

# Cricket Event Recognition and Classification from Umpire Action Gestures using Convolutional Neural Network

Suvarna Nandyal<sup>1</sup>

Department of Computer Science & Engineering  
P. D. A. College of Engineering (Affiliated to Visvesvaraya  
Technological University, Belagavi-590018)  
Kalaburagi-585102, Karnataka, India

Suvarna Laxmikant Kattimani<sup>2\*</sup>

Department of Computer Science & Engineering  
B.L.D.E.A's V.P.Dr.P.G.Halakatti College of Engineering  
and Technology (Affiliated to Visvesvaraya Technological  
University, Belagavi-590018), Vijayapur-586103  
Karnataka, India

**Abstract**—The advancement of hardware and deep learning technologies has made it possible to apply these technologies to a variety of fields. A deep learning architecture, the Convolutional Neural Network (CNN), revolutionized the field of computer vision. One of the most popular applications of computer vision is in sports. There are different types of events in cricket, which makes it a complex game. This task introduces a new dataset called SNWOLF for detecting Umpire postures and categorizing events in cricket match. The proposed dataset will be a preliminary help, it was assessed in system development for the automatic generation of highlights from cricket sport. When it comes to cricket, the umpire has the authority to make crucial decisions about on-field incidents. The referee signals important incidents with hand signals and gestures that are one-of-a-kind. Based on detecting the referee's stance from the cricket video referee action frame, it identifies most frequently used events classification: SIX, NO BALL, WIDE, OUT, LEG BYE, and FOUR. The proposed method utilizes Convolutional Neural Networks (CNNs) architecture to extract features and classify identified frames into Umpire postures of six event classes. Here created a completely new dataset of 1040 images of Umpire Action Images containing these six events. Our method train CNNs classifier on 80% images of SNWOLF dataset and tested on 20% of remaining images. Our approach achieves an average overall accuracy of 98.20% and converges on very low cross-entropy losses. The proposed system is a influential answer for generation of cricket sport highlights.

**Keywords**—Cricket match; computer vision; deep learning; SNWOLF dataset; umpire recognition; umpire action images; CNN; event classification

## I. INTRODUCTION

The emerging areas like Artificial intelligence (AI) and machine learning (ML) are transforming modern society. An important subset of machine learning is deep learning, which can be employed to recognize images and speech. Deep learning is rapidly being applied in sports for a variety of applications, due to the advancements in computer vision. The use of AI is ubiquitous, from helping executives match in the decision-making process to helping athletes train on physical aspects. In the last few years, advances in TV channel transmission and CCTV technology have made it possible to exploit vast volumes of data and vision in the computer

research for sporting activities is growing. CNN is a highly efficient deep learning feature extractor and classifier for detecting various details and patterns in images. Time related physical movements of the fingers, hands, arms, head, face, or torso that are expressive and significant aimed at assigning meaningful information or interacting with the environment.

In recent years, gestures have become widely used by humans to interact with computers and machines and most everyday devices such as televisions, smart phones, and car dashboards can now be controlled with simple hand gestures. Gestures of cricket sports helps the umpire to take decision, gesture recognition for event recognition, development of assistive devices for the hearing impaired, sign language recognition, patient emotional medicine that allows very young children to interact with computers. It is also used in many areas such as gesturing [1] condition or stress level, lie detection, driver attention / fatigue monitoring, etc.

Incorporating gesture recognition technology into the sport makes game play fairer and more proficient. Gestures made by sports people show what meaningful information can be derived. Gestures are intended by those who move parts of their body in a predefined way to help with specific highlights, video context labeling, and more importantly, decision making and automatic extraction of highlights. Cricket is an important sport in many countries, the second most popular sport in Asia, the fourth most popular in Europe and the second most popular sport in the world [2]. However, since the 19th century, the same old manual method has been used to update the scoreboard. This puts a lot of strain on the scoreboard. The manner of viewing the rating has modified loads in time, however the simple rating updating manner remains the equal and achieved with the aid of using a person. Therefore, in the 21st century, there is an urgent need for automated systems in this area. Automated systems are replacing many of the tedious manual tasks performed by humans five or ten years ago. Therefore, there is an urgent need to develop such a fully functional automation system. Using state-of-the-art equipment and machine learning algorithms, one can improve the accuracy of decisions and provide perfect results that are evasive to the naked eye. This has led to the design of a system that recognizes cricket referee gestures in real time. In

\*Corresponding Author.

addition, the reason for choosing cricket is that there are not enough technical applications to make cricket more impartial and accurate. Therefore, cricket has become an interesting topic for deep learning researchers doing computer vision-based research to identify and recognize the referee's various postures to extract the highlights of the cricket match.

