International Journal of Advances in
Soft Computing and Intelligent Systems (IJASCIS)
2024, Vol 03, Issue 02, 251-262
ISSN: 3048-4987

Original Paper

A HYBRID APPROACH FOR CRICKET HIGHLIGHT GENERATION BASED ON PLAYER-SPECIFIC ANALYSIS

Tariq Shah[a], Rabia Maham[a], Naeem Ahmed[b*], Muhammad Saeed[b]

[a] Department of Computer Science, University of Engineering and Technology Taxila Pakistan.
[b] School of Software, Nanjing University of Information Science and Technology, China, naeem.uoh@gmail.com

ABSTRACT

Highlight generation involves extracting the most engaging clips from a sports video. In the context of video summarization, the entire video is condensed into a shorter format, retaining the most critical content. For example, a complete match video in cricket includes various actions such as fours, sixes, and wickets. Highlights, on the other hand, still these significant events—fours, sixes, and wickets—into a cohesive and essential highlights package. We use recorded cricket videos and player images from publicly accessible online sources to conduct our study. Data pretreatment is done first, and then cleaned data is ready for deep learning model training. CNN and VGG-16 are the two deep-learning models that we trained. Following the creation of the models, standard assessment metrics are used to compare the two models. We trained the model for 70 epochs with SGD optimizer, and categorical cross-entropy loss function to optimize the model parameters. We trained two models for the task of cricket player recognition and highlight generation: a CNN model for frame extraction and a pre-trained VGG-16 model for feature extraction. The results suggest that both models are adequate for the task of cricket player recognition and highlight generation. The proposed model achieved 96% accuracy in the testing phase of the study. However, the VGG-16 model achieved higher accuracy than the CNN model, indicating that the pre-trained model is more effective at extracting relevant features from the frames. The ROC curves also suggest that both models have good discrimination ability, which is essential for generating accurate player-specific highlights.

KEYWORDS

Highlight Generation, Deep Learning, Data Analysis, Video Processing.

1. INTRODUCTION

Cricket is the national sport of England and now it is played around the world. According to a survey, cricket is the second most popular game after football in the world. A 2.5 billion strong fan base exists for cricket. The UK and various former British colonies, particularly India, Pakistan, and Australia, are where the game is most popular. It uses a bat, a big field, two teams, and scoring runs, just like baseball. At some point, we've all watched highlights of sporting events. Even if you aren't particularly interested in sports, you have probably seen highlights on television while dining out, relaxing in a hotel, etc. [1]. The process of selecting the most captivating segments from a sports video is known as highlight generation [2]. This can be viewed as a typical use of video summarization. In the video summary, the entire video is reduced to a manageable length while retaining the key points. The whole match video of a cricket game includes shots like fours, sixes, wickets, and so forth. Even uninteresting events like defences, leaves, wide balls, byes, etc., are captured in the unedited version. Highlights are the short videos that contain the main events of any sports video. The purpose of highlights is to quickly see the main stories of the game and save crucial time [2,3]. The classic highlights package in cricket combines all the key talking points, including fours, sixes, and wickets.

Artificial Intelligence significantly accelerates and simplifies the process, benefiting data scientists tasked with collecting, analyzing, and interpreting vast amounts of data [4,5]. Automatically collecting highlights from a whole match video saves both the developer and the user time [5-7]. So, this article will discuss how to generate automated highlight generation from cricket videos. The purpose of selecting Cricket is its popularity; as we know, it has a handsome fan base of 2.5 billion around the world [8]. Figure 1 illustrates the fan base of different popular sports in the world.

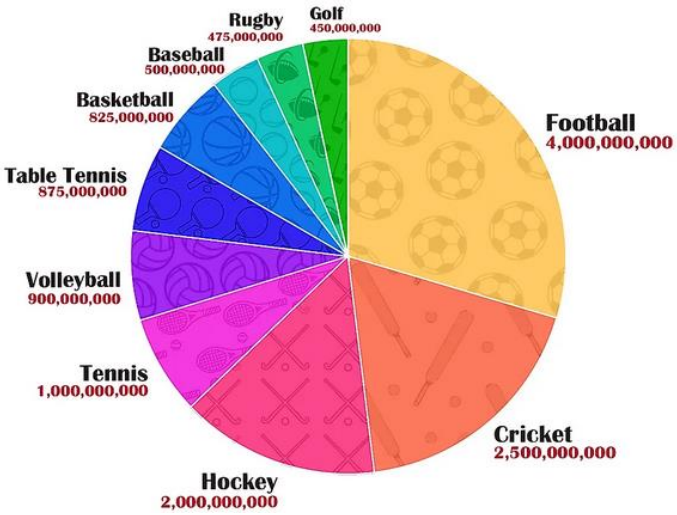

Figure 1: Most Famous Sports Around the World

Computer vision and Natural Language Processing (NLP) [9,24] are two popular methods that we can employ to solve various problems [10, 11]. This work makes several critical contributions to the field of automated sports highlight generation. First, it proposes a novel two-stage approach using CNN and VGG-16 models for player-specific analysis and highlight extraction. This allows for personalized, player-focused summarization. Second, it provides a comparative evaluation of different deep-learning architectures for this task. The findings demonstrate the superiority of transfer learning with VGG-16 over training a CNN from scratch. To end with, the work introduces a new automated methodology tailored to cricket, an understudied sport in video summarization research. The main contributions of the work are listed below.

