Video Based Cricket Shot Classification using modified LRCN Approach

Sriraam C
School of Computer Science andEngineering
Vellore Institute of Technology
Chennai, TamilNadu
sriraam.c2021@vitstudent.ac.in

Krithin Ragav
School of Computer Science andEngineering
Vellore Institute of Technology
Chennai, TamilNadu
krithin.r2021@vitstudent.ac.in

Lionel Donato L
School of Computer Science andEngineering
Vellore Institute of Technology
Chennai, TamilNadu
lioneldonato.l@vitstudent.ac.in

Abstract— **In recent years, advancements in Artificial Intelligence (AI) have catalyzed profound transformations in the realm of cricket analysis. From player performance evaluation to strategic decision-making, AI technologies have revolutionized how cricket is understood and played. In this project, we focus on the task of classifying cricket shots using video data, leveraging deep learning methodologies. Our model aims to classify cricket videos into five distinct shot categories, utilizing the CricShot10 dataset originally comprising ten shot classes. With approximately 100-150 video clips per shot type, we streamline the classification task by selecting five fundamental shots for analysis. To accomplish this, we propose an innovative architecture combining Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks, commonly referred to as LRCN. Additionally, we integrate techniques such as MaxPooling2D, Dropout, Flatten, LSTM, and Dense layers to enhance the model's performance and robustness. Through this project, we not only contribute to the field of cricket analysis but also demonstrate the efficacy of deep learning approaches in video-based sports analytics.**

Keywords— AI developments, cricket analysis, deep learning,classification, video data, CNNs, LSTM.

I. Introduction

In the ever-expanding landscape of data communication, the surging tide of video content on platforms like YouTube and Vimeo presents both an ocean of opportunities and formidable challenges. Within this deluge, the analysis of video data emerges as a pivotal task, particularly within the realm of sports. Cricket, with its sprawling global audience, stands as a quintessential example, epitomized by monumental events like the ICC Men's Cricket World Cup 2019, which captivated an average global viewership exceeding 1.6 billion. This colossal viewership underscores the commercial significance of scrutinizing cricket data, offering broadcasters invaluable insights into viewer preferences and behaviour, thus highlighting the pivotal role of data analysis in the sports broadcasting arena.

However, despite cricket's colossal popularity, there remains a significant gap in leveraging data analytics from a commercial perspective. Yet, within this untapped landscapelies immense potential, particularly in developing unbiased, sensor-based commentary systems and shot detection algorithms. Recognizing the transformative impact of such endeavours, we focused our attention on cricket as our foundational sport, with a keen emphasis on analysing different batting shots. These shots, ranging from classic cover drives to deftly executed sweeps, present a unique challenge due to the inherent similarities in video frames across different shot types.

Our contributions to this field include proposing the use of a deep learning methodology to classify different cricket shots. This approach involves combining Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks in a CNN-LSTM model, also known as the LRCN model. This architecture was first proposed in the paper titled "Long-term Recurrent Convolutional Networks for Visual Recognition and Description" by Donahue et al [17]. We explored various convolutional techniques and transfer learning models to optimize the classification process. Our proposal uses the existing CricShot10 dataset [16], a comprehensive repository that provides an unparalleled variety of cricket shots, as a foundation for our approach.

Moving forward, the subsequent sections of this paper serve to elucidate our methodology, beginning with an in-depth exploration of related work (Section II). We then pivot to a detailed exposition of our proposed CNN-LSTM architecture for cricket shot categorization (Section III), followed by a comprehensive analysis of experimental results and discussions on dataset and model evaluation (Sections IV, Vand VI). Finally, we conclude by charting a course for future research endeavours (Section VII), underscoring the vast potential that lies ahead in the intersection of data analytics and cricket.