In this work there is no existing record of cricket referee behavioral gestures. Created and named own dataset- "SNWOLF"- to train and evaluate the performance of the proposed method. The dataset includes 1040 umpire action images and proposed a method which uses SNWOLF dataset for identifying the umpire event postures from umpire action images and classified into six event classes for Highlights Extraction using CNNs. Initially the Convolutional neural network classifier was pre-trained on 80% Umpire Action Event images, like SIX, NOBALL, WIDE, OUT, LEG BYE and FOUR, from SNWOLF dataset and remaining 20% Umpire Action images are tested. Our method can be noted as novel approach for state-of-the art techniques.

**SNWOLF Dataset:** SNWOLF is collections of Umpire Action images. We use this dataset to train CNN machine learning and computer vision algorithm. SNWOLF dataset has 1040 Umpire Action colored images. It has Six classes, and they are an SIX, NO BALL, WIDE, OUT, LEG BYE, FOUR. The images are of size 32x32 pixels. The dataset consists of 832 training and 208 testing examples. It is a database for people who want to try learning techniques and pattern recognition methods on real-world data while spending minimal efforts on preprocessing and formatting. We will use this database in our experiment.

## II. RELATED WORKS

In the area of cricket activity detection, some work is being done using computer vision. For ball-by-ball cricket video classification, Dixi et al. [3] compared three separate CNN architectures. To classify each ball into distinct results, they used the VGG16-CNN framework, which had already been pre-trained for transfer learning. In a game of cricket, Batra et al. [4] suggested an automated multifaceted perceptual approach for detecting balls thrown. To extract the highlight event of the cricket match, Harry Al. [5] employed the intensity projection profile of the referee. Hari et al. [5] used the referee's intensity projection profile to extract the highlight event of the match of cricket. To recognize foot crossings, Chowdhury et al. [6] recommended using video processing technologies. Lazarescu et al. [7] categorized cricket footage in broad terms based on camera movement metrics. The batsman's motion vector was employed to detect cricket shots by Karmaker et al. [8]. Semwaletal et al. [9] used saliency and optical flow to emphasize static and dynamic cues from cricket video, and then used CNN to extract feature representations for these cues. Finally, they adopted a Support Vector Machine (SVM) [10] to categories' cricket shots premised on those characteristic representations.

There exists limited work for classification of cricket sports images to detect cricket events. It has been noticed that an approach, based on Convolutional neural networks, provides the most precise detection. Very few explorations have been conducted by experts in the of cricket event

detection. However, none of them have attempted to recognize Events – like SIX,NOBALL,WIDE,OUT,LEG BYE,FOUR - using deep learning(CNN) method. Our findings may be the basis for future research on cricket video summarization and query-oriented highlight extraction.

All the current vision systems mainly rely on deep learning approach [11] which is a part of machine learning[12] attracted many researchers and the results of applying to this technique in many fields [13], [14], [15] are becoming more encouraging. L. Y. Deng and Y. Liu [16] provides event feature manipulations at multiple levels and correctly detects interested events, video indexing with merge for what user preferences. Aravind Ravi et al.[17] used pre-trained VGG19 and Inception V3 networks for feature extraction and highest classification results are obtained using linear SVM.Rabia A. Minhas et al.[19] proposed an effective shot classification method based on AlexNet Convolutional Neural Networks (AlexNet CNN) for field sports videos and achieves the maximum accuracy of 94.07%. In [20-22] provides large scale image Net for visual recognition, umpire-signals for different activities and scene classification for video summarization. Karen Simonyan et al.[23] suggested an evaluation of networks of increasing depth using an architecture with very small (3x3) convolution filters. In deep learning, a machine models learns to perform independently, the classification tasks from images, text, or sound with almost accurate results as compared to past. The models get trained by huge labeled data templates and neural network structure with many layers.