- Gathering and cleaning video datasets for cricket Highlight generation.
- Design and implementation of a deep learning-based framework for player-specific cricket highlight generation.
- Design of hybrid technique using CNN for efficient feature extraction for cricket videos.
- Evaluation of proposed models based on various evaluation metrics, including accuracy, precision and recall, and F1-score.

The paper is organized into 5 Sections. Section 2 presents the related work for the proposed problem, while Section 3 discusses the proposed solution for highlight generation. Section 4 presents the results and analysis of proposed models, while Section 5 concludes the article. daunting, the simplest approach is to use this template and insert headings and text into it as appropriate.

2. RELATED WORK

The task of automatically generating highlights for sports videos has attracted growing research interest in recent years [12,13]. However, existing literature on cricket highlight generation remains relatively limited compared to more widely studied sports like soccer and basketball. While some progress has been made, most prior cricket summarization techniques rely predominantly on audiovisual cues and domain-specific heuristics. This section reviews relevant studies employing traditional sports video analysis and initial explorations of data-driven methods for cricket highlight extraction.

Gaikwad et al. [12] presented a method for producing highlights that pre-processes the video frames that were retrieved. The evaluation of these targeted frames is subsequently performed using convolutional neural networks. The suggested method extracts and computes the characteristics needed to produce a summary of videos. Cricket data was utilized to train deep neural networks. Experimental findings of this study demonstrate that the suggested technique outperforms other cutting-edge video summarization methodologies. The proposed method for summarizing videos is simple to apply and effective in capturing the critical moments from cricket matches. Tang et al. [8] suggest a novel method for identifying highlights from sports. Based on an unsupervised activity identification and detection approach, the videos are dynamically divided into several events. Cricket video categorization is used to show the efficiency of the suggested representation. Using activity statistics based on CH and HOG features, they gain a minimal error rate of 12.1%. Midhu et al. [13] suggest several algorithms to create highlights of cricket footage. The proposed procedure consists of two steps. Keyframe identification at level one is done using the hue histogram difference. Then, by identifying the absence of a scorecard, they categorize the frames as replay or actual frames. Then, based on the Dominant Grass Pixel ratio, actual frames are divided into field view and non-field view categories. The second section is then subjected to concept mining that uses the Apriori technique, using input from labelled frame events [14,15]. GAN uses a generator model to create images, outperforming standard GANs in prediction accuracy conditionally.

Shukla et al. [16] suggest a model that can automatically produce sports highlights. Cricket is a game with more intricate regulations and a longer season than many other games. In this research, they suggest a strategy to identify and clip significant occurrences in a cricket game that takes into account both event-based and excitement-based attributes. Cues used to record such events include replays, sound intensity, player excitement, and ground scenarios. Emon et al. [17] designed the deep cricket summarization network (DCSN). They developed a new dataset called CricSum because the few datasets that are already available in this field have certain limitations. The resulting summary's quality heavily depends on the viewers' perceptions. Therefore, they use the Mean Opinion Score (MOS) method to show the effectiveness of the suggested summarizing system. The summarizing videos that were produced automatically received a MOS score of 4 out of 5 in this study. Guntuboina et al. [18]. proposed an object detection method using YOLO, trained using a dataset of 1300 images. The scorecard was then cut out of the image for every frame of the video once it was found via YOLO [19]. The trimmed scoreboard was then processed to image processing to minimize distortion and false positives. To obtain the score, the final step involved running the processed image through an OCR (Optical Character Recognizer). The result of the OCR was passed through a rule-based technique to produce timestamps for significant game-based occurrences. In the experiments, an overall F1 Score of 97% was achieved.

A technique for identifying the critical events and summarizing them was proposed by [20]. The excitement snippets are first extracted using audio characteristics. Then, from each clip, the crucial moments, such as replay, athletes, umpires, fans, and player gathering, are taken out. In this case, an HDNN-EPO hybrid deep neural network is suggested for automatically identifying

the enthusiasm ideas shown in the cricket video based on the observed occurrences. Tejero-De-Prablos et al. [22] provided a method for summarizing user-generated sports footage and used the LSTM network to recognize different user activities in the video. Videos were categorized as engaging or uninteresting. The system's effectiveness was examined with a variety of attributes. Evaluation outcomes show proposed model performs better than the current summary methods.

