AN INTERACTIVE APPROACH TO IDENTIFY CRICKET SHOTS THROUGH DEEP LEARNING MECHANISM

Atharv Pushkraj Nirgude*1, Rohit Dinesh Sonone*2, Sahil Vikram Sonawane*3, Rushikesh Sham Ahire*4, Prof. Balaji Bodkhe*5

*1,2,3,4UG Student, Department of Computer Engineering, MESCOE, Pune, Maharashtra, India.
*5Professor, Department of Computer Engineering, MESCOE, Pune, Maharashtra, India.

ABSTRACT

In recent years, numerous sports have received significant attention and appeal. In the height of the current epidemic, the unavailability of any athletic events had left a significant number of people yearning to observe some game being played. Cricket is arguably the most popular sport in India, with millions of followers who watch the games obsessively. Fans are enthralled by the game and conduct in-depth analyses of the individual players and their abilities, particularly their shot choices. With the growth of fantasy leagues as well as other similar services, there is a growing interest in evaluating individuals who are playing well so that they may be selected for their teams. The manual procedure of batter shot recognition is amongst the most time-consuming and labor-intensive processes that might be automated. As a result, this study provides a successful cricket shot assessment approach that makes use of deep learning in the form of Convolutional Neural Networks to fulfill its objectives.

Keywords: Image Normalization, Convolutional Neural Networks, Cricket Shot Identification.

I. INTRODUCTION

Cricket is one of the most common and highly popular game being played in the Indian subcontinent. Cricket is a complex game that is played between two teams. These teams have to go through a toss to decide their play, if it is batting or bowling. The game of cricket is exhilarating and quite strategic in nature. This allows the game to be fairer without any significant advantage to a particular team throughout. The game of cricket can go on for several hours testing the player’s strength, endurance and resilience. A large number of Indians are fanatical about the game of cricket. These fans are extremely into the game and are well aware of the technological advancements that have been happening in this space.

There have been improved usages of the advances in technology that have been crucial in the development of effective methodologies that have been crucial in the enhancement of the sport and its experience to the players as well as the audience as a whole. This is an effective strategy to allow for the game to evolve and become better over the years. This has also led to a lot of technicalities that have entered the game permitting the effective implementation of various rules and analysis of the player’s abilities. This enables a much better control and fine tuning of the performance and achieving the results that are expected towards the maximum potential.

The proliferation of multimedia information such as photographs and videos via the internet in recent years has resulted in an expansion of concepts and scientific investigation across numerous disciplines. Athletics and sports are one such subject where a great amount of information may be analyzed. The cricket game, as one of the most popular sports on the planet, is extensively aired on television and other media platforms, with a diverse variety of camera motion views and events, as well as review and evaluation. In terms of indexing cricket games, categorizing distinct types of cricket strokes is a challenging but necessary task.

Shot detection techniques are extremely useful in situations where the batter needs to enhance his game and limit the deviation of his strokes from the conventional approach. The difficulty in distinguishing different cricket shots stems from the limited amount of data frames available for extraction of diverse characteristics and signals. As a result, automated algorithms that can accurately recognize cricket shots from real-time recordings must be developed.

Determining the sort of shot executed by the batter has its own set of difficulties. Because hitting a certain shot does not always culminate in a successful hit, we can only comprehend the desired shot. Other elements that contribute to the task being more difficult include occlusion, the tiny size of the ball, and the velocity of motion.
Furthermore, extracting only the necessary information from the data is a time-consuming and difficult effort, especially given the size of the data set provided. As a result, the deep learning paradigm is the optimal way for achieving successful cricket shot analysis.

The approach stipulated in this research article describes the cricket shot analysis that allows for a much better analysis of the shots automatically through the use of deep learning mechanisms. The presented approach utilizes the video input of any cricket shot being played through the use of Convolutional Neural Networks. The CNN model is trained using the dataset consisting of various different shots played by a batsman. The dataset is initially preprocessed and the resultant preprocessed images are normalized and then provided to the CNN model for the training. Once the trained model is ready, it is used for the purpose of achieving the testing using the input video which is effectively preprocessed and normalized before being subjected for the detection through CNN. The outcomes are effectively classified using the decision making mechanism to achieve the cricket shot analysis.

II. LITERATURE REVIEW

A. Javed et al. [1] devise a computationally efficient hybrid technique for detecting replays in video summarization. The score captions (SC) in RSs are omitted by broadcasters. Furthermore, replay frames feature several slow transitions (GTs). The current work makes use of these two insights to detect replays. More precisely, the suggested framework detects replays using GTs and SCs, which are subsequently utilized to generate highlights. The dual-threshold-based technique is employed for GT detection to attain this purpose. Identified GT frames are utilized to extract possible RSs. SC detection relies on candidate RSs. The approximated SC is utilized to distinguish between replay video frames and live video frames. The suggested solution is resistant to camera modifications, replay speed, logo design, size, and location, among other things.