Table I shows the Research Gaps in this. So we have carried out the work considering the current scenario of digitization of cricket sports video. In this paper umpire gestures are recognized and classified for highlights generation, video summary creation and further work is query based event retrieval from the cricket sports video at global level.

Contributions of the proposed work:

- In this work a novel deep learning framework has been designed to recognize Umpire Action Image such as SIX, NOBALL, WIDE, OUT, LEGBYE, FOUR etc which can be used for Cricket Video Highlight Summarization. It consists of 14 layers of Convolution, pooling, flattening and fully connected layer.
- A designed dataset consisting of 1040 Umpire Action Images of events like SIX, NOBALL, WIDE, OUT, LEGBYE, FOUR has been built to train and evaluate the model.
- This work mainly concentrates on calculating Feature map for each Umpire Action Image and classify it among six Umpire Event Classes- SIX, NOBALL, WIDE, OUT, LEGBYE, FOUR etc. Hence CNN has been applied to reach the acceptable accuracy of the model.
- Rigorous experimentation has been carried out to tune the model and the relationship between model accuracy and parameters is presented. 80% dataset images used for huge training set and remaining for testing to achieve model accuracy of 96.48%.

TABLE I. IDENTIFICATION OF RESEARCH GAPS AND LIMITATIONS FROM EXISTING MODEL

Sl.No.	Author	Algorithm	Merits	Research Gaps	Accuracy
1	Kalpit Dixit et al.[3]	Single Frame Based architecture using softmax probabilities	Softmax function translates these raw scores into softmax scores which helps for classification	The model can be extended to more general problem of commentary generation	80%
2	N. Batra et al.[4]	Trajectory Approximation Algorithm	Detection of no-ball and wide ball making accurate decisions	Technique could be extended to check for faults, order of service in doubles and aces.	85%
3	M.Lazar escu et al.[7]	Incremental learning Algorithm	Camera motion parameters define the trajectory of the ball.	straight drive shot was unable to accurately classify the shot	77%
4	Md Nafee Al Islam et al.[18]	CNN model with transfer learning	Model classifies Cricket bowlers based on bowling actions	Extend our work and include bowlers from all the cricket playing nations.	93%
5	Shahid Karim et al.[24]	Inverted residual network architecture	SSD method quickly and accurately recognize the multi-gesture hands in video	Design more advanced CNN network to improve the accuracy of gesture recognition.	90%

### III. PROPOSED MODEL FOR CLASSIFYING UMPIRE ACTION FRAMES

#### A. SNWOLF Dataset

The existing image dataset for event classification of the umpire action frame could not be found. Work accumulated pictures of Umpire completing a variety of tasks actions related to types of events as an example: “SIX”, “NO BALL”, “OUT”, “LEG BYE”, “FOUR” and “WIDE”. There are six types of data in the dataset. To train and test the model, we had to create our own dataset called "SNWOLF". Each class is made up of 174 images, and each of the six classes provides a total of 1040 images. Fig. 2 depicts some of the photographs from the dataset for each of the six types of events. From the dataset, we trained the model with 832 images and remaining 208 images were used as a validation set to achieve test accuracy.

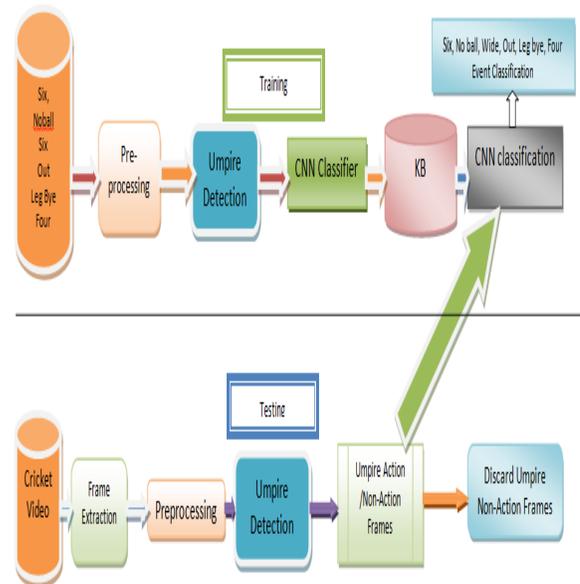


Fig. 1. System Architecture for Cricket Event Classification.



