0.5 and 0.4 and transition benchmark is set to 1.5 seconds. The third entropy, i.e. is 1.8 is higher than its immediate left and right neighbors (i.e. 0.6 and 0.7). Similarly, the eighth entropy i.e. is 1.9 which is higher than its immediate left and right neighbors (i.e. 0.3 and 0.5). These suggest that clips 3 and 8 are transition points. The transitions at clip clips 3 and 8 means these two clips show peak in prediction uncertainty and they suggest possible changes in the activities of players in a given basketball video. Besides, understanding of this segment boundaries can enable video editor to avoid errors due to arbitrary clip cuts. Actions in basketball games are not often distributed evenly over time. Entropy thresholding can improve clip relevance. It can enable coaches, players and video editors to ensure segmented basketball videos result into clips that are semantically consistent and achieve high classification accuracy.

Mean Average Precision (mAP) is a performance metric that can be used to rank and predict all actions (clips) detected in basketball videos [5]. This metric is used in this paper to measure how well the proposed model ranks the occurrence of every correct class of all actions identified in the basketball videos. Mean Average Precision calculates the average precision for each class and then computes its average over all classes. Mathematically, mAP is computed as Equation 2.

$$\text{mAP} = \frac{1}{N}\sum_{j=1}^{N} AP_j \tag{2}$$

$$\text{Top-k: } \{j_1, \ldots, j_k\} \text{ such that } x_{j_1} \geq x_{j_2} \geq \cdots \geq x_{j_k} \tag{3}$$

The interpretation is that higher mAP indicates better precision and recall balance across all classes observed in a given basketball video. Top-p Accuracy ranks the correct class label on how they appear in the model's top p classes in the evaluation datasets [34]. Mathematically, Top-p Accuracy is calculated as shown below: For instance, in Top-5 Accuracy, the correct label must rank (belong to) the 5 topmost actions in each clip of the basketball video. Similarly, in Top-1 Accuracy, the correct label must be the topmost prediction of all clips in the basketball video. However, clip-wise prediction can disclose some topmost actions that vary from one clip to another even in the same basketball video and across different basketball videos.

3. Methodology

We use Python language and associated libraries shown in the table below to implement the above algorithm. The model comprises pretrained rules, initialization rules, video-level classification rules, clip classification rules, metric computation rules and results rules. The model begins by using its pretrained rules to load the R3D-18 model with pretrained weights required for clip-level feature extraction. Thereafter, model uses its initialization rules to initialize the RNN-based classifier. This action makes the classifier to handle temporal modeling and perform video-level classification. The video-level classification rules are used to classify four online basketball videos (i.e. basketball_video1, basketball_video2, basketball_video3, and basketball_video4).

Table 1. Summary of Model Components and Python Libraries

S/N	Component	Python Libraries Used
1	R3D-18 (3D CNN)	torch, torchvision
2	CNNs (spatial features)	torch, torchvision
3	RNNs (temporal modeling)	torch.nn (LSTM, GRU)
4	Entropy thresholding	numpy, scipy.stats
5	Clip-wise & video-level evaluation	sklearn.metrics, pandas
6	Accuracy metrics (mAP, Top-k)	sklearn.metrics, custom metric functions
7	Video processing	cv2, moviepy, or decord

For each video, the rules segment the video into fixed-length clips to optimize clip classification. Its clip classification rules receive each clip and passed it through the R3D-18 model to extract deep visual (or spatial-temporal) features. The rules CNNs to perform feature extraction from all clips aggregate and fed them into the RNNs classifier to generate a prediction for the full video.

The model then uses its metric computation rules to generate average entropy, average confidence, the top-1 class, top-3 class top-5 class and top-10 class predictions based on the video-level output. It uses a threshold of 1.5 detect transition after computing entropy of the prediction distribution that evaluates the confidence and repetitiveness of the prediction. The result rules organize the outputs from the above metrics into the video name, top-5 predictions and entropy value. The rules subsequently trigger a flag indicating whether the prediction is repetitive and saved all these results in four respective files. Finally, the above processes are repeated for each of the above four basketball videos and can be extended or repeated further if required. The main results obtained in the above simulations are shown and exhaustively discussed in the next section.

4. Findings and Discussion

Clip-wise prediction details for basketball_Video1 to basketball_Video4 were 27, 11, 8 and 10 respective. Tables 2 and 3 illustrate the assessment of the model's performance on the content of all the four basketball videos used for evaluation of the proposed mode. Tables 2 and 3 indicates top-k (top-10) classification of actions by their probabilities in the above four datasets as Basketball games.