3. PROPOSED SYSTEM

To improve the analysis of cricket videos we present a novel technique using Dl models in this research. We gather the cricket video data and clean it for better processing and training of the Dl models. First of all, data preprocessing is performed then cleaned data is prepared for the training of deep learning models. We trained two deep learning models, i.e., CNN and VGG-16. After model building, both models are compared to standard evaluation measures. The proposed system is illustrated in Figure 2.

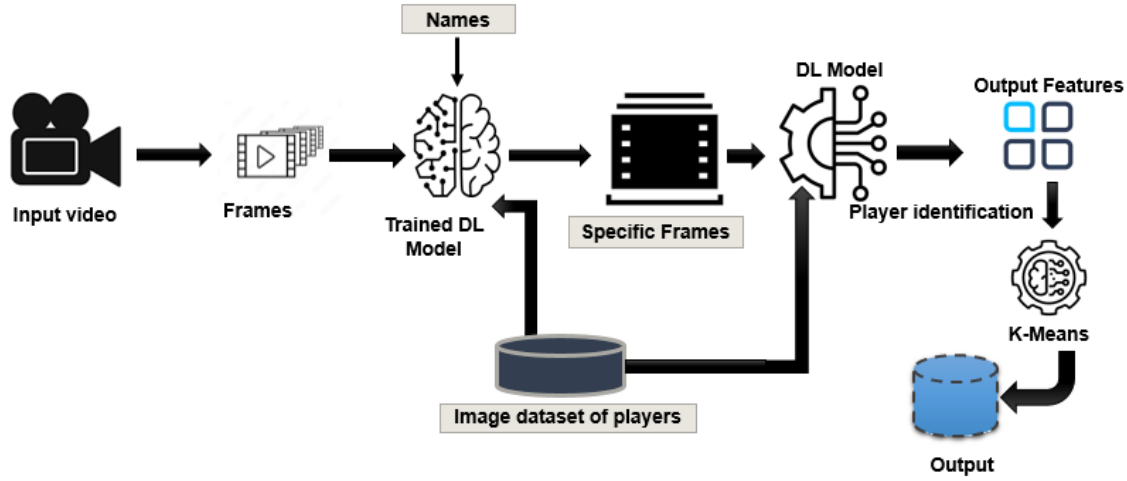

Figure 2: Process of proposed work

The developed CNN model is designed to recognize specific cricket players in a video and generate player-specific highlights. The model is trained on images of cricket players and takes input videos. When a particular player appears in the video, the model extracts frames and passes them to another model, VGG-16, which focuses on recognizing the player. Once the player is identified, the frames that belong to that player are extracted, and feature extraction is performed. Finally, the K-Nearest Neighbors (KNN) algorithm is used to summarize the video and generate player-specific highlights.

3.1. Input: Cricket Match Videos

We collected images of different cricket players from various online sources, including social media, news websites, and official team websites. We tried to collect images of players from other countries, teams, and playing positions to ensure that the model can recognize players from diverse backgrounds. We selected a variety of videos that represent different types of matches, including test matches, one-day internationals, and T20 matches. We also chose videos from various tournaments and competitions, including international matches and domestic leagues.

2.2. Data Preprocessing and Preparation

Since the videos were in different formats and resolutions, we standardized the videos by converting them to a standard format and resolution. We used the FFmpeg tool to convert the videos to the MP4 format and the 720p resolution. This standardization ensured video compatibility and consistent quality across different sets, followed by dividing videos into training, validation, and test sets.

2.3. Player Detection

Edge detection is a commonly used technique in computer vision that can help to enhance the features of an image and reduce the amount of noise in the data. We used a canny edge detection technique in our system. The Canny edge detector is a commonly used algorithm that can help to enhance the features of an image and reduce noise in the data. The example of edge detection is shown in Figure 3.

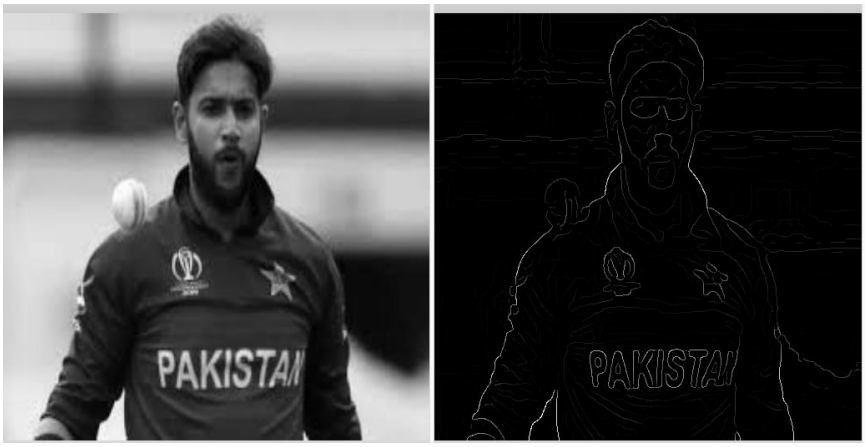

Figure 3: An example of Edge detection

After applying the Canny edge detector, the frames are fed into the deep-learning models for player detection and recognition. The enhanced edges could help the models to identify better and classify the players, leading to more accurate and effective highlight generation.