Fig. 2. An Example of a SNWOLF Dataset showing an Image of a Referee for Actions such as: “Six”, “No Ball”, “Wide”, “Out”, “Leg Bye” and “Four”.

Fig. 1 shows the proposed model is a deep learning based model which makes use of convolution neural networks. There are two phases, one is training phase and another is testing phase. During training, the model is trained on 80% of SNWOLF dataset images and knowledge base is created to test remaining Umpire Action Images for classification among six classes of events. The CNN and its layers are explained in Section subsequent sections.

### B. Convolution Neural Network

CNNs (Convolutional Neural Networks) [11, 12] are a deep learning architecture that is being used to classify images [13]. Yann Le Cunetal. [14] first, introduced the idea of a Convolutional neural network that can be trained by back propagation. Jing Yu, Hang Li, Shou-Lin Yin, and Shahid Karim [24] offer a deep learning method for recognizing dynamic gestures in Human-Computer Interfaces using Transfer learning and to recognize different gestures quickly and effectively but the accuracy improvement is required in gesture recognition. Convolutional neural network's design became well-known with researchers that specialize in deep learning, with the introduction of LeNet5 in [15] and demonstrated outstanding performance in handwriting recognition and our proposed framework of Convolutional Neural Network is as shown in Fig. 3.

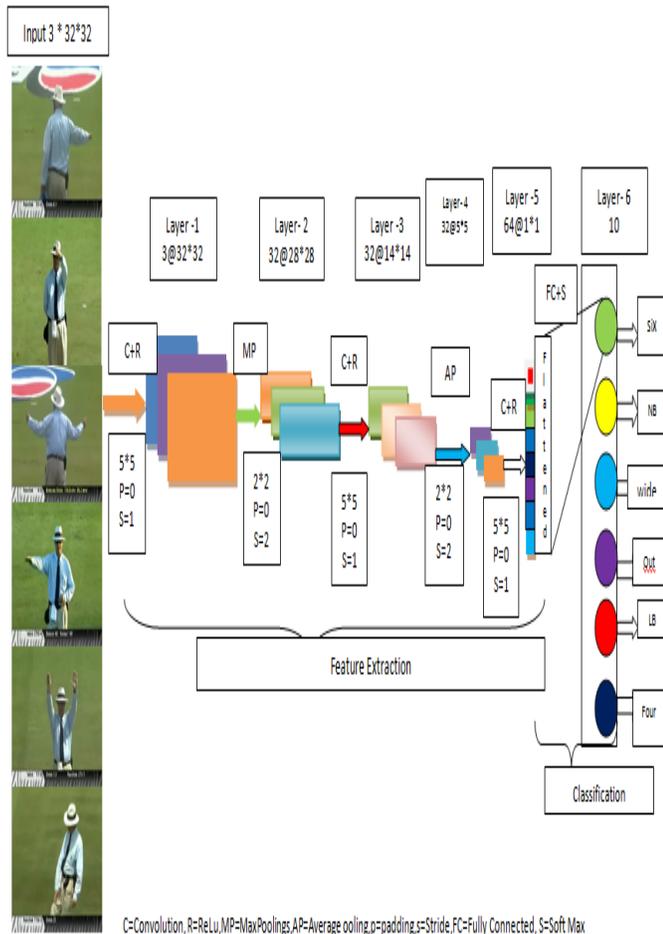


Fig. 3. Framework of Convolutional Neural Networks.

