Classification of Cricket Shots from Cricket Videos using Deep Learning Models

[1] Ishani Bhat, [2] M Anjan Yajur, [3] Taarun Sridhar, [4] Venkatesh B R, [5] Sujatha R Upadhyaya

[1][2][3][4][5][6] Department of Computer Science, PES University, Bengaluru, India
E-mail: [1] ishanibhat4@gmail.com, [2] yajurthejas@gmail.com, [3] iamtaarun@gmail.com, [4] brvenkatesh7@gmail.com, [5] sujathar@pes.edu

Abstract— ***As a part of the AI revolution, Activity Recognition has found numerous applications over the decade. Recognition of cricket shots is one of the less researched domains, however, it has profound applications. Instant classification of cricket shots can be of immense help in training videos, performance analysis of the players or the game itself, and also in enriching the watching experience of the sport. This paper addresses issues involved in manual cricket shot classification, to achieve a state of automated shot recognition in the game. The manual techniques used for shot recognition are time-consuming and require expert assistance at all levels of the sport. Instant recognition of the shots avoids expert involvement and translates into a reduction of time and effort. Furthermore, its application has the potential to increase grassroots engagement and improve the game's analytical component. The success of deep learning algorithms with video and image data is one of the motivations for exploring similar algorithms for cricket shot classification. The suggested methodology uses several deep learning algorithms best suited for activity recognition and compares their performance. Specifically, the study delves deeply into CNN + RNNs, Attention Networks, and Vision Transformer, a specific type of CNN, utilizing temporal and spatial information to improve classification results. The experiments showed exceptional accuracies of 99.19% for Attention Networks, 99.2% for CNN + RNNs and 98.9% for ViTs on the PES dataset. The results obtained for the reference dataset had no significant difference to the PES dataset for ANN and CNN+RNN while slight difference for ViT model, validating our tested model. This technology adds a new level of knowledge by facilitating accurate shot detection and contextual relevance, which helps decision-making processes related to shot selection in all formats of the game which can be extended to other sports as well.***

Index Terms— ***Attention networks, CNN, Cricket shot classification, Deep learning, RNN, Vision Transformers***

I. INTRODUCTION

Video processing integration for intelligent actions in human activity has become industry standard across several sectors [1] [2]. Cricket, being one of the most watched games in the world and specifically the subcontinent region, has seen many attempts to use technology to improve player performance as well as the fan experience. Of these efforts, shot and delivery classification are particularly noteworthy [3].High accuracy shot classification models are vital, as seen by the growing use of AI in coaching and training [4]. In one of the previous reports, backlift technique in batting has been automated in the cricket batting video footage using deep learning Architectures [5] and video footage was used for using various techniques [6]. Unfortunately, precise data is hard to come by at all levels of the sport due to the laborious, time- consuming, and frequently specialized nature of existing manual approaches for identifying cricket strokes. Acknowledging that automated shot detection has the potential to transform cricket statistics and player development [7], our approach aims to accelerate this process without the need for new technology or high-end cameras.

Our methodology is based on the understanding that easily accessible and precise shot detection is essential to the advancement of the sport. It aims to provide players and coaches at grassroots levels with valuable insights and analyses, democratizing access to accurate data and insights beyond the elite level and encouraging widespread improvement and involvement within the cricketing community. In order to obtain more accuracy and improve overall model performance, our study emphasizes training machine learning and deep learning models with large data sets. Efforts were made in dataset development where the volume of clips has been significantly increased. Furthermore, an exhaustive investigation and comparison of different models and algorithms is performed, giving a more nuanced knowledge of their effectiveness in cricket shot classification. The models developed for our dataset were applied to Reference dataset obtained Three sets of experiments are conducted using different model for each and are run on both datasets to get the accuracy, F1 score, precision and recall to compare the performances of each model to decide the best. Thus, the current research focusses on building and training three different models to classify cricket shots accurately. Also, a set of experiments comparison with other datasets to show the robustness of our algorithms.