2.4. Extracting Relevant Frames

We used CNN model for the extraction of relevant features. The stochastic gradient descent (SGD) optimizer with a learning rate of 0.001 and a momentum of 0.9 was used for this model. We trained the model for 70 epochs, with a categorical cross-entropy loss function to optimize the model parameters. We fine-tuned the trained model using the validation dataset to improve its performance. We tune the hyperparameters of the model, such as the learning rate, the number of epochs, and the batch size, to improve the model's accuracy. We evaluated the performance of the trained model on a separate validation dataset. We extracted the frames from the input video at a fixed frame rate using OpenCV. The extracted frames were then fed into the player detection and recognition model to detect and recognize the players in the video frames.

2.5. Specific Player Identification and Highlight Generation

The second model is a pre-trained VGG-16 model that focuses on recognizing the specific cricket player whose frames were extracted in the previous step. The proposed VGG-16 architecture consists of 13 convolutional layers followed by three fully connected layers. A ReLU activation function and a max pooling layer follow each convolutional layer. The last three fully connected layers are used for classification, with the final layer providing the class probabilities. When the extracted frames for a specific player are passed to the VGG-16 model, the model analyzes the frames and generates a feature vector that represents the unique

characteristics of the player in each frame. Once the feature vectors are generated for all the frames belonging to a specific player, the next step is to use a machine learning algorithm, that is, KNN, to summarize the video and generate player-specific highlights. The KNN algorithm compares the feature vectors of each frame with those of the neighbouring frames and identifies the most similar frames. These similar frames are then grouped to create a highlight for the specific player.

4. Results and Analysis

We evaluated the performance of both models on a test set of cricket videos. The size of the videos ranges from 10 to 25 minutes, while others are 2.5 hours long. During our simulations, we applied both optimistic and realistic parameters in the context of video summarization. Figure 4 presents the training dataset employed in this study. These are some videos that are part of the dataset that we utilized to train and test our proposed models.

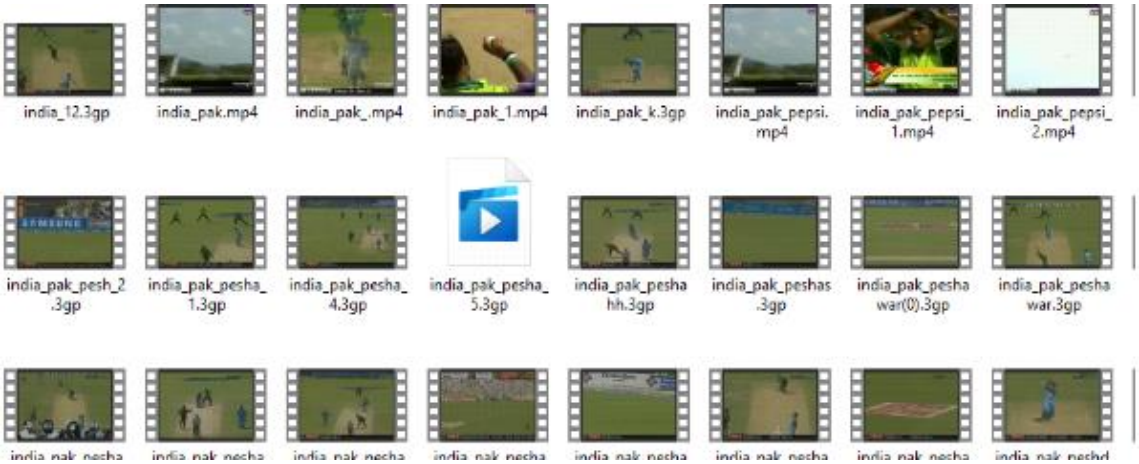

Figure 4: Training data set

Individual frames were extracted from the videos, as displayed in Figure 5, converting the unstructured video data to a structured format for analysis. This dataset and frame-based representation provided the necessary input for developing machine learning models within a supervised learning paradigm, enabling the algorithms to learn real-world object-tracking challenges.

Figure 5: Extracted frames of a video from the dataset

The most crucial classification metric in a machine learning task is accuracy. It is well-suited for both binary and multiclass classification problems. The most widely used statistic for evaluating a model, however, it is not a reliable indicator of its performance. The formula for accuracy is given below:

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \quad (1)$$

The training and validation accuracy of the CNN model is illustrated in Figure 6.