Table 2. Activity Probabilities for Basketball_Video1 and Basketball_Video2

Activity	Probabilities	
	Basketball_Video1	**Basketball_Video1**
Basketball	0.9952	0.993
Skateboarding	0.0007	0.0011
Tennis	0.0007	0.001
Running	0.0006	0.0009
Soccer	0.0005	0.0008
Cycling	0.0005	0.0008
Swimming	0.0005	0.0007
Boxing	0.0005	0.0007
Skiing	0.0004	0.0005
Billiards	0.0004	0.0005

Based on Table 2, the values are 0.9952 (99.52%), 0.9930 (99.30%), 0.9938 (99.38%) and 0.9932 (99.32%) respectively. The model is extremely confident that the activities in all these videos are relevant to basketball. Visual analysis of these datasets corroborates the above results. For instance, these findings suggest that all the visual and motion features across the clips in all the above datasets strongly match the learned patterns for basketball games. The remaining activities such as Skateboarding, Tennis, Running, Swimming, Boxing and Skiing have very low probabilities that range from 0.0004 to 0.0007 in all the basketball videos. These values indicate that the model finds such classes of activities negligibly. This implies that there is little or no confusion between basketball and other activities in all the datasets.

Table 3. Activity Probabilities for Basketball_Video3 and Basketball_Video4

Activity	Probabilities	
	Basketball_Video1	**Basketball_Video1**
Basketball	0.9938	0.9932
Skateboarding	0.001	0.0011
Tennis	0.0009	0.001
Running	0.0008	0.0009
Soccer	0.0007	0.0007
Cycling	0.0007	0.0007
Swimming	0.0007	0.0007
Boxing	0.0006	0.0007
Skiing	0.0004	0.0005
Billiards	0.0004	0.0005

International Journal of Artificial Intelligence, vol. 13, no. 1, pp. 1-15, June 2026. **DOI: 10.36079/lamintang.ijai-01301.883**

Essentially, the results suggest that the model performs clip-wise and video-level action classification with exceptional confidence, achieving 99.92% and 99.52% certainty that all the evaluative footages depict basketball videos. Figure 2 to Figure 6 illustrate the video-wise classification of the above four datasets. The entropy of basketball_Video4 is 3.1976. The dataset has lowest entropy. This means the model made very confident and consistent predictions across its clips. The dataset is mostly likely to be the cleanest and least ambiguous among other three basketball videos. The entropy of basketball_Video2 is 3.5341 while Basketball_Video3 has 3.7912. Moderate entropy values suggest fairly confident predictions. However, these values are less variable compared to Basketball_Video4. Similarly, the entropy of Basketball_Video1 is 4.787. This dataset has the highest entropy. This indicates the model experienced more uncertainty across clips and this could possibly due to noisy content, unclear actions and footage that are less famous among be basketball games.