II. Related Works

Several studies have investigated various approaches to enhancing accuracy of cricket shot classification systems. In this study, we delve into a comprehensive analysis of methodologies and associated drawbacks across various research papers in the domains of computer vision, machine learning, and digital image processing. In the domain of gesture recognition for sports, recent advancements have been made with the integration of deep learning techniques. Notably, A. Semwal investigated cricket bat stroke classification using a combination of saliency, optical flow analysis, and Deep Convolutional Neural Networks (DCNN) [1]. While their methodology achieved high accuracy in categorizing right-handed shots, it encountered challenges in generalizing effectively to left-handed shots, potentially due to biases or limitations in the diversity of training data.Shifting focus to healthcare applications, Abedi delved into vision-based posture recognition using convolutional neural networks (CNNs) [2]. Despite offering a non-invasive approach without wearable sensors, their methodology faced difficulties in accurately distinguishing subtle posture variations and operating robustly in complex environments, presenting significant hurdles for practical implementation. In the field of scene image analysis, [P. Promvijittrakarn] present an innovative approach utilizing eye tracking and saliency maps [3]. Despite its novelty, the methodology encounters challenges in accurately interpreting gaze signalsand ensuring broad applicability across diverse scene images, suggesting areas for further refinement and validation. Shifting focus to machine learning in soccer image analysis, a recent study combines deep learning and histogram analysis for player detection and team classification [4]. While demonstrating promising outcomes, the methodology's reliance on single frames and sensitivity to scene complexity pose significant challenges that necessitate in-depth investigation and optimization for robust performance. Additionally, a comprehensive survey of optical flow estimation techniques sheds light on this critical aspect of computer vision [5]. However, the lack of practical implementation insights and empirical validations may restrict the real-world applicability of these techniques, underscoring the importance of rigorous experimentation and comparative analysis to evaluate their effectiveness across diverse scenarios.

In the domain of human action recognition, Thomas presents a study using MediaPipe Holistic, introducing innovative landmark extraction techniques [6]. However, potential limitations in accuracy are noted due to the extraction methods employed and the choice of classification algorithms, indicating areas for further refinement and investigation. Furthermore, discussions on digital image processing in real-time systems [7] and computer vision applications in sports analysis [8] underscore crucial challenges and opportunities in these areas, emphasizing the need for more detailed methodologies and empirical evaluations to bolster their contributions. Additionally, a literature analysis on machine learning styles in computer vision [9] and data screening for quality subjective assessment [10] provide valuable insights, yet their effectiveness could be strengthened with comprehensive practical implementations and assessments of proposed methodologies. Lastly, a study on shot segmentation in sportsvideo analysis [11] proposes a method for detecting cricket video shot boundaries but could benefit from empirical validations and comparisons with existing approaches to enhance its robustness and effectiveness across diverse cricket match scenarios.

III. PROPOSED METHODOLOGY

The proposed approach for video-based cricket shot classification begins with the importation of necessary modules to facilitate data processing and model training. Libraries such as OpenCV for video manipulation and scikit- learn for dataset splitting establish a framework for our classification task. The CricShot10 dataset contains videos showcasing various cricket shots, which undergo processing to extract frames. This is a crucial step in preparing the data for input into our classification model.

Central to this process is the frames_extraction() function. It iterates through each video, reading frames at regular intervals, resizing them to a predetermined height and width,and normalizing pixel values to a range between 0 and 1. Standardizing the dimensions and intensity levels of the frames ensures uniformity in the input data, essential for effective model training. Additionally, the function focuses on the centre of the 1280x720 pixel video frames to avoid unnecessary noise. By capturing only the central portion of the frame, this approach helps reduce the influence of peripheral elements, streamlining the dataset and enhancing the relevance of the input data.

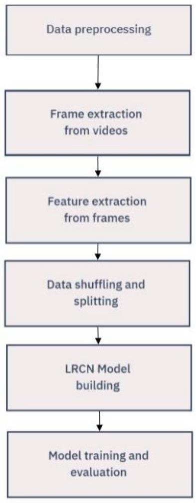

Figure 1: Diagram demonstrating the flow of activities

Once frames are extracted and pre-processed, shuffling of the features and labels is performed to introduce randomness into the dataset. This step helps mitigate any inherent biases in the original ordering of the data, enabling the model to learn more robustly. Additionally, one-hot encoding is employed to transform categorical labels into a format suitable for classification. Each class is represented as a binary vector, with a value of 1 at the index corresponding to the class and 0s elsewhere.

Having pre-processed and encoded the dataset, the next step is to split it into training and testing sets using the train_test_split() function. The training set, comprising 80% of the data, serves as the primary source for model training, while the testing set, encompassing the remaining 20%, is reserved for evaluating model performance on unseen data. Importantly, data shuffling occurs during splitting to ensure that both the training and testing sets contain a representativedistribution of cricket shots, thereby avoiding biases arising from fixed ordering.