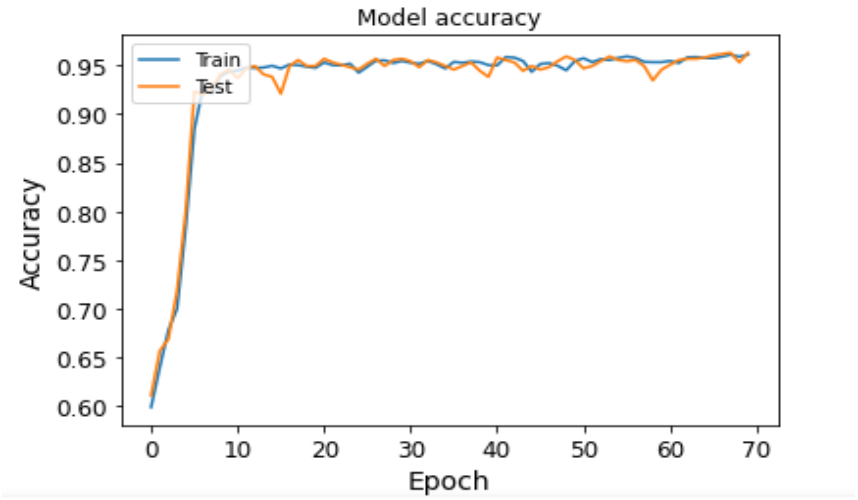

Figure 6: Training and validation accuracy of the proposed CNN model

As shown in Figure 6, the proposed model achieved 95.6% accuracy for the extraction of specific frames to detect the players so that these frames can be used for video summarization. The curves show that the model achieved high accuracy on both the training and validation sets without overfitting. The Training and validation loss of the proposed CNN model is shown in Figure 7.

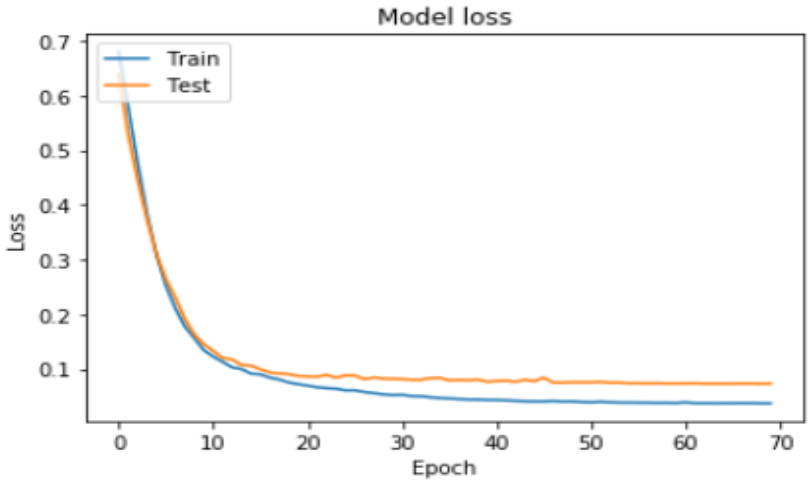

Figure 7: Training and validation loss of the proposed CNN model

Figure 7 shows that the CNN model was able to converge during training and achieve a low loss value rapidly. The consistently decreasing loss on both the training and validation sets again

provides evidence that the model successfully learned meaningful representations for detecting players without overfitting. The training and validation accuracy of the proposed VGG-16 model for feature extraction of specific frames is illustrated in Figure 8.

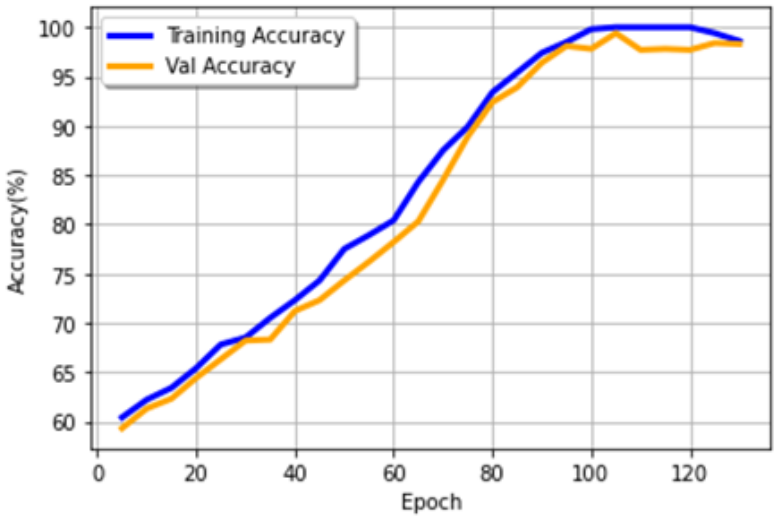

Figure 8: Training and validation accuracy of the proposed VGG-16 model

On the test set, the VGG-16 model achieved 97.3% accuracy with a 0.1 loss. Figure 9 depicts the training and validation accuracy curves for the VGG-16 model. The VGG-16 model's loss curve further indicates that the model converged rapidly and attained a low loss.