A. S. Rao et al. used Riemannian manifolds to construct a probabilistic recognition of crowd events on OFMs. A motion-based, probabilistic paradigm for crowd action identification has been suggested. A novel technique for detecting walking and running events, which exploits image sequences vector durations in OFMs, has been published in particular [2]. In particular, the framework includes a mechanism for detecting merging and dividing events that makes advantage of Riemannian connections in the OFB. The method fared well in detecting all events and beat other algorithms in merging, splitting, and dispersion.

D. Tang introduce the Hybridized Hierarchical Deep Convolutional Neural Network that has been designed to improve picture segmentation accuracy in sports athletics exercise rehabilitation. In the convolutional neural network, designing an image segmentation technique to sample the multi-layer convolution output. The image may be segmented into several super pixels using a super pixel segmentation technique. The classifier is trained to use the hierarchical characteristics of the super pixels, and the classification results are transferred to the pixel. Finally, a random field method with paired potential energy and one potential energy is available for a completely linked layer [3]. The energy function is utilized to smooth the pixel classification result and improve the pixel label's consistency and regional dependability. The results of the experiments show that it improves convergence speed, decreases training time, and improves segmentation accuracy.

The deep learning technique and big data of sports medicine are evaluated and investigated by H. Ma et al. The major goal is to address the flaws in classic deep learning algorithms in the framework, as well as the problem of training mode loss. Simultaneously, this research provides a semi-physical cloud-end fusion simulation model that may be utilized for genuine cloud-end fusion. To accomplish accurate and effective mining of massive data in sports medicine, the system provides reference and technical assistance [4]. In terms of practical details, this research utilizes deep learning algorithms to achieve effective prediction and risk assessment of sports medicine-related diseases, then uses a resampling algorithm with a self-regulation function and a tensor convolution self-coding neural network model to help with multi-dimensional sports medicine data analysis.

W. R. Johnson et al. use a subset of marker trajectories recovered from historical motion capture sidestepping trials to see if pre-trained CNN models can be applied to enhance 3D GRF/M predictions beyond what PLS could achieve. The approach’s accuracy and validity were evaluated by comparing mean alignment between GRF/Ms produced from ground truth-force plate data to those predicted by a series of novel models ranked using the
CNN algorithm [5]. It was expected that fine-tuning a pre-trained CNN model will beat prior PLS accuracy findings, especially for GRMs with higher noise and non-linearity.

R. Ji investigates basketball shooting gesture recognition. In the realm of machine vision, human motion tracking and posture recovery have several applications. However, because of the inherent variability of human posture, the dimensionality of the observation data space is too large, the expression of human video image features is complex, and the influence of different experimental environments, etc., human motion tracking remains an unsolved urgent need in the field of machine vision. In light of the aforementioned issues, this research begins with the active principle of basketball shooting and then goes on to explain the research's history, motivations, and present state [6]. The precise detection of basketball shooting movements is obtained by expanding the theory of picture feature recognition using Gaussian hidden variables. The excellence of the gesture recognition technology obtained is demonstrated by a practical case study and vibration impact evaluation.

M. Moness et al. suggested two techniques, Canny and Fuzzy algorithms, and developed a system that automatically extracts 101 feature points from front and side photos of a human body. These feature points may assess 36 different body measures. The swimming game explained in this research requires just 8 body dimensions out of the 36 available. Therefore, the technique is not confined to swimming sports but may be used in a variety of other activities based on their anthropometric measures [7]. The eight body measures and weight are utilized to identify junior swimmers who match the best performance standards.

J. Ševčík et al. developed an image prior with adaptive correlations, which relates to the automated estimate of filter coefficients from data. The authors present a hierarchical Bayesian Gaussian prior model with unknown correlations amongst each pixel and its neighbors [8]. The correlation coefficients are computed from the observed picture, which may be viewed as locating the best appropriate transformation, especially the coefficients of heterogeneous filters. The suggested model also permits for an accurate Variational Bayesian solution without approximation, and the resultant method is devoid of tuning parameters. The authors used the suggested before outperform approaches depending on first order difference prior and total variation before the challenge of super-resolution reconstruction.

T. Matsui et al. suggest a self-contained de-fencing algorithm. The suggested technique can precisely locate the fence zone and recover the lost section organically. The suggested network can recognize fence zones and recover the concealed background without any user involvement by merging unique CNN algorithms and standard image processing approaches. On several real-world fence photos, experimental findings show that the presented technique outperforms other state-of-the-arts [9]. Furthermore, because the authors define this fence identification task as a regression issue rather than a classification challenge, their suggested detection network can deal with uneven fence patterns. Furthermore, during the picture completion step, they generate synthetic fence images to train the network. Because of their learning-based methodology, the network is resistant to a wide range of fence pictures.