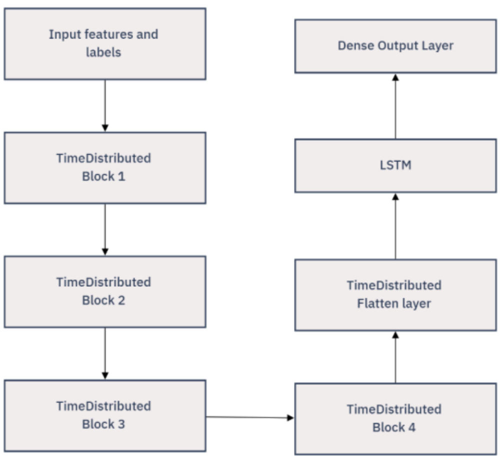

Figure 2: The architecture of the proposed model

The architecture of the Long Short-Term Memory Convolutional Neural Network (LRCN) model is defined, and training and evaluation processes are initiated. The model comprises a sequence of convolutional layers, each wrapped in a TimeDistributed layer to process video frames over time, followed by max-pooling and dropout layers for regularization as displayed by Figure 3. A bidirectional LSTM layer is incorporated to capture temporal dependencies, and a softmax-activated dense layer serves as the output layer for multi-class classification. Regularization is enforced using L2regularization throughout the model.

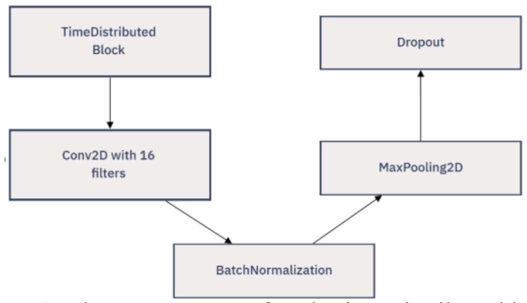

Figure 3: The components of each TimeDistributed layer

For training, the model is compiled with the categorical cross-entropy loss function, the Adam optimizer, and accuracy as the evaluation metric. Additionally, early stopping is implemented with a patience of 15 epochs to prevent overfitting and restore the best weights when the validation loss stops decreasing. Training commences with adefined number of epochs (in this case, 70) and a batch size of
4. Data shuffling is applied during training to enhance model generalization, and a validation split of 20% is employed to monitor model performance during training.

Following training, the model's performance is evaluated on the test set using the evaluate() method. The evaluation metrics, including loss and accuracy, are computed, providing insights into the model's generalization capability on unseen data. This comprehensive training and evaluation pipeline ensures that the LRCN model is effectively trained and rigorously assessed, allowing for reliable assessments of its performance in classifying cricket shot videos.

IV. EXPERIMENTAL RESULTS AND OBSERVATIONS

In this research study, our Deep learning model achieved commendable performance metrics across multiple evaluation metrics. Specifically, the precision of the model was recorded at 0.90, indicating a high proportion of correctlypredicted positive cases out of all cases predicted as positive.

This signifies the model's ability to minimize false positives and make accurate positive predictions. Moreover, the model demonstrated an accuracy of 0.90, reflecting the overall correctness of predictions across both positive and negative cases. This high accuracy underscores the effectiveness of the model in making correct classifications. Additionally, the recall of the model was measured at 0.93, representing the proportion of actual positive cases that were correctly identified by the model. A high recall value indicates that the model effectively captures most of the positive cases, minimizing false negatives. Furthermore, the F1-score, which combines precision and recall into a single metric, was calculated at 0.92. This F1-score indicates a robust balance between precision and recall, highlighting the model's overallstrong performance in terms of both correctly identifying positive cases and minimizing false classifications. These performance metrics collectively demonstrate the efficacy and reliability of our machine learning model in the task under consideration, showcasing its potential for practical applications in real-world scenarios.

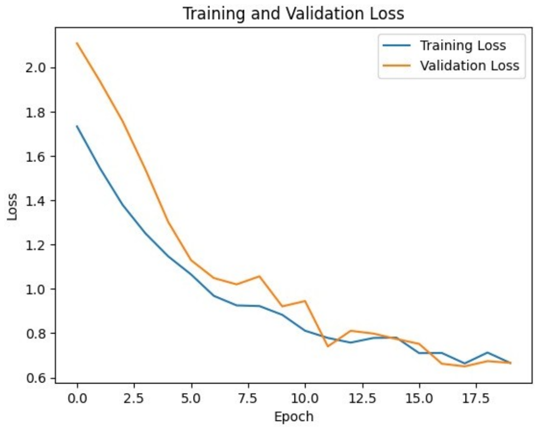

Figure 4: Training and Validation Loss

Figure 4 shows the training and validation loss over epochs, indicating the performance of our model as it learns from data. The trend of both curves decreasing over time suggests that the model is improving its ability to predict outcomes on both the training and unseen validation data. This is a positivesign, as it shows that the model is learning patterns from the data effectively. The rate at which the validation loss decreases is slower compared to the training loss, which is typical because the model is learning from data it has not seen before in the validation set. The continuous decrease in validation loss is encouraging, as it suggests the model is generalizing well. Overfitting would be indicated by a significant increase in validation loss, which is not evident in this graph.