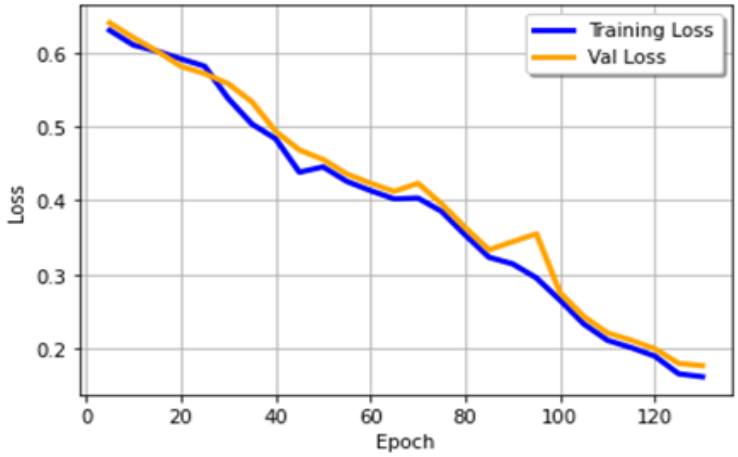

Figure 10: Training and validation loss of the proposed VGG-16 model

To combat overfitting, the models were trained on sizable datasets of player images and cricket match videos. The ROC curve of the proposed CNN model is presented in Figure 10.

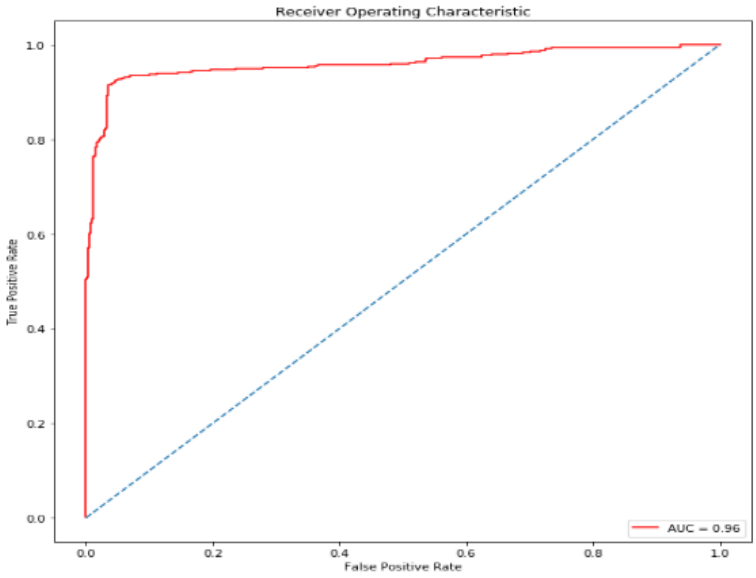

Figure 10: ROC Curve proposed CNN model

We also evaluated the performance of both models using ROC curves. The ROC curves for both models show that they can achieve high TPR values while maintaining low FPR values, which is a desirable characteristic for our task. The AUC score for the VGG-16 model is slightly higher than that of the CNN model, indicating that the pre-trained model is more effective at distinguishing between frames that belong to the specific player and frames that do not. The ROC curve of the proposed VGG-16 model is illustrated in Figure 11.