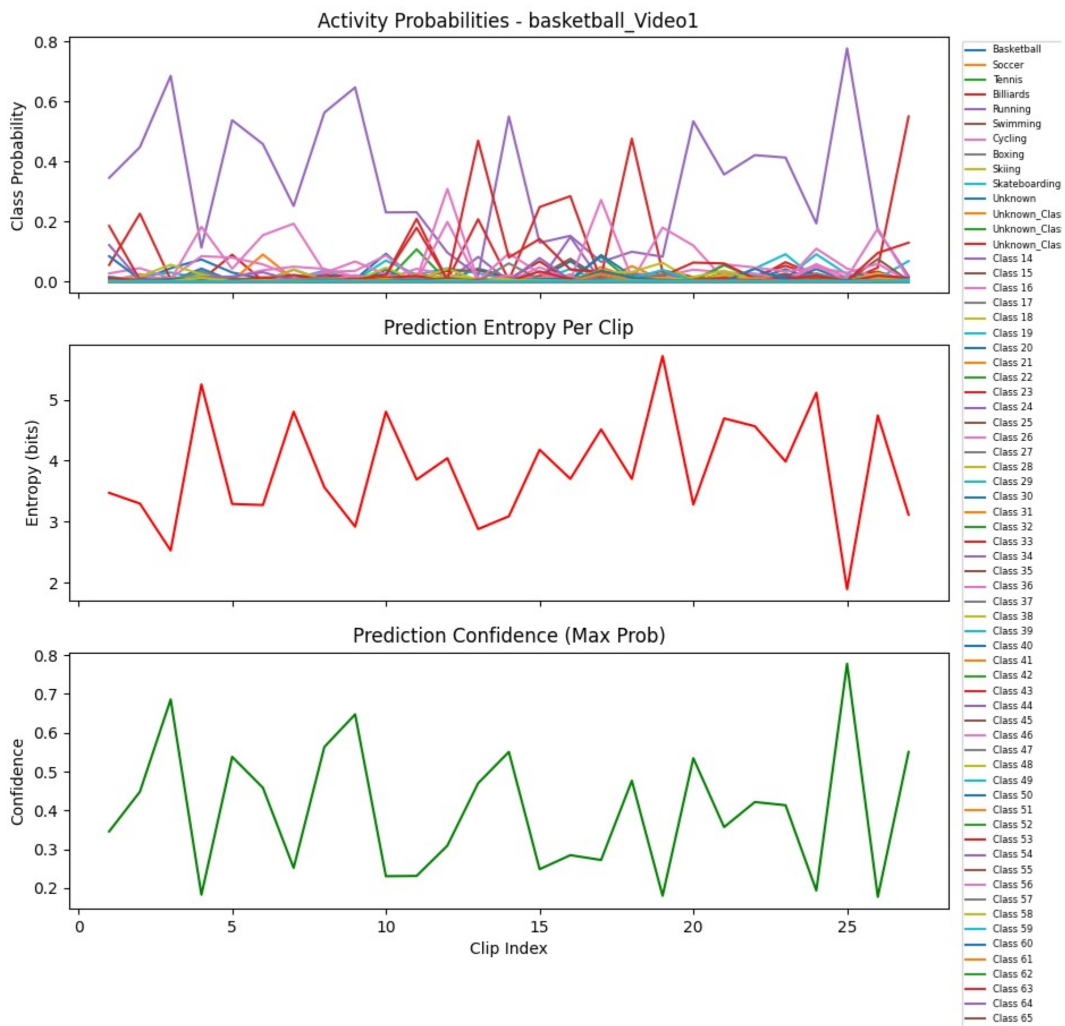

Figure 2. Performance of the Combined Metrics on Basketball video1

The training loss of the proposed deep learning model was measured across five training epochs and specifically during the analysis of basketball_Video1 to basketball_Video4. For epoch 1/5, its loss was 1.9757 while for epoch 2/5, its loss was 0.9749. Similarly, for epoch 3/5, its loss was 0.7231, for epoch 4/5, its loss was 0.5785 and for epoch 5/5, its loss was 0.4372. Loss essentially measures how poorly (or well) the proposed model was performing. It is a single number that reflects the difference between the predicted output and the true (i.e., ground truth) output. A higher loss indicates worse performance while a lower loss means better performance. Each epoch means one complete pass through the training data. Loss is calculated on the training data during each epoch. These results show the loss decreases consistently over the epochs. This trend indicates that the model is learning and improving in its predictions. Meanwhile, a drop in loss from 1.9757 to 0.4372 suggests a significant improvement in the ability of the proposed model to analyze the above four basketball videos. Average loss is equivalent to (1.9757+0.9749+0.7231+0.5785+0.4372)/5 or 0.9379. This is a positive sign and it the proposed model is learning effectively, and improving at classifying actions in the basketball videos with each epoch reducing the error. Average loss is the mean value of the loss function over a batch, an epoch or the entire basketball video dataset during training or evaluation. It reflects how well the model's predictions match the ground truth labels on average. Average loss is used to monitor training progress of the model. A lower average loss basically indicates better model performance while comparing model performance across experimental simulations.