Convolutional Neural Network has four layers. They are convolution, pooling, flattening and fully connected layer. We assume a gray scale or RGB image 'I' represented by size  $n_1 * n_2$  denoted by function given in Eq. (1).

$$I: \{1, \dots, n_1\} * \{1, \dots, n_2\} \rightarrow W \in \mathbb{R}, (i, j) \rightarrow I_{i,j} \quad (1)$$

Given the filter  $K \in \mathbb{R}^{2h_1 + 1 * 2h_2 + 1}$ , the discrete convolution of the image I with filter K is given by Eq. (2)

$$I \times K = \sum_{u=-h_1}^{h_1} \sum_{v=-h_2}^{h_2} K_{u,v} I_{r+u,s+v} \quad (2)$$

Where the filter K is given by Eq. (3).

$$K = \begin{matrix} K_{-h_1, -h_2} & \dots & K_{-h_1, h_2} \\ \dots & K_{0,0} & \dots \\ K_{h_1, -h_2} & \dots & K_{h_1, h_2} \end{matrix} \quad (3)$$

A filter which is most familiar for smoothing is the discrete Gaussian filter  $K_G(\sigma)$  which is given by Eq. (4).

$$K_G(\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{r^2+s^2}{2\sigma^2}\right) \quad (4)$$

Convolution Layer: The small window with width and height spatially convolves through the input volume and dot products are computed between the entries of the filter during the forward pass. Assume layer 1 be a Convolutional layer. The input to this layer constitutes  $m_1$  ( $I-1$ ) feature maps from the preceding layer with  $m_2$  ( $I-1$ ) \*  $m_3$  ( $I-1$ ) size. When  $I=1$ , image I is an input consisting of one or more mediums. This is how raw images are fed as input to the convolution neural network. The output from the layer 1 constitutes  $m_1$  ( $I$ ) feature maps  $m_2$  ( $I$ ) \*  $m_3$  ( $I$ ) sized Eq. (5) is used to compute the feature map ( $i^{th}$ ) of layer 1.

$$Y_i^{(1)} = B_i^{(1)} + \sum_{j=1}^{m_1} m_1^{(i-1)} K_{i,j}^{(1)} * Y_j^{(i-1)} \quad (5)$$

Where, B is a bias and K is the filter of size  $2h_1+1 * 2h_2+1$  coupling the  $j^{th}$  activation map in layer ( $I-1$ ) with the  $i^{th}$  activation map in layer 1.  $m_2$  ( $I$ ) and  $m_3$  ( $I$ ) are determined by border influences. If layer 1 is not linear,  $m_1$  ( $I$ ) activation maps are the inputs to it and  $m_1$  ( $I$ ) =  $m_1$  ( $I-1$ ) activation maps are the outputs, each of size  $m_2$  ( $I$ ) \*  $m_3$  ( $I$ ) along with  $m_3$  ( $I-1$ ), given by Eq. (6).

$$Y_i^{(1)} = f(Y_i^{(i-1)}) \quad (6)$$

Pooling layer: Pooling is nothing but a down-sampling with non-linearity. The output of pooling is maximum value of such sub-region and reduces size of representation, memory, computation in network and controls the over fitting.

Flattening Layer: Flatten our pooled feature map into a column and insert the data into Artificial Neural Network.

Fully Connected Layer: It takes the output from several convolution and max pooling layers and performs affine transformation with matrix multiplication and bias offset to compute its activations. The output of this layer is calculated using Eq. (7).

$$Y_i^{(1)} = f(Z_i^{(1)}) \quad (7)$$

C. Model Implementation

The provided training data set was being used to train the CNN model. Initially the image features are extracted from each umpire action training images and model creates feature map- knowledge base - for testing. If the image in the frame was categorized as being one of the six classifications as : SIX, NO BALL, WIDE, OUT, LEG BYE or FOUR, then the processed frame is accumulated in one of the six classes and further used for generating the Event video summary of Cricket Sport. Algorithm 1 depicts the way in which the model is created for the purpose.

**Algorithm 1 : Proposed work Model**

**Input:** Upload dataset- SNWOLF-consisting of 80% images for training purpose

**Output:** trained CNN model

**Begin**

**Step1:** Data Pre-processing

**Step2:** Input Layer

This step reshapes the image into 3\*32\*32 pixels

**Step3:** Convolution and Pooling layers

P=32

For i = 1 to 4 do

1.Add six convolution layer with P feature maps of size 3\*3 and a rectifier activation function. Also dropout layer at 20% in between these six layers

2. Add 2\*2 is the greatest pool layer size and stride 3

3. P=P\*2

End

**Step4:** Flatten Layer

It is going to flatten our pooled feature map into a column. Set dropout layer at units and a 20%

**Step5:** Dense layer:

Create a 512-unit completely linked layer with a rectifier activation function.