II. RELATED WORK

In the realm of cricket shot recognition, several methodologies have been explored. A lightweight Convolutional Neural Network based method was applied for replay detection in video to summarize and interpret the outcome [8]. The work on Cricket Stroke Extraction focuses on temporal action localization, utilizing a random forest model and linear SVM camera models [9]. However, it lacks extensive consideration of diverse stroke styles. To address this, our approach involves meticulous dataset curation using the VGG annotator, ensuring a comprehensive representation of all different strokes.

The work on "Outcome Classification in Cricket Using Deep Learning"[10] focuses on ball-by-ball outcomes, achieving a commendable accuracy of 70%. However, limitations include the absence of a standardized dataset. Our approach emphasizes a meticulously curated dataset and explores attention mechanisms, compensating for challenges posed by varying cricket conditions.

In "Deep CNN-based Data-driven Recognition of Cricket Batting Shots,"[11] the use of 2D CNN followed by RNN and 3D CNN achieved high accuracy. Yet, limitations such as robustness under different conditions and a small and imperfect dataset were identified. Our method leverages attention networks, enhancing accuracy and adaptability while addressing these limitations.

"CricShotClassify"[12] introduces a hybrid CNN-GRU architecture for shot classification, achieving an impressive accuracy of 93%. Our approach extends beyond this by incorporating attention networks and accurate dataset creation, minimizing misclassifications due to extraneous frames and different angles.

Our methodology builds upon and addresses gaps identified in related works, contributing to the evolution of cricket shot recognition techniques. Here, we discuss relevant prior research and subsequently introduce our approach, emphasizing attention networks for improved accuracy, adaptability, and a comprehensive dataset for robust evaluation.

Our experiments demonstrate the effectiveness of attention networks, CNN + RNNs, and Vision Transformers in cricket shot recognition. These models exhibit promising accuracy, addressing limitations identified in previous methodologies.

In-depth discussions provide insights into the strengths and limitations of implemented algorithms.

Attention networks, CNN + RNNs, and Vision Transformers offer improved accuracy and adaptability, contributing to cricket shot classification research.

The current research discusses the extension of the dataset created and published by the authors [13] and efforts made to rigorously evaluate various algorithms and models for cricket shot classification. Attention networks, CNN + RNNs, and Vision Transformers emerge as promising contributors to accurate shot categorization.

III. PROPOSED METHODOLOGY

In our endeavor to advance cricket shot recognition, we implemented various algorithms and models, assessing their performance in real-time categorization. The primary goal was to compare common video recognition algorithms, including CNN + RNNs, ViT, and Attention Networks. We utilized TensorFlow, Keras, and scikit-learn for implementation.

For experimentation purposes, two datasets were chosen. The first dataset is the PES dataset. The secondary dataset used is the CricShot10. The PES dataset is made up of 6 shots, 'sweep', 'drive', 'pull', 'slog', 'flick', 'cut', a total of 1922 clips split as shown in Fig. 1.

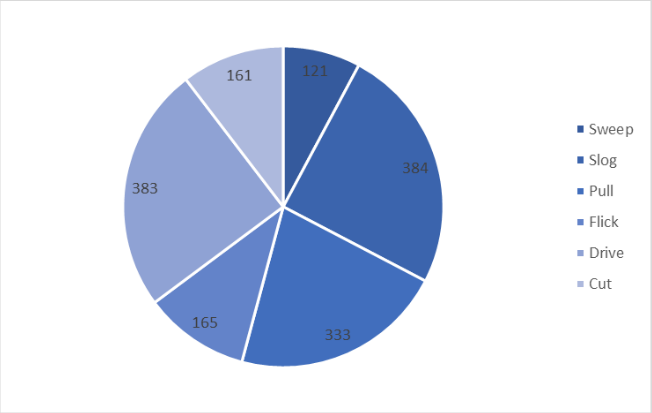

Fig. 1 Shots Split for PES Dataset

The reference dataset is made up of 10 shots, 'sweep', 'cover', 'pull', 'defense', 'flick', 'straight', ‘late_cut’, ‘square_cut’, ‘hook’, ‘lofted’, a total of 1888 clips split as shown in Fig. 2.