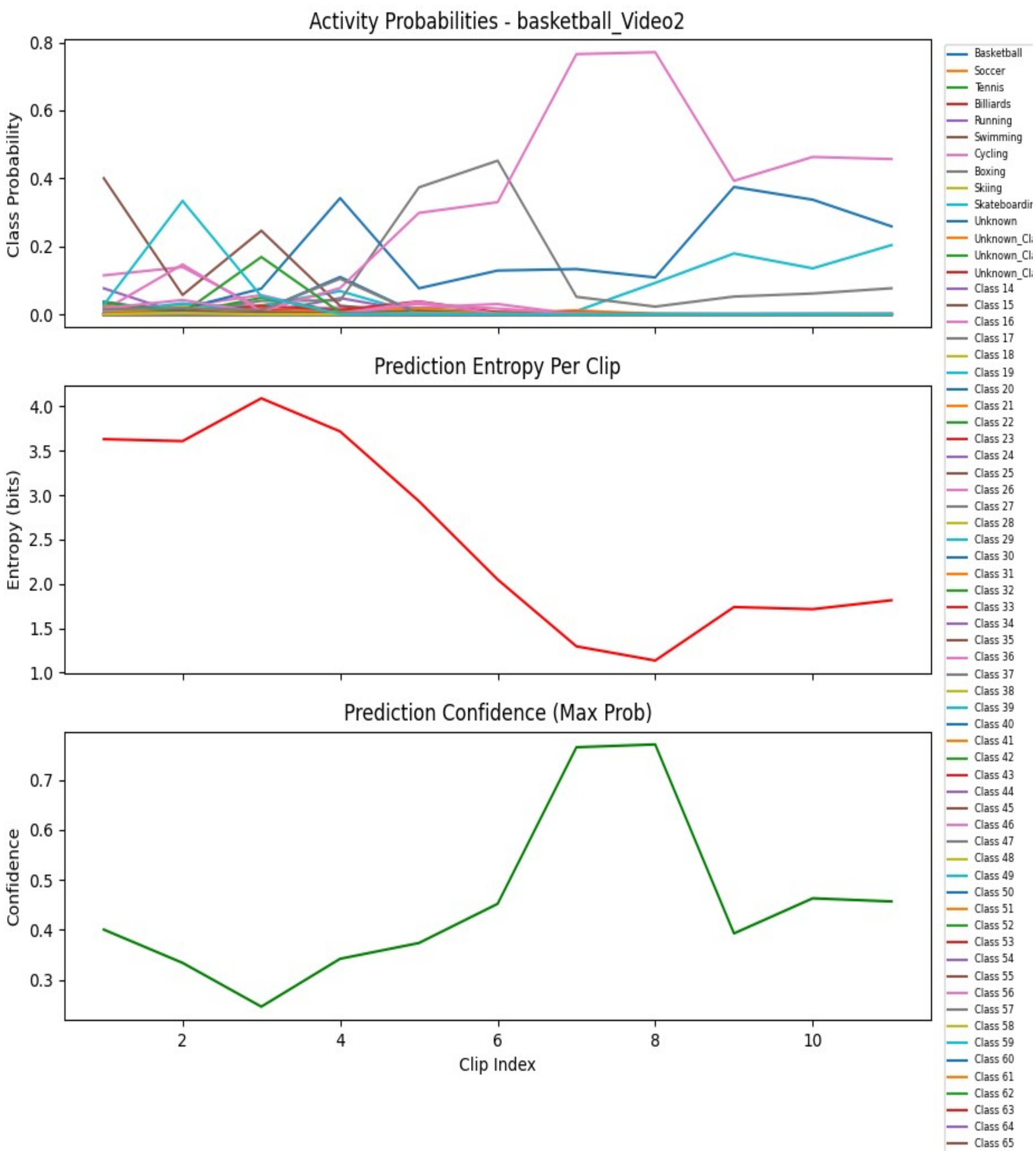

Figure 4. Performance of Combined Metrics on basketball_Video2

The average entropy scores of 4.787, 3.5341, 3.7912 and 3.1976 across datasets were reported, reflecting potentially elusive variations and uncertainties within clip-level predictions due to nuanced activity shifts despite strong overall (93.3%) classification confidence. Figure 7 suggests the top-10 predicted classes by the RNN model arranged by confidence or probability score. The figure shows how confident the standard deep learning model is in assigning the video to each class. Basketball was 0.9992. The proposed model is very certain (or confidence) that the video shows basketball. It further shows low confidence in running (0.0004) and very low confidence in Tennis or Soccer (0.0002) and Skateboarding (0.0001).