R. Yuan et al. investigates the development of a sport virtual video scene and compares and analyzes the visual algorithm model of its sports training. The human movement style is an essential component of animation, film, and video games. The display must be accomplished by the collection of real-world human motion data, followed by manual processing at a later stage. Therefore, this essay investigates human movement synthesis and style transmission in depth. A style transfer model was developed by extracting the spatiotemporal aspects of human movement data to automatically synthesize the relevant human movement style data [10]. To achieve style transfer, this research offers a combination of RBM and self-encoding motion style transfer model to map the original human motion data to the feature space, impose style transfer constraints in the feature space, and map the features to the original human motion space. The model employs a self-encoding network with symmetric encoding and decoding properties, making it simple to encode the original human motion data, map it to the feature space, and then decode to acquire the produced human motion data.

J. Chen et al. examine the less-studied topic of acquiring individualized intransitive visual preferences [11]. The author’s present multi-criterion models depend on the intuitions that users have various criteria to choose from when making preference judgments on different pairings of images. They leverage the observed pairwise comparison findings as well as image content aspects to describe customized intransitive preferences. The
authors create a fresh dataset of pairwise comparison findings on photos to test the performance of the suggested models. The suggested models outperform all baselines in terms of accuracy, according to the experimental data.

T. Guo et al. present a two-phase cascaded CNN approach for detecting individual ice hockey players during ice hockey games. Phase I of the cascaded architecture detects the targeted players by filtering out most of the distracting details, such as the audience and touchline advertising bars, while Phase II refines the results derived from Phase I's output by incorporating specific information, such as overlapping areas of body position caused by individual player movements and the uniform colors of the two teams. Images from the 2018 Winter Olympics were used to construct a dataset that was divided into training and testing datasets [12]. The distribution of all players' aspect ratios was then determined from the training data using a deep learning framework to design an adequate bounding box to manage the complex situation when players demonstrate a range of postures. Following player detection, the regions containing the detected players' uniforms are cropped, and the features of the uniform color distribution are represented through five color channels that are preliminarily divided based on the uniform color features' statistics to recognize the team affiliation.

T. T. Kim et al. introduce analysis and method for finding critical video frames in a whole golf swing using vision-based posture estimate to aid in delivering feedback for improvement. The suggested technique evaluates crucial times in the golf swing utilizing both still images and movies and may determine parameters of the player like posture, swing pace, and swing stability. By developing a route projection, these crucial frames can also forecast the swing consequence [13]. The photos and videos that were processed were utilized to examine the golf swing from a down-the-line standpoint. To do assumption and data processing, the authors use a low-cost tensor processing unit, which sets the execution baseline for the video recording system. Through the gathering of footage at a driving range, the authors generated some data, highlighting the need for data filtering and a deeper understanding of camera technology. Then, to anticipate swing outcomes, a KFS plotter of key points is utilized.

III. PROPOSED SYSTEM

The proposed methodology for the Cricket shot identification and analysis through the use of Convolutional Neural Networks is depicted in the above figure 1. The below mentioned steps detail the step by step execution of the methodology.

Step 1: Preprocessing – The proposed model for Cricket shot identification and analysis utilized a dataset that is generated manually for this purpose. The dataset generation process was initiated by downloading videos from the popular video sharing website YouTube pertaining to those shots. The proposed approach is defined to identify 5 different types of shots. Among those shots are the Cover Drive, Pull Shot, Slog Sweep, Square Cut, and Straight Drive from both left and right handed batsmen. The collected videos are effectively preprocessed and the frames containing the shot are extracted. The resultant dataset is achieved consisting of 2074 images of
various shot types which are segregated equally into training and testing directories utilized in the next step of the methodology.

**Step 2: Image Normalization** – Before beginning of the training, the cricket shot images are resized to the dimension of 170 X 170.

In this phase of the proposed approach, an ImageDataGenerator object is built utilizing the ratio 1/255 and the libraries TensorFlow and Keras for detailed evaluation. This protocol is used to initiate training and testing on cricket shot images. The ImageDataGenerator object is initialized using characteristics such as training and testing folder locations, image sizes, batch size of 32, and categorical class mode with grayscale as the color mode.

**Step 3: Training with Convolutional Neural Network** – The Sequential class of the TensorFlow package is used to create a sequential neural network architecture. After that, as the very first layer of the Neural Network only for corresponding dimension of the pictures, a convolution layer with 32 kernels of size 3 X 3 is inserted with "ReLU" activation function. After that, an additional Convolution layer with 64 kernels of size 3 X 3 and the "ReLU" activation function is introduced. A maximum pooling layer of size 2 X 2 with a dropout rate of 25% is established.