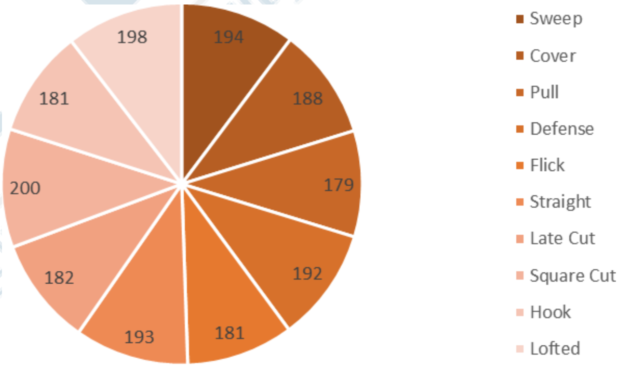

Fig. 2 Shots Split for Reference Dataset

The PES dataset was extended, following the procedure shown in Fig. 3.

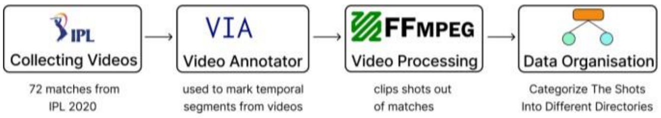
IPL
Collecting Videos
72 matches from IPL 2020
VIA
Video Annotator
used to mark temporal segments from videos
FFMPEG
Video Processing
clips shots out of matches
Data Organisation
Categorize The Shots Into Different Directories

Fig. 3 Dataset extension

The data from the dataset is then processed.

1. Each folder containing a type of shot is iterated through.
2. Each video clip is then read by OpenCV2.
3. Each frame is then resized to 100px by 100px and the data is converted into a numpy array, normalized to a value between [0, 1], and stored.
4. The data is then used to train each model
5. Three kinds of models are used.

A. CNN + RNNs

Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) offer a promising approach to cricket shot categorization, leveraging spatial information. The experiment is setup as shown in Fig. 4:

1. Import necessary libraries for data handling and deep learning. Define shot types and sample fraction.
2. Model Architecture: Define a CNN + RNN model incorporating convolution layers, and a dense layer for classification.
3. Training and Evaluation: Compile the model with an appropriate optimizer and loss function. Train the model using training data, and validate on the test set for a defined number of epochs and batch size. Evaluate the model's accuracy on the test set.

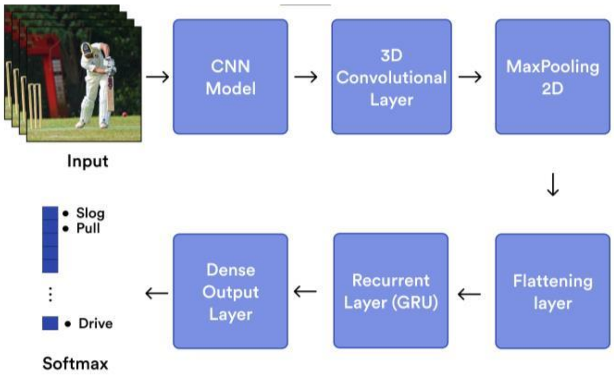

Fig. 4 CNN + RNN Architecture

B. Attention Networks for Cricket Shot Classification

Attention Networks present a powerful methodology for cricket shot classification, focusing on both spatial and temporal features. The architecture includes convolutional layers, reshaping for a time-distributed layer, and bidirectional LSTM with an attention mechanism.

The experiment is set up as shown in Fig. 5:

1. Import necessary libraries for data handling and deep learning. Define shot types and sample fractions.
2. Model Architecture: Define an attention model using Convolutional layers, Reshape, Bidirectional LSTM with Attention, and Dense layers.
3. Training and Evaluation: Compile the model with Adam optimizer and sparse categorical cross-entropy loss. Train the model using training data, and validate on the test set for 20 epochs with a batch size of 32. Save the trained model for future use and evaluate accuracy on the test set.