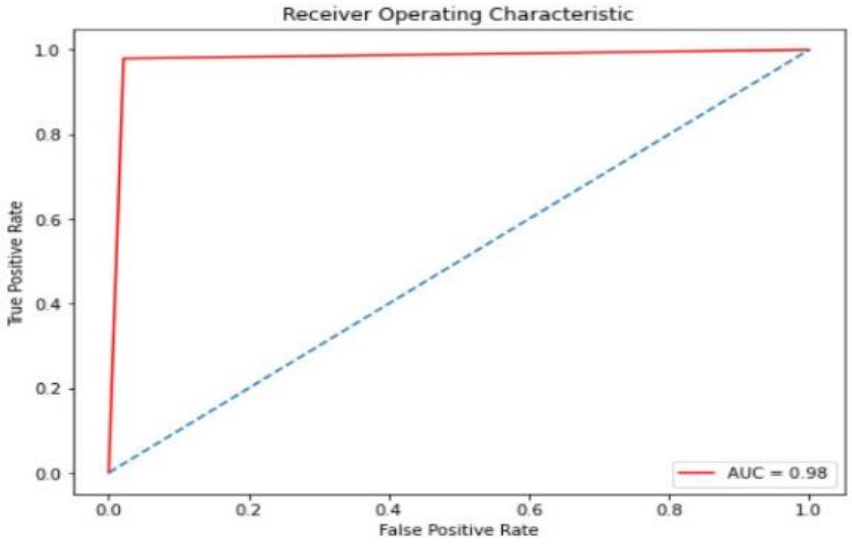

Figure 11: ROC Curve proposed VGG-16 model

The KNN model achieved an accuracy of 94.44%, indicating that it was able to identify player-specific highlights in the video with high accuracy correctly. The precision of 92.13% suggests that out of all the highlights identified by the model, 92.13% were genuinely relevant to the specific player being analyzed. The recall value of 93.16% indicates that out of all the relevant

player-specific highlights in the video, the model was able to identify 93.16% of them correctly. Finally, the F1 score of 90.63% provides a balanced metric that takes into account both precision and recall.

5. COMPARISON WITH OTHER STUDIES

Khan et al. [31] present a novel framework for sports video highlight generation using deep learning. They introduced SPNet, a deep learning-based system that generates highlights and can identify thrilling sporting events by utilizing sequences of high-level visual attributes. For precise activity recognition, the proposed model makes use of the strengths of Inception blocks and 3D convolution networks. They separate the thrill of sports videos into scenarios, viewpoints, and actions. Their suggested approach achieved 76% accuracy on the SP-2 dataset and 82% accuracy on the C-sports dataset. Shingrakhia et al. [32] proposed a Stacked Gated Recurrent Neural Network with Attention Module (SGRNN-AM), a hybrid deep learning model for summarizing cricket videos. Key occurrences from the cricket video are identified by analyzing the excitement, object, and event-based aspects. Their model achieved 96% accuracy in their proposed dataset.

The proposed model outperforms these previous models by achieving 96.56% accuracy in the evaluation phase of the study. The evaluations indicate that the proposed models can be used to summarize long-form cricket videos. The proposed VGG-16 model achieved higher accuracy than the CNN model, indicating that the pre-trained model is more effective at extracting relevant features from the frames. The ROC curves also suggest that both models have good discrimination ability, which is essential for generating accurate player-specific highlights. The summary and comparison with previous models are discussed in Table 2.

Table 3: Comparison with previous studies

Author	Methodology	Key Findings	Accuracy
Karpathy et al. [27]	CNN for video classification	CNN-based method lacked temporal analysis, resulting in lower accuracy.	65.4%
Banjar et al. [29]	An automated framework based on enthusiasm detection	Used audio stream analysis for key moment identification;	97%
Narwal et al. [30]	Multimodal Technique	Automated framework with an acoustic feature descriptor;	87.61%
Khan et al. [31]	SPNet	Utilized Inception blocks and 3D convolution networks; separated thrill into scenarios, viewpoints, and actions.	76% on the SP-2 dataset, 82% on the C-sports dataset
Shingrakhia et al. [32]	Stacked Gated Recurrent Neural Network with Attention Module (SGRNN-AM)	Identified key occurrences using excitement, object, and event-based analysis.	96%
Proposed Model	VGG-16 model for player-specific highlight generation	Focus on player-specific analysis and achieve higher accuracy with better feature extraction compared to CNN models.	**96.56%**

6. CONCLUSIONS

In this paper, a video summarization method using a convolution neural network has been studied. For user consumption, a summary method creates an abstract form of its inputs. Due to the abundance of social media data, it is essential to delve into the text to uncover knowledge that can be applied to a wide range of people. The goal of this study is to close the gap between the quantity of data that is produced and the quantity that can be effectively examined manually. This research presents a novel technique for extracting highlights from cricket films. The process involves two steps. Each cover of the video is used to generate a different event in the first phase. Highlights are created through the proper association of the occurrences using deep learning techniques. All of these stages are tested experimentally, and the outcomes are confirmed. Future upgrades to the technology could include audio commentary detection and umpire gesture recognition.

REFERENCES

[1] Kolekar, M.H. and S. Sengupta. Event-importance-based customized and automatic cricket highlight generation. in 2006 IEEE International Conference on Multimedia and Expo. 2006. IEEE.

[2] Kolekar, M.H. and S. Sengupta, Bayesian network-based customized highlight generation for broadcast soccer videos. IEEE Transactions on Broadcasting, 2015. 61(2): p. 195-209.

[3] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. nature, 2015. 521(7553): p. 436-444.

[4] Brooks, C. and C. Thompson, Predictive modelling in teaching and learning. Handbook of learning analytics, 2017: p. 61-68.

[5] Khan et al. "SPNet: A deep network for broadcast sports video highlight generation." Computers and Electrical Engineering 99 (2022): 107779.