Also set dropout layer at 20%

**Step6:** layer output

With 50 units and a softmax activation mechanism, create a fully connected output layer.

**Step7:** logic layer

The final step is the prediction.

**End**

IV. RESULTS AND EVALUATION OF SNWOLF DATASET

After the model has been trained, we tried to achieve almost 100% performance using the validity set. Fig. 4 displays the accuracy of instruction and verification progressively improves with the range maximum iteration. Our system was able to achieve an exactness of 98.2% which demonstrates classification ability on the practice set. Fig. 5

depicts the model's confusion matrix after being put through its paces with the test set. A matrix of uncertainty was used to measure the Accuracy, Precision and Recall of each class. The model's macro was then used to determine the F1 score average of the fit rate score and the recall score. Table I provides a summary of the model evaluation.

Table II shows the summary result of the classification assessment of the model on the practice set (SNWOLF).

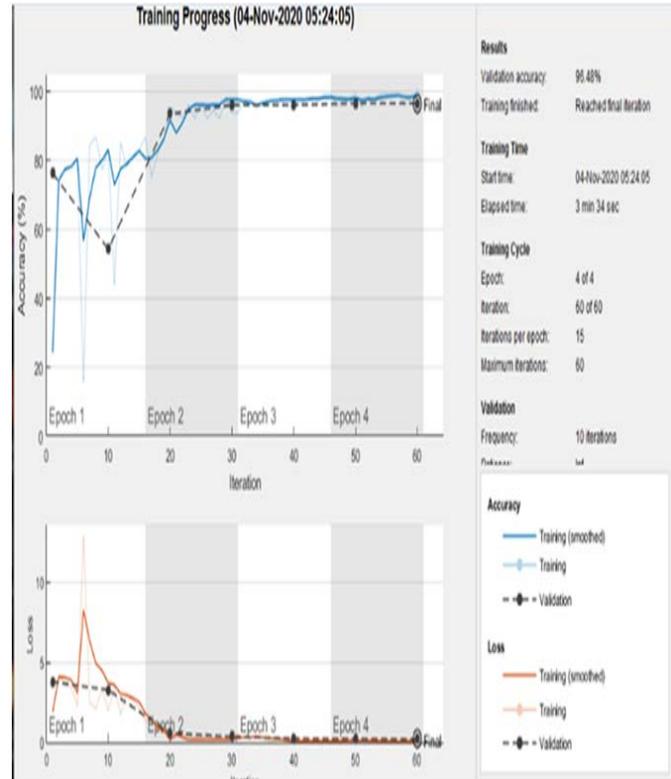


Fig. 4. Training and Validation Accuracy.

TABLE II. SUMMARY RESULT OF THE CLASSIFICATION ASSESSMENT OF THE MODEL ON THE PRACTICE SET –SNWOLF

Event Name	Accurac y for individu al event	Error rate for individu al Event	Precisi on	Reca ll	F1 scor e	Accura cy
1.Action.No Ball	83.3 %	16.7%	83	52	92%	98.20%
2.Action.Wid e	98.5%	1.5%	97	96		
3.Action.Out	97.9%	2.1%	97	99		
4.Action.Fou r	98.7%	1.3%	98	100		
5.Action.Six	100%	0%	100	100		
6.Action.Leg Bye	98.2%	1.8%	100	91		