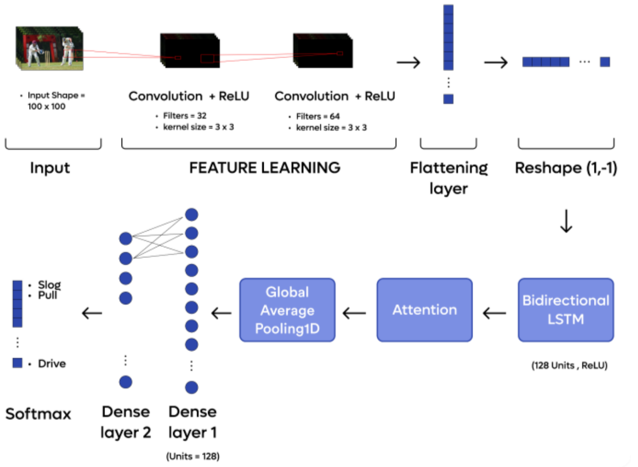

Fig. 5 Attention Networks Architecture

C. Vision Transformer for Cricket Shot Recognition

Vision Transformers (ViTs) represent a paradigm shift in image analysis, emphasizing global picture understanding through self-attention mechanisms and can take advantage of both spatial and temporal information. The model is employed for cricket shot categorization.

The experiment is set up as shown in Fig. 6:,

1. Library Setup: Import necessary libraries, including the ViT model implementation.
2. Model Construction: Build the ViT-based model, configuring the ViT part and adding RNN and output layers.
3. Model Compilation: Compile the model with optimizer, loss function, and metrics.
4. Training and Evaluation: Train the model using training data, and validate on the test set for a specified number of epochs and batch size. Evaluate the model's accuracy on the test set.

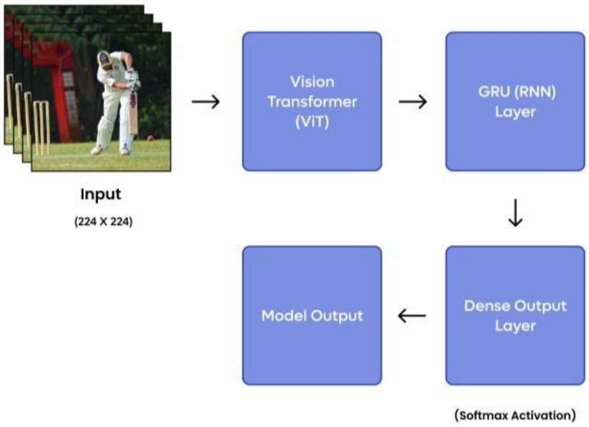

Fig. 6 ViT Architecture

These experiments collectively contribute to our comprehensive evaluation of cricket shot recognition, emphasizing the strengths and adaptability of different models in capturing both spatial and temporal features.