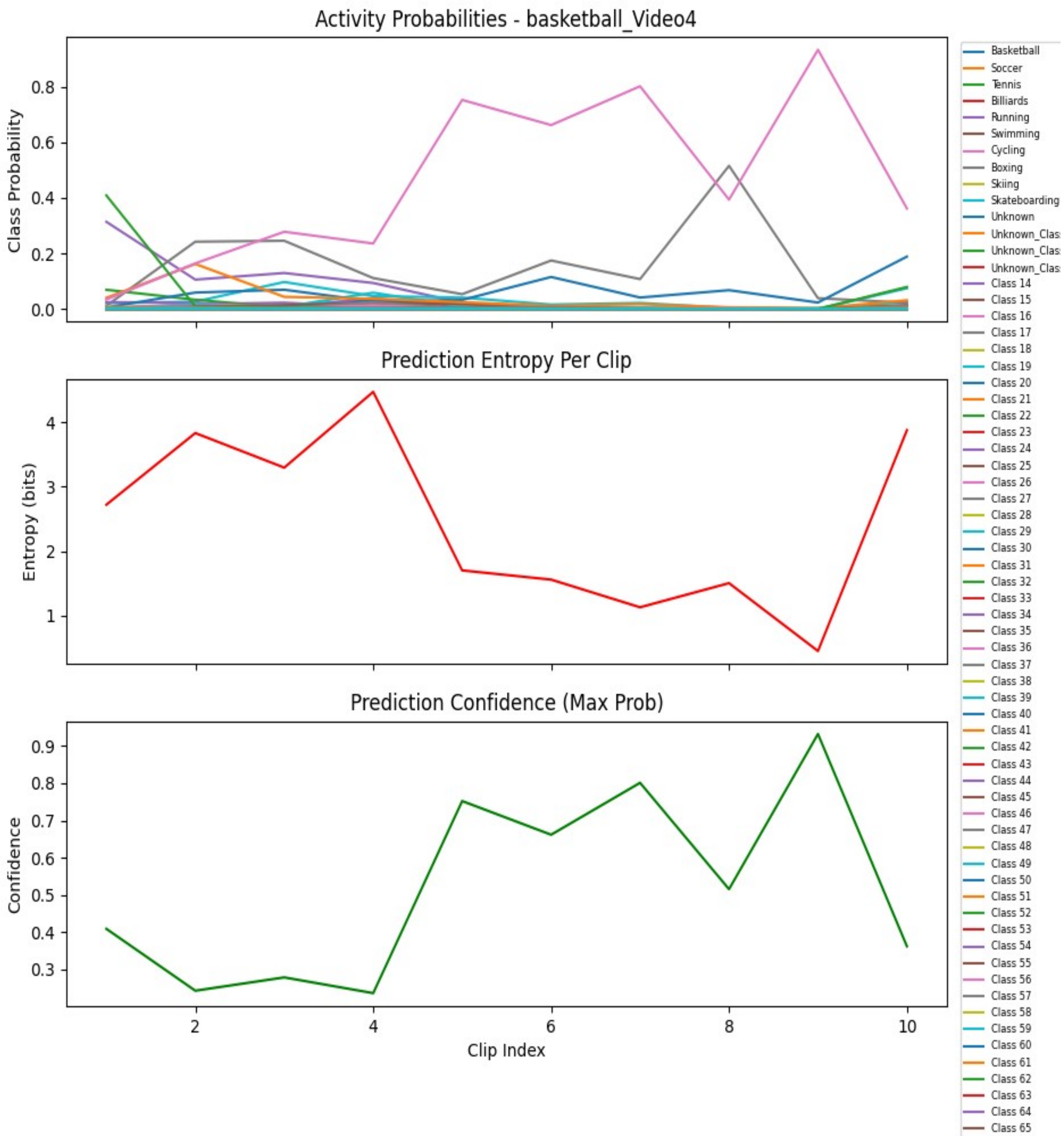

Figure 5. Performance of the Combined Metrics on basketball_Video 3

In addition, clip-wise prediction details for 27 clips found in basketball_Video1 are 3.4738, 3.2983, 2.5299, 5.2502, 3.2916, 3.2743, 4.8023, 3.5642, 2.9196, 4.8021, 3.6910, 4.0412, 2.8802, 3.0903, 4.1800, 3.7023, 4.5120, 3.7006, 5.7132, 3.2830, 4.6930, 4.5646, 3.9858, 5.1134, 1.8964, 4.7413 and 3.1162. The model did not detect transitions at any clip of the dataset. Clip-wise prediction details for 11 clips found in basketball_Video2 are 3.6316, 3.6095, 4.0919, 3.7188, 2.9326, 2.0486, 1.2949, 1.1348, 1.7379, 1.7142 and 1.8148. The model detected transitions at clips. Clip-wise prediction details for 8 clips found in basketball_Video3 are 3.6533, 2.7872, 2.2494, 4.1781, 4.5472, 0.9753, 0.9855 and 0.9231.