### Results of CNN Classification

		Confusion Matrix						
Output Class								
	01	02.1Action	02.2Action	02.3Action	02.4Action	02.5Action	02.6Action	
01	114 5.1%	2 0.1%	0 0.0%	1 0.0%	8 0.4%	0 0.0%	0 0.0%	91.2% 8.8%
02.1Action	2 0.1%	10 0.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	83.3% 16.7%
02.2Action	2 0.1%	0 0.0%	131 5.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	98.5% 1.5%
02.3Action	4 0.2%	0 0.0%	0 0.0%	191 8.6%	0 0.0%	0 0.0%	0 0.0%	97.9% 2.1%
02.4Action	8 0.4%	7 0.3%	5 0.2%	1 0.0%	1687 75.8%	0 0.0%	1 0.0%	98.7% 1.3%
02.5Action	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	41 1.8%	0 0.0%	100% 0.0%
02.6Action	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	11 0.5%	100% 0.0%
	87.7% 12.3%	52.6% 47.4%	96.3% 3.7%	99.0% 1.0%	99.5% 0.5%	100% 0.0%	91.7% 8.3%	98.2% 1.8%
		Target Class						
		01	02.1Action	02.2Action	02.3Action	02.4Action	02.5Action	02.6Action

Fig. 5. Confusion Matrix.

### V. CONCLUSION AND FUTURE WORK

A CNN deep feature model is proposed in this research that can identify and classify six unique cricket EVENTS (SIX, NO BALL, WIDE, OUT, LEG-BY, FOUR) based on the Umpire Action Frame of the dataset. Using SNWOLF as a dataset, we trained the model on 80% of the images to identify and classify six events related to a cricket match between the two countries. Our model, shown in Fig. 6, test set was 98.2% accurate and the F1 score was 92.0%, which worked very well. The sample outputs of six classes Gestures are shown in Fig. 7. We intend to broaden our scope of study in the future to cover all forms of cricket competition country events where different types of cricket matches are played.

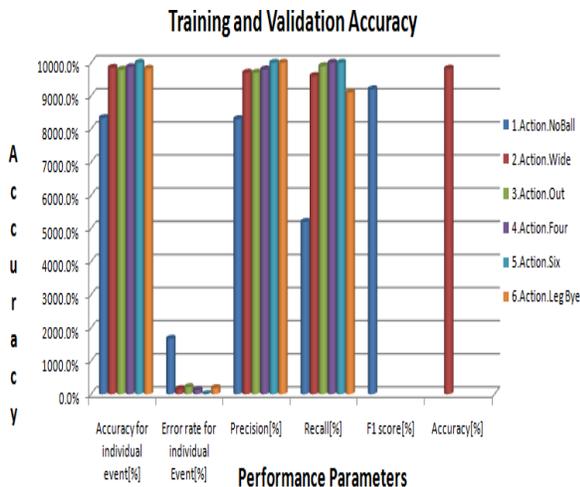


Fig. 6. Classification Model Evaluation Results on Accuracy, Error, Precision, Recall, F1 Score and Accuracy.



Fig. 7. Sample Images of Output Classes Six, No Ball, Wide, Out, Leg Bye and Four Events.

We intend to broaden our scope of study in the future for detection of other events, like Ball Catch, Baler Detection, Boundary, Third Umpire Decision, etc. will be suggested. Further covering all forms of cricket competition country events where different types of cricket matches are played. In this work open problems identified for future work are Text Tagging for events detected and Audio Annotation with Commentary.