IV. RESULTS AND DISCUSSIONS

The PES and reference datasets were tested under the same experimental conditions for comparison. The total clips considered in the current research are 1922 and 1888 for PES and reference dataset respectively. The complete comparative analysis was done for both the datasets across various performance metrics such as accuracy, recall, precision and F1 measure and the experiment is set-up as represented in Fig. 7. The videos are loaded from the stored dataset and used to train each of the three models. Once the training is done the models are tested with new data from the same dataset to validate and test the. In CNN + RNN the reference dataset outperforms the PES dataset as they have split their shots into 10 categories leading to more clear separation and thus more accuracy while classifying the shot. It resulted in 99.2% accuracy on the PES dataset and 98.65% on the reference dataset (Fig. 8). As expected, a very high accuracy of 99.19% and equal F1 score, recall and precision on the PES dataset and 99.05% on the reference dataset. This is due to the excess frames present after the shot has been played, leading to more irrelevant information. This shows that taking spatial and temporal information together is important and gives more accuracy (Fig. 9). ViT incorporates attention networks which takes into account both spatial and temporal information and has other benefits such as scaling and global perspective. It gives us an accuracy of 98.9% on the PES dataset and 97.16% on the reference dataset (Fig.10). The experiments showcased exceptional accuracies — 99.19% for Attention Networks, 99.2% for CNN + RNNs, and 98.9% for ViTs on the PES dataset while the reference dataset too showed accuracies close to PES dataset (Fig. 11). The statistical analysis of the performance metrics between two datasets showed no significant difference in ANN and CNN+RNN models while ViT model showed slight significant difference for Reference and PES dataset. The difference in performance metrics for ViT model between two datasets may be due to imbalances in class distribution. The PES dataset had a more balanced distribution of classes, the model may perform better in terms of precision, recall, and F1 score. Accuracy is sensitive to imbalances; if one class dominates, a model might achieve higher accuracy by simply predicting the majority class. On the other hand, precision, recall, and F1 score consider the performance of each class independently, providing a more distinct evaluation, especially when dealing with imbalanced datasets. Thus, these results validate that the models tested in the current research can be applied for any such dataset with confidence (Table 1). The event detection in the videos of sports such as baseball, cricket and tennis can be demanding and extremely hard. This is majorly due to unexpected motions of the players with similar outfits as well as movement of camera [14]. It is well proven that in videos of various sports the posture of the player changes every moment [15]. The present research aims to overcome few such challenges through Machine intelligence in one of the well-known sports cricket. The experiments are conducted on the dataset created and compared with another published dataset. The dataset was subjected to Attention Networks, CNN+RNN and one of the recent models that associate the image analysis with self-attention-based architectures, the ViTs

There are reports of creation of large dataset [13][16] that are prerequisite for accurate ML based predictions in sports. In one of the earlier reports deep CNN has been used to recognize cricket shots. The dataset involved 800 batting shot covering various forms such as pull, drive, cut, sweep, hook and flick strokes. The accuracy of the trained model was up to 90% [20]. Deep CNN has been used for classification of cricket shots in dataset comprising of 429 video clips with accuracy of about 98% and 93% for a left-handed and right-handed batsman respectively [17].

In one of the recent reports, researchers have applied a combination of LSTM and time-distributed 2D CNN layers for extracting and training the cricket data obtained from the input sequences. It is known that RNN alone provides less efficient results as compared to other models or RNN in combination with CNN. This has been shown in one of the reports wherein their proposed model was better than RNN and achieved a recall of 91%, F1 score, 91% and accuracy of 92% [18]. LSTM (Long Short-Term Memory) is a well known recurrent neural network (RNN) type of architecture that has been widely applied in domain other than sports such as automatic Speech Recognition[19].

The combination of two deep learning techniques such as convolutional and recurrent neural networks (CNN+RNN) has been applied to classify five different classes of sports including cricket. This analysis run for multiple experiments resulted in classifying test data with accuracy up to 96.66% [20].

Feature extraction distinguish CNNs and ViTs and the latter adopts more creative and innovative strategy for classification and prediction.

The vision transformers (ViTs) have emerged as a promising strategy and being more powerful than CNNs. Though the technology has been exploited in healthcare such as classification diabetic retinopathy and detection of glaucoma[21], it has not been explored in the domain of sports yet. In one of the study, an effective transformer-based strategy for spotting actions in videos of soccer games is reported. To extract features from the videos, a multi-scale vision transformer was applied that lead to outstanding results as compared to the baseline [22].

The three models were applied to both the reference as well as PES dataset and the results of accuracies validated the experiments as no significant difference was observed between the two for the models. This further increases the prospect of using the models for sports having similar outcomes of shots including cricket.