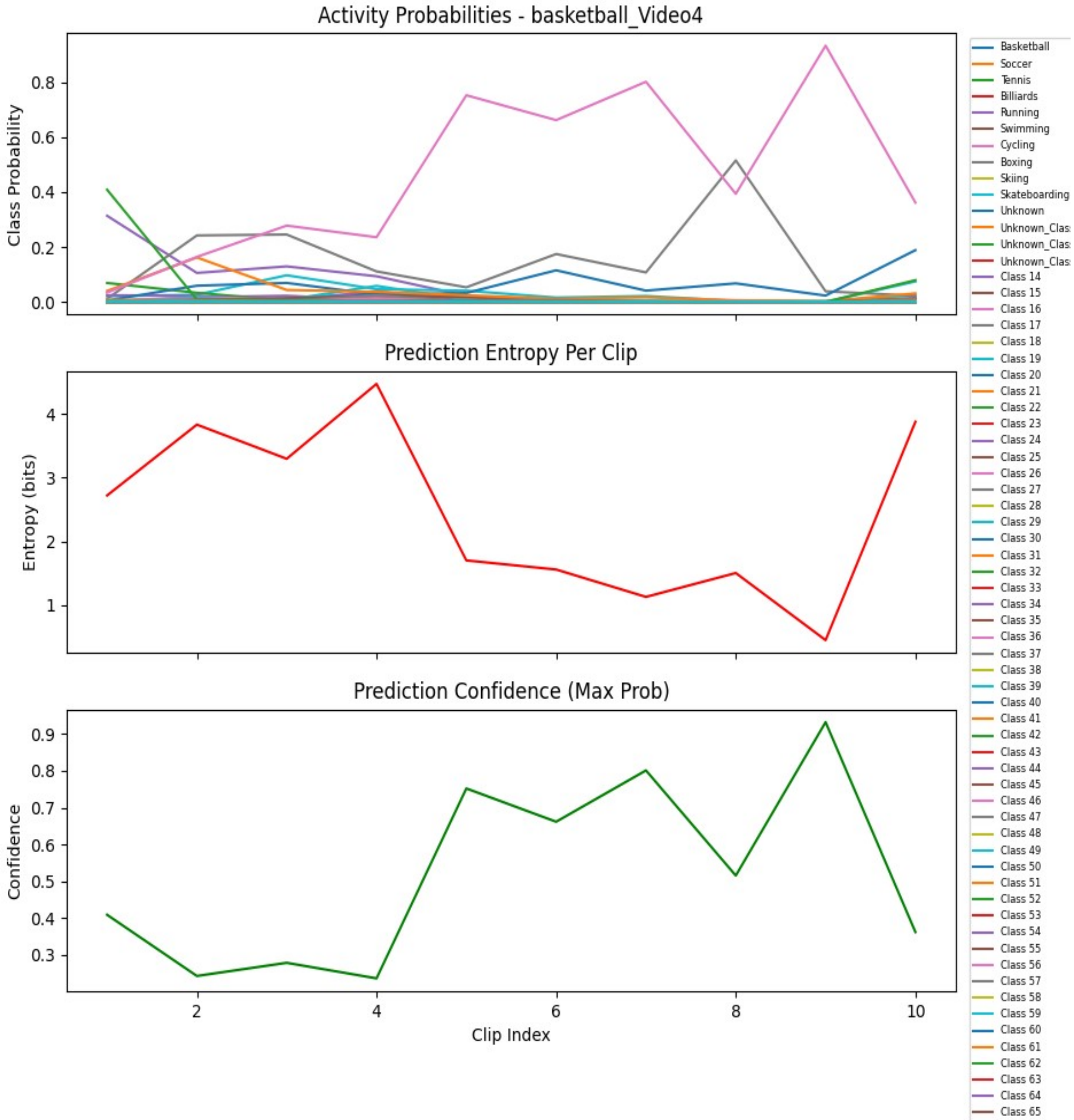

Figure 6. Performance of the Combined Metrics on basketball_Video4

The model did not detect transitions at any clip. Clip-wise prediction details for 10 clips found in basketball_Video4 are 2.7196, 3.8305, 3.2946, 4.4689, 1.7018, 1.5589, 1.1298, 1.5043, 0.4521 and 3.8748 respectively. Detected transitions at clips. The implications of the above results are many. The top-k (top-10) Accuracy and RNN confidence (i.e., over (99.3%) that the above four evaluative datasets are basketball videos. Entropy basically measures uncertainty in the model's clip-wise predictions. Lower entropy clips the threshold below 1.5 indicate higher confidence while higher entropy means less confident and more uncertain predictions. From Table 6, these results indicate that the model detected no transitions in basketball_video1 and basketball_video3 while notable transitions appeared at clips 8 and 7 in basketball_video2 and basketball_video4. These respectively corresponding to spikes in uncertainty likely caused by activity changes like switching from basketball to running.

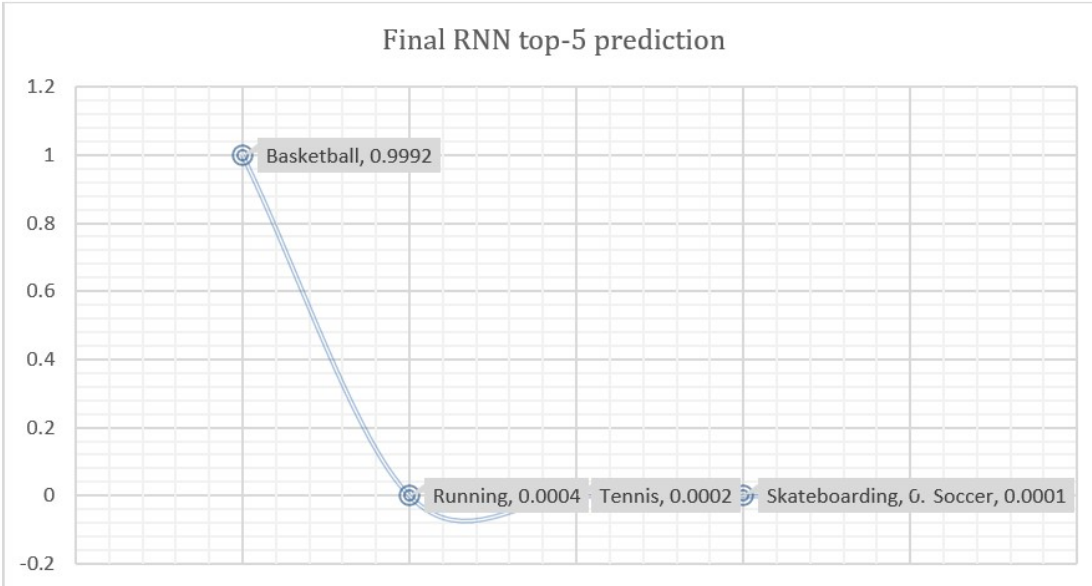

Figure 7. Final RNN Top-5 Prediction for Basketball Videos

5. Conclusion

Clip classification in basketball video datasets is a highly impactful task. Basketball is a fast-paced, dynamic sport characterized by rapid transitions, diverse camera angles and complex actions. Automating the classification of video clips into semantically meaningful categories such as offense, defense, replay, free throws and timeouts can offers numerous practical applications to couches, team members and fan or supporters in reviewing past performance with precision. It can allow broadcasters to generate automatic highlights and summaries. Nevertheless, conventional methods often show low accuracy with dynamically complicated sports footages. Thus, this paper explores the integration of entropy thresholding to achieve adaptive transition detection by leveraging on deep learning architectures that model spatiotemporal dependencies. The paper further shows researchers can significantly improve the performance and reliability of automated basketball video analysis systems with the use of standard deep learning models.