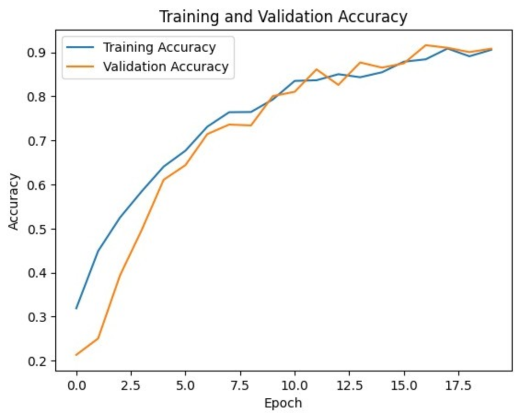

Figure 5: Training and Validation Accuracy

The Figure 5 is a graph depicting validation accuracy and training accuracy over epochs shows a consistent increase in both metrics as training progresses. This trend suggests that the model is successfully learning from the data and improving its ability to correctly classify inputs. The rising validation accuracy indicates an improvement in the model'sperformance on the unseen validation set, which is a positive sign of generalization. The steady increase in accuracy acrossboth training and validation data sets implies effective learning and adaptability of the model. Monitoring these trends is essential for ensuring the model's performance remains robust. Future research may focus on enhancing thisperformance further through techniques such as data augmentation or exploring more diverse training datasets.

V. RESULTS AND DISCUSSIONS

The model underwent training for 20 epochs, processing a total of 509 batches. Despite the batch size not being explicitly specified, the training duration of 1405 seconds implies an approximate batch size of 4. Each epoch exhibitedvariable durations, ranging from 53 to 82 seconds, indicating potential fluctuations in computational load or dataset complexity across epochs. These variations may reflect dynamic adjustments in learning dynamics or fluctuations in resource utilization during the training process. Overall, the training procedure demonstrated efficiency, culminating in the model's convergence to a satisfactory performance level across the training epochs.

The model training was conducted on an HP Pavilion Gaming 15-ec2000 Series laptop featuring a dedicated NVIDIA GeForce GTX 1650 GPU. Leveraging the computational power of this hardware configuration, the training process navigated through the complexities of deep learning tasks with relative ease. The GPU's parallel processing capabilities facilitated rapid execution of intensive numerical computations, significantly enhancing the training performance.

Parameters	**Values**
Epochs	20
Total Batches	509
Batch size	4
Training time/ epoch	120s
Total	2400 s

Table 1: Tabulates the training results of the experiment

VI. PERFORMANCE EVALUATION

In the study outlined in the paper "Cricket Shot Detection from Videos,"[1], the classification of shots performed by right-handed players yielded an accuracy of 83.098%. This achievement was made possible through the utilization of advanced methodologies such as saliency and optical flow analysis, which effectively capture both static and dynamic cues present in the video data. Moreover, the integration of Deep Convolutional Neural Networks (DCNN) played a pivotal role in extracting intricate representations from the visual data, contributing significantly to the accuracy achieved.

Another paper that we add to our comparison is "Cricshotclassify: an approach to classifying batting shots from cricket videos using a convolutional neural network and gated recurrent unit" published in 2018 [9]. This paper presents a hybrid deep neural network architecture for classifying 10 different cricket batting shots. The authors experimented with various models, including conventional CNN, dilated CNN, and transfer-learning models such as VGG16, InceptionV3, Xception, and DenseNet169. The VGG16–GRU model achieved 86% accuracy. Further refining the VGG16 model by freezing all but the final 4 andfinal 8 layers improved the performance, achieving 93% accuracy on the CricShot10 dataset.

In contrast, our proposed model, based on the Long Short-Term Memory (LSTM) Recurrent Convolutional Network (LRCN) architecture, surpasses these results with a remarkable accuracy of 96% across five distinct classes of shots executed by right-handed players. This enhanced performance underscores the efficacy of the LRCN architecture in discerning subtle temporal patterns inherent in cricket shots. By leveraging the temporal information encoded in the sequential data, our model demonstrates superior discriminative capabilities, thereby offering a robustsolution for shot classification tasks in cricket videos.