The third convolution layer is added with the size of 3 x3 of each of the 128 kernels. The activation function being used for this purpose is the ReLU activation function. The size of 2x2 is given for the max pooling layer. After the third layer, another layer, the fourth layer is deployed with the ReLU activation function using 128 kernels of size 3x3. Another Max pooling layer is introduced with the dropout of 25% and the size set as 2x2.

Finally, through using flatten function and a dense layer of size 1024 with the "ReLU" activation function, the neural network is terminated. With two dense layers and the "softmax" Activation Function, a dropout percentage of 50 is specified at the conclusion of the convolution neural network.

The Adam optimizer is often used to improve the outcome with 500 Epochs even during learning phase for all 5 distinct shots, including the Cover Drive, Pull Shot, Slog Sweep, Square Cut, and Straight Drive from both left and right handed batters. The trained data is saved in a H5 file after training and is consumed by the model during testing. The architecture of the Convolutional Neural Network is shown in Figure 2 below.

**Step 4: Testing through Decision Making** - The saved trained model data in h5 file format is loaded into the testing image neural network object during the testing procedure. Integer predictions are made using this information. The dictionary of classes recognizes the cricket shot predicated on the integer index and displays it together with the assessment to the user.

**IV. RESULTS AND DISCUSSIONS**

The suggested approach for Cricket Shot identification and analysis is written in the Python programming language on a Windows-based workstation. The Spyder IDE is used for coding this strategy. The deployment machine features an Intel Core i5 CPU, 8GB of RAM, and a 1TB hard drive.

For a successful deployment of the Convolution neural Network, the reliability of the Cricket shot recognition and analysis technique must be evaluated. This method uses a picture containing five distinct sorts of cricket shots from left and right handed batsman as input as shown in the figure 3 below.
The RMSE performance indicator is successfully used to evaluate the performance of the Cricket Shot Detection. The experimental evaluation is discussed in the section given below.

Performance Evaluation through Root Mean Square Error

The root mean square error (RMSE) is calculated to determine the error rate of the proposed approach. The RMSE is utilized in this experiment to calculate the error rate between the actual cricket shot detection and the expected cricket shot detection using the CNN module. The RMSE technique is depicted in equation 1 below.

\[
RMSE_f = \left[ \frac{1}{N} \sum_{i=1}^{N} (Z_i - Z_{oi})^2 \right]^{1/2}
\]

Where

- \(\Sigma\) - Summation.
- \((Z_i - Z_{oi})^2\) - Differences Squared for the Cricket Shot Detection.
- \(N\) - Number of Images.

The Mean Square Error, or MSE, must be computed first before the Error Rate of the Approach can be estimated using RMSE. The MSE is the variation among the actual cricket shot detection accomplished and the expected cricket shot detection. The entire project is being tested with an expanding quantity of trials, with the results recorded in Table 1 below. The results acquired are used to create the graph shown in Figure 4 below.

<table>
<thead>
<tr>
<th>Cricket Shot</th>
<th>Number of Iterations</th>
<th>Correctly identified Cricket Shot</th>
<th>Incorrectly identified Cricket Shot</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cover Drive</td>
<td>10</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pull Shot</td>
<td>10</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Square Cut</td>
<td>10</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Straight Drive</td>
<td>10</td>
<td>9</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Slog Sweep</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

![Figure 3: Different Cricket shots for the input](image1)

![Figure 4: No of proper expected of scores V/s No of Obtained Scores](image2)
The average MSE is calculated using MSE measured value following extensive testing with the CNN component of the Cricket Shot detection technique. The RMSE value of 1.264 is calculated by square rooting the average MSE. A low mistake rate implies that the CNN Model was implemented effectively. As a result, the cricket shot detection execution accuracy improves dramatically. The model accuracy achieved by CNN in the system is shown in the figure 5 below.

Figure 5: Model Accuracy

V. CONCLUSION

The process described in this research paper proposes a technique for automatically analyzing cricket shots that uses deep learning strategies to reach a significantly better result. Convolutional Neural Networks are used in the described technique to utilize the video input of each cricket shot being played. A dataset of numerous distinct shots performed by a batsman is used to train the CNN model. The dataset is first preprocessed, and the preprocessed pictures are then normalized before being fed into the CNN model for training. Once the trained model is complete, it is utilized to do the testing utilizing input video that has been adequately preprocessed and normalized without first being exposed to CNN detection. To perform the cricket shot assessment, the results are efficiently categorized utilizing the decision-making process. The outcomes achieved by the prescribed approach are outlined in the research article in the results and discussion sections. The results have demonstrated the superiority of the proposed approach.

For the purpose of future research directions, the presented technique can be effectively converted into an API for easier integration and allow for a real time application in the future.

VI. REFERENCES