The standard deep learning model detected transitions in (clips 7 and 8 in Videos 2 and 4) that are likely indicative of moments whereby the model perceived sudden changes in actions of players captured both footages. It was observed that lower entropy clips (e.g., below 1.5) indicate higher confidence predictions while higher entropy clips (above 4) can suggest ambiguous and complex frames. The RNNs classified all the above four basketball videos with extremely high confidence. The other actions in the footages have insignificant probabilities, suggesting the model has learned to distinguish basketball from similar sports like tennis, swimming and soccer. This also suggests that integrated model from 3D ResNet feature extraction to RNNs sequence modeling is working correctly. Nonetheless, four basketball videos may not big enough for learning a robust representation about all basketball videos. One way to improve the accuracy of the final RNNs top-k prediction for all the basketball videos is to add non-basketball videos from other classes so that the above standard deep learning model can learn contrast. Hence, these will be our future research to learn long-term dependencies in basketball video sequences.

Acknowledgements

This research is fully supported by Tertiary Education Trust Fund (TETFUND) Grant. The authors specially acknowledged TETFUND for the approved fund which makes this important research viable and effective.

References

[1] J. De Bock and S. Verstockt, "Video-Based Analysis and Reporting of Riding Behavior in Cyclocross Segments," *Sensors*, vol. 21, no. 22, Art. no. 7619, 2021. doi: 10.3390/s21227619.
[2] J. Zhang, H. Zeng, L. Zhang, J. Feng, J. Li, X. Chen, and Y. Cheng, "Deep Learning-Based Automated Tactical Clip Extraction and Performance Evaluation of Basketball Game Videos," *Appl. Sci.*, vol. 16, no. 11, Art. no. 5683, 2026. doi: 10.3390/app16115683.
[3] K. Q. Fang and H. Jiang, "A machine learning framework for evaluating basketball player skill development using video analysis," *Egypt. Inform. J.*, vol. 32, Art. no. 100762, Dec. 2025. doi: 10.1016/j.eij.2025.100762.
[4] D. Rösch, F. Schultz, and O. Höner, "Decision-Making Skills in Youth Basketball Players: Diagnostic and External Validation of a Video-Based Assessment," *Int. J. Environ. Res. Public Health*, vol. 18, no. 5, Art. no. 2331, Feb. 2021. doi: 10.3390/ijerph18052331.
[5] K. Host and M. Ivašić-Kos, "An overview of Human Action Recognition in sports based on Computer Vision," *Heliyon*, vol. 8, no. 6, Art. no. e09633, Jun. 2022. doi: 10.1016/j.heliyon.2022.e09633.
[6] M. Ausloos, "Shannon Entropy and Herfindahl-Hirschman Index as Team's Performance and Competitive Balance Indicators in Cyclist Multi-Stage Races," *Entropy*, vol. 25, no. 6, Art. no. 955, 2023. doi: 10.3390/e25060955.
[7] G. Bertasius, H. S. Park, S. X. Yu, and J. Shi, "Am I a Baller? Basketball Performance Assessment from First-Person Videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2177–2185. doi: 10.1109/CVPR.2016.239.
[8] Z. Cui, "3D-CNN-based Action Recognition Algorithm for Basketball Players," *Informatica*, vol. 48, no. 13, pp. 97–110, Sep. 2024. doi: 10.31449/inf.v48i13.6100.
[9] Y. Feng and H. Sun, "Basketball Footwork and Application Supported by Deep Learning Unsupervised Transfer Method," *Int. J. Inf. Technol. Web Eng.*, vol. 18, no. 1, Art. no. 334365, Jan. 2023. doi: 10.4018/IJITWE.334365.
[10] M. Lippi, R. F. Carpio, M. Contarini, S. Speranza, and A. Gasparri, "A Data-Driven Monitoring System for the Early Pest Detection in the Precision Agriculture of Hazelnut Orchards," *IFAC-PapersOnLine*, vol. 55, no. 32, pp. 42–47, 2022. doi: 10.1016/j.ifacol.2022.11.112
[11] A. A. Sangüesa, C. Ballester, and G. Haro, "Single-Camera Basketball Tracker through Pose and Semantic Feature Fusion," Jun. 2019, *arXiv:1906.02042*. doi: 10.48550/arXiv.1906.02042.
[12] Q. Zhong, "Application of digital-intelligent technologies in physical education," *Front. Public Health*, vol. 13, Art. no. 1420516, Jan. 2025. doi: 10.3389/fpubh.2025.1420516.
[13] L. Zhao, Z. Lin, R. Sun, and A. Wang, "A Review of State-of-the-Art Methodologies and Applications in Action Recognition," *Electronics*, vol. 13, no. 23, Art. no. 4733, 2024. doi: 10.3390/electronics13234733.
[14] C. Hong, "Basketball Video Image Segmentation Using NeutrosophicFuzzy C-means Clustering Algorithm," *Informatica*, vol. 48, no. 9, pp. 143–154, Jun. 2024. doi: 10.31449/inf.v48i9.5929.
[15] S. Spiegler, R. Hoda, and A. Pant, "Images of AI: How AI practitioners view the impact of Artificial Intelligence on society, now and in the future," *Technol. Soc.*, vol. 84, Art. no. 103109, Mar. 2026. doi: 10.1016/j.techsoc.2025.103109.
[16] J. M. Alves and R. S. Barbosa, "Machine Learning for Basketball Game Outcomes: NBA and WNBA Leagues," *Computation*, vol. 13, no. 10, Art. no. 230, Oct. 2025. doi: 10.3390/computation13100230.
[17] T. Yang, C. Jiang, and P. Li, "Video Analysis and System Construction of Basketball Game by Lightweight Deep Learning under the Internet of Things," *Comput. Intell. Neurosci.*, vol. 2022, Art. no. 5565319, Mar. 2022. doi: 10.1155/2022/5565319.
[18] K. C. Paul, D. Waldmann, C. Chen, Y. Wang, and T. Zhao, "Artificial Intelligence for DC Arc Fault Detection in Photovoltaic Systems: A Comprehensive Review," *IEEE Access*, vol. 13, pp. 1–21, Jan. 2025. doi: 10.1109/ACCESS.2025.3526144
[19] T. Kopalidis, V. Solachidis, N. Vretos, and P. Daras, "Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets," *Information*, vol. 15, no. 3, Art. no. 135, Mar. 2024. doi: 10.3390/info15030135.
[20] P. Pietraszewski, A. Terbalyan, and A. Gołaś, "The Role of Artificial Intelligence in Sports