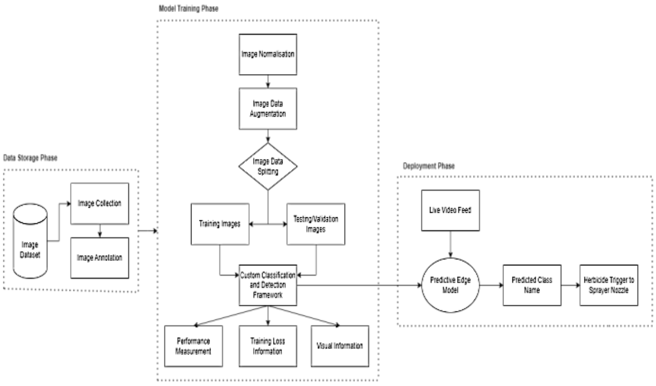

Fig. 7: The experiment set-up in the current study.

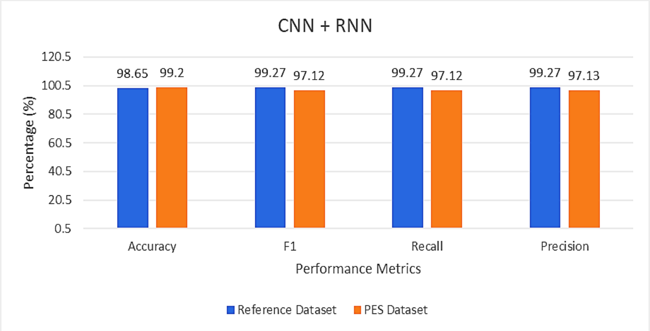

Fig. 8: The performances of the datasets with the CNN + RNN model.

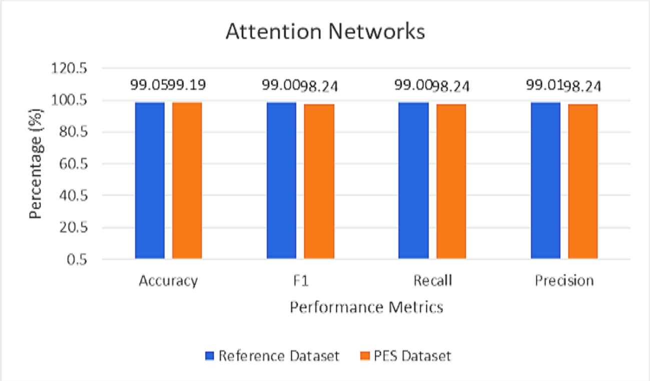

Fig. 9: The performances of the datasets with the attention networks model.

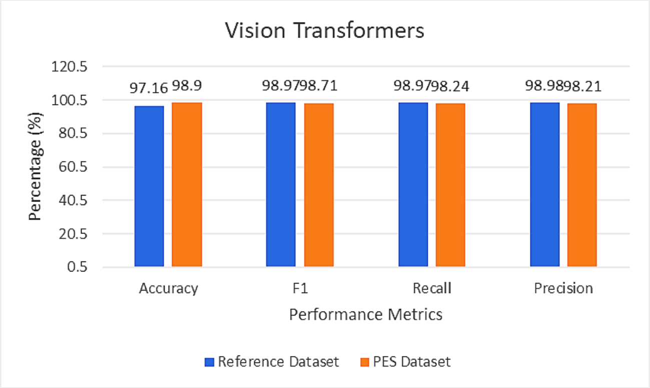

Fig. 10: The performances of the datasets with the ViT model

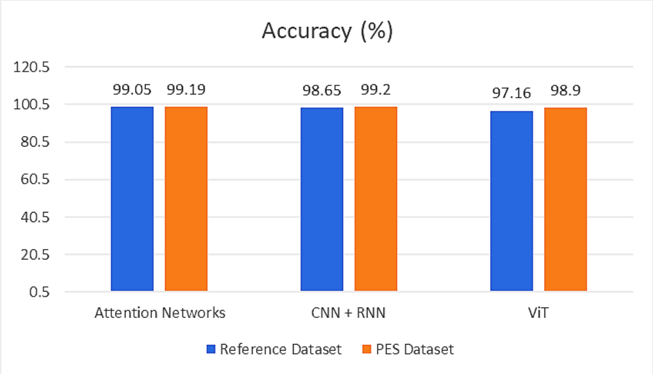

Fig. 11: Comparative analysis of accuracies of three ML models in PES and Reference datasets