### REFERENCES

- [1] S. Mitra and T. Acharya, "Gesture recognition: A survey, IEEE Transactions on Systems, Man, and Cybernetics", Part C (Applications and Reviews), vol. 37, no. 3, pp. 311-324, 2007. DOI: 10.1109/TSMCC.2007.893280.
- [2] Top10 list of the internet world's most popular sports,http://www.topendsports.com/world/lists/popular-sport/fans.htm.
- [3] K.Dixit and A.Balakrishnan, Deep learning using CNNs for ball-by-ball outcome classification in sports, Report on the course: Convolutional Neural Networks for Visual Recognition, Stanford University, 2016.
- [4] Batra et al."an automated multi-dimensional visual system which detected no-balls bowled in a cricket match". DOI:10.1109/ICACCI.2014.6968378.
- [5] R.Hari and M.Wilsey, Event detection in cricket videos using intensity projection profile of umpire gestures, in 20/4 Annual IEEE India Conference (INDICAN). IEEE, 2014, pp.1-6. DOI: 10.1109/INDICON.2014.7030519.
- [6] A.E.Chowdhury,M.S. Rahim, and M.A.U. Rahman, Application of computer vision in cricket: Foot over step no-ball detection, in 2016 3rd International Conference on Electrical Engineering and information Communication Technology(ICEEICT). IEEE, 2016, pp.1-5. DOI: 10.1109/CEEICT.2016.7873086.
- [7] M.Lazarescu,S.Venkatesh,and G.West, Classifying and learning cricket shots using camera motion, in Australasian Joint Conference on Artificial Intelligence. Springer, 1999, pp.13-23. DOI: 10.1007/3-540-46695-9\_2.
- [8] D. Karmaker, A. Chowdhury, M.Miah, M.Imran, and M.Rahman, "Cricket shot classification using motion vector", in 2015 2nd International Conference on Computing Technology and Information Management (ICCTIM). IEEE, 2015, pp.125-129. DOI: 10.1109/ICCTIM.2015.7224605.
- [9] A.Semwal,D. Mishra,V.Raj, J.Sharma, and A.Mittal, Cricket shot detection from videos, in 2018 9th International Conference on Computing, Communication and Networking Technologies(ICCCNT). IEEE,2018,pp.1-6. DOI: 10.1109/ICCCNT.2018.8494081.
- [10] C.J.Burges, A tutorial on support vector machines for pattern recognition, Data mining and knowledge discovery, vol.2,no.2,pp.121-167,1998. doi.org/10.1023/A:1009715923555.

- [11] James Wang, DEEP LEARNING: An Artificial Intelligence Revolution, A white ARK Invest, 2011, pp.1-41.
- [12] V.N.Pawar and S.N.Talbar, Machine learning approach for object recognition, International Journal of Modeling and Optimization, Vol.2, No.5, pp.622-628, October 2012. DOI:10.7763/IJMO.2012.V2.196.
- [13] Y.LeCun,K.Kavukvuoglu, and C. Farabet., Convolutional Networks and Applications in Vision. In Circuits and Systems, International Symposium on, pages 253-256, 2010. DOI: 10.1109/ISCAS.2010.5537907.
- [14] Simonyan, K., Zisserman, A.: Very deep Convolutional Networks for large-scale image recognition. arXiv:1409.1556 (2014).
- [15] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016). DOI:10.1109/CVPR.2016.90.
- [16] Deng, L.Y., Liu, Y.: Semantic analysis and video event mining in sports video. In: 22nd International Conference on Advanced Information Networking and Applications-Workshops (aina workshops2008), Okinawa,pp.1517–1522(2008). <https://doi.org/10.1109/WAINA.2008.167>.
- [17] Ravi, A., Venugopal, H., Paul, S., Tizhoosh, H.R.: A Dataset and Preliminary Results for Umpire Pose Detection Using SVM Classification of Deep Features. ArXiv: 1809.06217 (2018).
- [18] Islam, M.N.A., Hassan, T.B., Khan, S.K.: A CNN-based approach to classify cricket bowlers based on their bowling actions. arXiv:1909.01228 (2019).
- [19] Minhas, R.A., Javed, A., Irtaza, A., Mahmood, M.T., Joo, Y.B.: Shot classification of field sports videos using alexnet Convolutional neural network. Appl. Sci. 9(3), 483 (2019).
- [20] Russakovsky, O., et al.: Image net large scale visual recognition challenge. Int. J. Comput. Vis. 115(3), 211–252 (2015).
- [21] Different activities in cricket. 2015, [Online].Available: <http://www.open.ac.uk/ouclub/main/six-sidecricket/umpire-signals>.
- [22] Rafiq, M., Rafiq, G., Agyeman, R., Jin, S.I., Choi, G.S.: Scene classification for sports video summarization using transfer learning, Sensors 20(6), 1702 (2020), <https://doi.org/10.3390/s20061702>.
- [23] Karen Simonyan and Andrew Zisserman, Very deep Convolutional networks for large-scale image recognition, arXiv preprint arXiv: 1409.1556, 2014.
- [24] Jing Yu, Hang Li, Shou-Lin Yin and Shahid Karim :Dynamic Gesture Recognition Based on Deep Learning in Human-to-Computer Interfaces, 2020 - Volume 23 Volume 23, Issue 1 01 March 2020 Reach: 769 ,DOI.org/10.6180/jase.202003\_23(1).0004.