Analytics: A Systematic Review and Meta-Analysis of Performance Trends," *Appl. Sci.*, vol. 15, no. 13, Art. no. 7254, Jun. 2025. doi: 10.3390/app15137254

[21] J.-H. Kim, J. Kim, H. Kang, and B.-Y. Youn, "Ethical implications of artificial intelligence in sport: A systematic scoping review," *J. Sport Health Sci.*, vol. 14, Art. no. 101047, Dec. 2025. doi: 10.1016/j.jshs.2025.101047

[22] Z. Zhou, "Basketball game tactics decision modeling using video analytics for action recognition," *Proc. Inst. Mech. Eng. P: J. Sports Eng. Technol.*, Nov. 2025, doi: 10.1177/1754337125135266.

[23] Z. Shao, "Retraction Note: Individual characteristics of physical and mental development and their connection with regular physical exercises when playing basketball," *Sci. Rep.*, vol. 14, Art. no. 14321, Jun. 2024. doi: 10.1038/s41598-024-65123-y.

[24] J. Liu, "EITNet: An IoT-enhanced framework for real-time basketball player performance analysis," *Measurement*, vol. 242, Art. no. 115712, Jan. 2025. doi: 10.1016/j.measurement.2024.115712.

[25] Z. Wu, M. Xu, and C. Qu, "The use of wearable devices in the assessment of basketball players' athletic performance," *J. Sports Sci. Med.*, vol. 24, no. 1, pp. 88–99, Mar. 2025. doi: 10.389840585/jssm.2025.0324.

[26] F. Chen, "Deep Learning-Based Behavior Analysis in Basketball Video," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 3, pp. 412–419, Mar. 2025. doi: 10.14569/IJACSA.2025.0160345.

[27] J. Pino-Ortega, D. Lapuente-Salar, and M. D. M. Goya-Arana, "Observational Analysis in Basketball: A Literature Review," *J. Hum. Kinet.*, vol. 99, no. 1, pp. 215–228, Aug. 2025. doi: 10.2478/hukin-2025-0042.

[28] M. Caruso, L. Cimmino, F. Narducci, C. Pero, and G. Ronga, "Advancements in basketball action recognition: Datasets, methods, explainability, and synthetic data applications," *Image Vis. Comput.*, vol. 162, Art. no. 105689, Oct. 2025. doi: 10.1016/j.imavis.2025.105689.

[29] J. Wang, "Basketball technique action recognition using 3D convolutional neural networks," *Sci. Rep.*, vol. 14, Art. no. 10254, Apr. 2024. doi: 10.1038/s41598-024-59281-2.

[30] L. Xiao, "Recognizing sports activities from video frames using deep learning frameworks," *Multimed. Tools Appl.*, vol. 82, no. 14, pp. 21455–21473, Jun. 2023. doi: 10.1007/s11042-023-14561-2

[31] L. F. Rossi, G. M. Di Nunzio, and G. A. Susto, "How AI-based computer vision algorithms are impacting the tactical analysis of team sports," *Signal Image Video Process.*, vol. 19, no. 4, pp. 1245–1255, Apr. 2025. doi: 10.1007/s11760-025-04123-x.

[32] M. Trigka and E. Dritsas, "A Comprehensive Survey of Machine Learning Techniques and Models for Object Detection," *Sensors*, vol. 25, no. 1, Art. no. 214, Jan. 2025. doi: 10.3390/s25010214.

[33] A. Shahverdy, M. Fathy, and A. Saberi, "Basketball-51: A Video Dataset for Activity Recognition in Basketball Games," *IEEE Access*, vol. 9, pp. 162145–162154, Dec. 2021. doi: 10.1109/ACCESS.2021.3132047

[34] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015. doi: 10.1007/s11263-015-0816-y.