Table 1: Comparative Statistical differences in the accuracies of three model under study

S. No.	Model tested	Total Number of video clips & Accuracy		Difference in percent required (95% confidence level)	Significant difference (YES/NO)
		Accuracy (%) Reference dataset (1888)	Accuracy (%) PES dataset (1922)		
1	ANN	99.05	99.19	0.59	NO
2	CNN+RNN	98.65	99.20	0.66	NO
3	ViT	97.16	98.9	0.88	YES

V. CONCLUSION AND FUTURE WORK

In this study, a thorough approach for classifying cricket shots is provided, utilizing Vision Transformers, mixed CNNs and RNNs, and attention networks. Accuracy measurements, which are visually presented in Fig. 11 and offer a clear depiction of our models' performance, have been used to validate our technique. Our study fills in important gaps in the existing methodology by creating and experimenting with precise datasets, providing insightful information for the development of cricket shot identification techniques. We have shown how several models are flexible and effective in capturing temporal and spatial characteristics that are essential for accurate shot classification. In the near future, our models have room for improvement and refinement in the following areas. We expect improved versatility and accuracy across a range of scenarios by training on different camera perspectives, different quality levels, and different lighting conditions. Continual work will enhance the usefulness and application of automated shot classification in cricket, which will help players and coaches at all levels of the game make better decisions.

ACKNOWLEDGEMENT

The authors express their gratitude to Anik Sen, Kaushik Deb, Pranab Kumar Dhar, and Takeshi Koshiba, the contributors of the dataset in "CricShotClassify: An Approach to Classifying Batting Shots from Cricket Videos

Using a Convolutional Neural Network and Gated Recurrent Unit" [12]. We extend our appreciation for providing their dataset, which has been instrumental in the comparative analysis with our dataset.

REFERENCES

[1] H. C. Shih, "A Survey of Content-Aware Video Analysis for Sports," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 5, pp. 1212-1231, May 2018, doi: 10.1109/TCSVT.2017.2655624.

[2] C. Nader, W. Hans, "Artificial Intelligence and Machine Learning in Sport Research: An Introduction for Non-data Scientists", in F*rontiers in Sports and Active Living, Vol 3, 2021, URL* https://www.frontiersin.org/articles/10.3389/fspor.2021.682287, DOI=10.3389/fspor.2021.682287, ISSN=2624-9367

[3] S. H. Emon, A. H. M. Annur, A. H. Xian, K. M. Sultana and S. M. Shahriar, "Automatic Video Summarization from Cricket Videos Using Deep Learning," *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, DHAKA, Bangladesh, 2020, pp. 1-6, doi: 10.1109/ICCIT51783.2020.9392707.

[4] M. Rabbia, A. Irtaza, S. Ur Rehman, T. Meraj, and H. T.Rauf. "A Player-Specific Framework for Cricket Highlights Generation Using Deep Convolutional Neural Networks", *2023. Electronics* 12, no. 1: 65. https://doi.org/10.3390/electronics12010065

[5] Moodley, T., van der Haar, D & Noorbhai, H, "Automated recognition of the cricket batting backlift technique in video footage using deep learning architectures", *Sci Rep*, 2022, **12**, 1895 https://doi.org/10.1038/s41598-022-05966-6.

[6] 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.

[7] M. F. A. Foysal, M. S. Islam, A. Karim, N. Neehal, "Shot-Net: A Convolutional Neural Network for Classifying Different Cricket Shots", *Recent Trends in Image Processing and Pattern Recognition*, 2019, Volume 1035, ISBN : 978-981-13-9180-4.