Experiment	**Methodology**	**Accuracy**
Cricket Shot Detection fromVideos [1]	DCNN	83.098%
Cricshotclassify [9]	CNN + LSTM	93%
Proposed Model	LRCN	96%

Table 2: Tabulating the performance of the proposed method and previous works

VII. FUTURE WORKS

In the burgeoning field of cricket data analytics, numerous avenues await further exploration and innovation. One promising area for future research involves fine-grained shot classification, delving deeper into the nuances of shot execution to provide more granular insights into player technique and performance. Additionally, there is a growing interest in developing real-time shot detection algorithms that could revolutionize live cricket broadcasting and analysis, enabling instant insights into player actions and facilitating on-the-fly adjustments by coaches and commentators. As part of this future work, integrating advanced object detection models such as the YOLO (You Only Look Once) Model could significantly enhance shot detection accuracy and speed, paving the way for more sophisticated and efficient cricket analytics systems.

Another fertile ground for exploration lies in player performance prediction, where researchers can delve into predictive modeling techniques to forecast individual and team performance based on historical data and contextual factors. Moreover, the integration of data from multiple sources, such as video footage, player biometrics, and match statistics, presents an opportunity for richer insights intoplayer behavior and game dynamics through multimodal datafusion.

Furthermore, the development of automated commentary systems that generate dynamic and contextually relevant commentary tailored to individual viewers represents an exciting frontier in cricket analytics. Leveraging natural language processing techniques and real-time data analysis, these systems have the potential to enhance the viewing experience for cricket fans worldwide. Additionally, advances in augmented reality and virtual reality offer opportunities for interactive fan engagement, enabling immersive experiences such as virtual stadium tours and interactive game simulations.

However, amidst these advancements, it is crucial to address ethical and legal considerations surrounding data privacy, consent, and algorithmic bias. Future research should prioritize the development of frameworks and guidelines to ensure the responsible and ethical use of cricket data analytics. By exploring these and other avenues of research, the field of cricket data analytics can continue to evolve, offering new insights, enhancing fan experiences, and drivinginnovation in the sport.

References

1. A. Semwal, D. Mishra, V. Raj, J. Sharma and A. Mittal, "Cricket Shot Detection from Videos," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India,2018, pp. 1-6, doi: 10.1109/ICCCNT.2018.8494081

2. Abedi, Wafaa M. Salih, D. Ibraheem-Nadher, and Ahmed T. Sadiq. "Modified deep learning method for bodypostures recognition." International Journal of Advanced Science and Technology 29.2 (2020): 3830-3841.

3. P. Promvijittrakarn and T. Charoenpong, "Soccer Player Detection and Soccer Team Classification using GoogLeNet and Histogram Analysis," 2023 15th Biomedical Engineering International Conference (BMEiCON), Tokyo, Japan, 2023, pp. 1-4, doi: 10.1109/BMEiCON60347.2023.10321826.

4. S. S. Beauchemin and J. L. Barron. 1995. The computation of optical flow. ACM Comput. Surv. 27, 3 (Sept. 1995), 433–466. https://doi.org/10.1145/212094.212141.

5. Singh, Amritanshu Kumar, Vedant Arvind Kumbhare, and K. Arthi. "Real-time human pose detection and recognition using mediapipe." International Conference on Soft Computing and Signal Processing. Singapore: Springer Nature Singapore, 2021.

6. Thomas, Graham, et al. "Computer vision for sports: Current applications and research topics." Computer Vision and Image Understanding 159 (2017): 3-18.

7. S. V. Mahadevkar et al., "A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions," in IEEE Access, vol. 10, pp. 107293-107329, 2022, doi: 10.1109/ACCESS.2022.3209825.

8. K. R. Raval and M. M. Goyani, "Shot Segmentation and Replay Detection for Cricket Video Summarization," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 933-938, doi:10.1109/CISES58720.2023.10183588.

9. Sen, Anik, et al. "Cricshotclassify: an approach to classifying batting shots from cricket videos using a convolutional neural network and gated recurrent unit." Sensors 21.8 (2021): 2846.

10. R. Kumar, S. D. and J. Barnabas, "Outcome Classification in Cricket Using Deep Learning," 2019 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), Bengaluru, India, 2019, pp. 55-58, doi: 10.1109/CCEM48484.2019.00012.

11. K. Soomro, A. R. Zamir, and M. Shah. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. [Online]

12. Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large- scale video classification with convolutional neural networks.In CVPR, 2014.

13. J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long- term recurrent convolutional networks for visual recognitionand description. CoRR, abs/1411.4389, 2014.

14. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In CVPR, pages 19. IEEE, 2015.

15. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In CVPR, pages 28182826. IEEE, 2016.

16. Sen, A., Deb, K., Dhar, P. K., & Koshiba, T. (2021). CricShot10 [Dataset]. https://github.com/ascuet/cricshot10

17. Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., & Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2625-2634).