[6] Fang et al., "Computer vision for behaviour-based safety in construction: A review and future directions." Advanced Engineering Informatics 43 (2020): 100980.

[7] Dange, B., et al. Automatic Video Summarization for Cricket Match Highlights using Convolutional Neural Network. in 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN). 2022. IEEE.

[8] Tyagi, S., et al. Enhanced predictive modeling of cricket game duration using multiple machine learning algorithms. in 2020 international conference on data science and engineering (ICDSE). 2020. IEEE.

[9] Akbik, A., et al. FLAIR: An easy-to-use framework for state-of-the-art NLP. in Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations). 2019.

[10] Shingrakhia, H. and H. Patel, Cricket Video Highlight Generation Methods: A Review. ELCVIA Electronic Letters on Computer Vision and Image Analysis, 2022. 21(2): p. 1-22.

[11] Shingrakhia, H. and H. Patel, Emperor penguin optimized event recognition and summarization for cricket highlight generation. Multimedia Systems, 2020. 26(6): p. 745-759.

[12] Gaikwad, D., S. Sarap, and D. Dhande, Video Summarization Using Deep Learning for Cricket Highlights Generation. Journal of Scientific Research, 2022. 14(2): p. 533-544.

[13] Midhu, K. and N. Anantha Padmanabhan, Highlight generation of cricket match using deep learning, in Computational Vision and Bio Inspired Computing. 2018, Springer. p. 925-936.

[14] Ramesh, P. N., & Kannimuthu, S. Context-Aware Practice Problem Recommendation Using Learners' Skill Level Navigation Patterns. Intelligent Automation & Soft Computing 2023, 35(3).

[15] Raja, K. C., & Kannimuthu, S. Conditional Generative Adversarial Network Approach for Autism Prediction. Computer Systems Science & Engineering 2023, 44(1).

[16] Shukla, P., et al. Automatic cricket highlight generation using event-driven and excitement-based features. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018.

[17] Emon, S.H., et al. Automatic Video Summarization from Cricket Videos Using Deep Learning. in 2020 23rd International Conference on Computer and Information Technology (ICCIT). 2020. IEEE.

[18] Guntuboina, C., et al., Deep learning based automated sports video summarization using YOLO. ELCVIA Electronic Letters on Computer Vision and Image Analysis, 2021. 20(1): p. 99-116.

[19] Yan, C., X. Li, and G. Li. A new action recognition framework for video highlights summarization in sporting events. in 2021 16th International Conference on Computer Science & Education (ICCSE). 2021. IEEE.

[20] Memon, J., et al., Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR). IEEE Access, 2020. 8: p. 142642-142668.

[21] Guntuboina, C., et al. Video Summarization for Multiple Sports Using Deep Learning. in Proceedings of the International e-Conference on Intelligent Systems and Signal Processing. 2022. Springer.

[22] Tejero-de-Pablos, A., et al., Summarization of user-generated sports video by using deep action recognition features. IEEE Transactions on Multimedia, 2018. 20(8): p. 2000-2011.

[23] Shih, H.-C., A survey of content-aware video analysis for sports. IEEE Transactions on circuits and systems for video technology, 2017. 28(5): p. 1212-1231.

[24] Karmaker, D., et al. Cricket shot classification using motion vector. in 2015 Second International Conference on Computing Technology and Information Management (ICCTIM). 2015. IEEE.

[25] Devi, V. S., & Kannimuthu, S. Author profiling in code-mixed WhatsApp messages using stacked convolution networks and contextualized embedding based text augmentation. Neural Processing Letters, 55(1) 2023, 589-614.

[26] Jesabhai,. "Automatic Cricket Highlight Generation Using Event-Driven and Excitement Based Features Using Deep Learning." PhD diss., AD Patel Institute of Technology, 2022.

[27] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 1725-1732).

[28] Ramanathan, V., Huang, J., Abu-El-Haija, S., Gorban, A., Murphy, K., & Fei-Fei, L. (2016). Detecting events and key actors in multi-person videos. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3043-3053).

[29] Banjar, A., Dawood, H., Javed, A., & Zeb, B. (2024). Sports video summarization using acoustic symmetric ternary codes and SVM. Applied Acoustics, 216, 109795.

[30] Narwal, P., Duhan, N., & Bhatia, K. K. (2023). A novel multi-modal neural network approach for dynamic and generic sports video summarization. Engineering Applications of Artificial Intelligence, 126, 106964.

[31] Khan, A. A., & Shao, J. (2022). SPNet: A deep network for broadcast sports video highlight generation. Computers and Electrical Engineering, 99, 107779.

[32] Shingrakhia, H., & Patel, H. (2022). SGRNN-AM and HRF-DBN: a hybrid machine learning model for cricket video summarization. The Visual Computer, 38(7), 2285-2301.