[8] A. Javed, A. Ali Khan, "Shot classification and replay detection for sports video summarization". *Front Inform Technol Electron Eng* 2022, 23, 790–800, https://doi.org/10.1631/FITEE.2000414

[9] I. Bandara, and B. Bačić, "Strokes Classification in Cricket Batting Videos", *In Proceedings of 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications, CITISIA 2020*, pages 1-6.

[10] R. Kumar, D, Santhadevi Barnabas, Janet. "Outcome Classification in Cricket Using Deep Learning", 2019, *IEEE International Conference on Cloud Computing in Emerging Markets (CCEM*), 55-58. 10.1109/CCEM48484.2019.00012.

[11] M. Z. Khan, M. A. Hassan, A. Farooq, and M. U. G Khan, "Deep CNN based Data-driven Recognition of Cricket Batting Shots", *In Proceedings of International Conference on Applied and Engineering Mathematics*, 2018, DOI:10.1109/ICAEM.2018.8536277

[12] A. Sen, K. Deb, P. K. Dhar, and T. Koshiba, "CricShotClassify: An Approach to Classifying Batting Shotsfrom Cricket Videos Using a Convolutional Neural Network and Gated Recurrent Unit", *Sensors 2021*, 21, 2846, https://doi.org/10.3390/s21082846.

[13] I. Bhat, T. Sridhar, V. B. R, M. A. Yajur and S. R. Upadhyaya, "Building a Video Dataset for Cricket Shot Analysis," *2023 International Conference on Network, Multimedia and Information Technology (NMITCON),IEEE*, Bengaluru, India, 2023, pp. 1-6, doi: 10.1109/NMITCON58196.2023.10276358.

[14] Ullah, J. Ahmad, K. Muhammad, M. Sajjad, S.W. Baik, "Action recognition in video sequences using deep bi-directional LSTM with CNN features", *IEEE access*,2017, 6, pp. 1155-1166.

[15] Nadeem, A. Jalal, K. Kim, "Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy markov model", *Multimedia Tools Appl*., 2021,80, pp. 1-34.

[16] Ahmad, Waqas & Munsif, Muhammad & Ullah, Habib & Ullah, Mohib & Alsuwailem, Alhanouf & Saudagar, Abdul & Muhammad, Khan & Sajjad, Muhammad, "Optimized deep learning-based cricket activity focused network and medium scale benchmark",2023, *AEJ - Alexandria Engineering Journal*, 73, 771-779. 10.1016/j.aej.2023.04.062.

[17] Khan, M. Z., Hassan, M. A., Farooq, A. & Khan, M. U. G. "Deep CNN based data-driven recognition of cricket batting shots", 2018, "*International Conference on Applied and Engineering Mathematics" (ICAEM), IEEE* 67– 71.

[18] Semwal, A., Mishra, D., Raj, V., Sharma, J., & Mittal, A., "Cricket shot detection from videos", 2018, *9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–6.

[19] Jane Ngozi Oruh,Serestina Viriri,Adekanmi Adegun, "Long Short- Term Memory Recurrent Neural Network for Automatic Speech Recognition" , 2022, *IEEE Access* 10:30069-30079,DOI: 10.1109/ACCESS.2022.3159339,

[20] M. A. Russo, A. Filonenko and K. -H. Jo, "Sports Classification in Sequential Frames Using CNN and RNN," 2018, *International Conference on Information and Communication Technology Robotics (ICT-ROBOT), Busan, Korea (South)*, pp. 1-3, doi: 10.1109/ICT-ROBOT.2018.8549884.

[21] Wu JH, Koseoglu ND, Jones C, Liu TYA. "Vision transformers: The next frontier for deep learning-based ophthalmic image analysis" 2023, *Saudi J Ophthalmol*., Jul 14;37(3):173-178. doi: 10.4103/sjopt.sjopt_91_23. PMID: 38074310; PMCID: PMC10701151.

[22] He Zhu, Junwei Liang,Chengzhi Lin,Jun Zhang, Jianming Hu, "A Transformer-based System for Action Spotting in Soccer Videos MMSports" 2022, *Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports*,Pages 103. 109https://doi.org/10.1145/3552437.3555693.