

# VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Belagavi – 590 018.



## PROJECT PHASE – I REPORT

on

**“AI CRICKET SCORE”**

*Submitted in partial fulfillment of the requirement for the curriculum of the 7<sup>th</sup> Semester*

**Bachelor of Engineering**

*in*

**Computer Science and Engineering**

*by*

<b>P RADHA RAI</b>	<b>:</b>	<b>1VI19CS066</b>
<b>PRAKHAR KUMAR CHADRAKER</b>	<b>:</b>	<b>1VI19CS070</b>
<b>SHAIK NOWSHEEN</b>	<b>:</b>	<b>1VI19CS098</b>
<b>SHANTANU KUMAR SINGH</b>	<b>:</b>	<b>1VI19CS100</b>

*Under the supervision of*

**Ms. VEENA G**  
**Assistant Professor**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**VEMANA INSTITUTE OF TECHNOLOGY**

**BENGALURU – 560 034**

**2022 - 23**

Karnataka ReddyJana Sangha®  
**VEMANA INSTITUTE OF TECHNOLOGY**  
Koramangala, Bengaluru-34.  
(Affiliated to Visvesvaraya Technological University, Belagavi)



**Department of Computer Science and Engineering**

**Certificate**

This is to certify that the project (phase – I) entitled “**AI CRICKET SCORE**” is a bonafide work carried out jointly by **P Radha Rai (1VI19CS066)**, **Prakhar Kumar Chandraker (1VI19CS070)**, **Shaik Nowsheen (1VI19CS098)** and **Shantanu Kumar Singh (1VI19CS100)**, during the academic year 2022-23 in partial fulfillment of the requirement for the 7<sup>th</sup> Semester course work of the **Bachelor of Engineering in Computer Science and Engineering** of the **Visvesvaraya Technological University, Belagavi**. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report. The project phase - I report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the said degree.

**Guide**  
Ms.Veena G

**HOD**  
Dr. M. Ramakrishna

**Principal**  
Dr. Vijayasimha Reddy. B. G

## ACKNOWLEDGEMENT

We sincerely thank **Visvesvaraya Technological University** for providing a platform to do the project phase - I.

Firstly, we would like to express our deep sense of gratitude to our institute “**Vemana Institute of Technology**” that provided us an opportunity to do a project phase - I entitled “**AI CRICKET SCORE**”.

We thank **Dr. Vijayasimha Reddy. B. G**, Principal, Vemana Institute of Technology, Bengaluru for providing the necessary support.

We would like to place on record our regards to **Dr. M. Ramakrishna**, Professor and Head, Department of Computer Science and Engineering for his continued support.

We would like to thank our project coordinators **Mrs. Mary Vidya John**, Assistant Professor and **Mrs. J Brundha Elci**, Assistant Professor, Dept. of CSE for their support and coordination.

We would like to thank our project guide **Ms. Veena G**, Assistant Professor, Dept. of CSE for her continuous support and valuable guidance towards successful completion of the project phase – I.

We would be failing in our duty if we do not thank the faculty members, batch mates, lab staffs, technicians and family members for their constant support and guidance.

	<b>P RADHA RAI</b>	<b>(1VI19CS066)</b>
<b>Date:</b>	<b>PRAKHAR KUMAR CHANDRAKER</b>	<b>(1VI19CS070)</b>
<b>Place: Bengaluru</b>	<b>SHAIK NOWSHEEN</b>	<b>(1VI19CS098)</b>
	<b>SHANTANU KUMAR SINGH</b>	<b>(1VI19CS100)</b>

## **ABSTRACT**

Gesture Recognition pertains to recognizing meaningful expressions of motion by a human, involving the hands, arms, face, head, and/or body. It is of utmost importance in designing an intelligent and efficient human–computer interface. The applications of gesture recognition are manifold, ranging from sign language through medical rehabilitation, monitoring patients or elder people, surveillance systems, sports gesture analysis, human behavior analysis etc., to virtual reality. In recent years, there has been increased interest in video summarization and automatic sports highlights generation in the game of Cricket. In Cricket, the Umpire has the authority to make important decisions about events on the field. The Umpire signals important events using unique hand on signals and gestures. The primary intention of our work is to design and develop a new robust method for Umpire Action and Non-Action Gesture Identification and Recognition based on the Umpire Segmentation and the proposed Histogram Oriented Gradient (HOG) feature Extraction oriented Non-Linear Support Vector Machine (NL-SVM) classification of Deep Features. Primarily the 80% of Umpire action and non-action images in a cricket match, about 1, 93, 000 frames, the Histogram of Oriented Gradient Deep Features are calculated and trained the system having six gestures of Umpire pose.

Key words: Python, Artificial Intelligence, Jupyter, Cricket Score, Sign Language Recognition, Umpire, Histogram of Oriented Gradients, Non-Linear SVM classifier

# TABLE OF CONTENTS

Chapter No.	Title	Page No.
	<b>Acknowledgement</b>	<b>i</b>
	<b>Abstract</b>	<b>ii</b>
	<b>List of Figures</b>	<b>v</b>
	<b>List of Tables</b>	<b>vi</b>
	<b>List of Abbreviations</b>	<b>vii</b>
<b>1.</b>	<b>INTRODUCTION</b>	<b>1</b>
	1.1 Scope	2
	1.2 Objective	2
	1.2.1 Plan of action for the project	2
	1.2.2 Current status of project	3
	1.2.3 Proposed plan for completion	3
	1.2.4 Outline of the chapters	3
<b>2.</b>	<b>LITERATURE SURVEY</b>	<b>4</b>
	2.1 Comparative Analysis	24
<b>3</b>	<b>SYSTEM REQUIREMENTS</b>	<b>26</b>
	3.1 Functional Requirements	26
	3.2 Non-Functional Requirements	26
	3.3 Hardware Requirements	27
	3.4 Software Requirements	27
<b>4.</b>	<b>DESIGN METHODOLOGY</b>	<b>28</b>
	4.1 System Architecture	28
	4.2 Dataflow Diagram	29

<b>5.</b>	<b>MODULE DESCRIPTION</b>	<b>31</b>
<b>6.</b>	<b>SUMMARY</b>	<b>33</b>
	<b>CONCLUSION</b>	<b>34</b>
	<b>REFERENCES</b>	<b>35</b>

## LIST OF FIGURES

<b>Fig. No.</b>	<b>Title</b>	<b>Page No.</b>
1.1	Timeline of the project work	2
4.1	System Architecture	28
4.2.1	Data Flow Diagram (Level 0)	29
4.2.1	Data Flow Diagram (Level 1)	30

## LIST OF TABLES

<b>Table. No.</b>	<b>Title</b>	<b>PageNo.</b>
2.1	Comparative Analysis	24



## LIST OF ABBREVIATIONS

AI	:	Artificial Intelligence
ANN	:	Artificial Neural Network
CbCr	:	Blue-Difference And Red-Difference Chroma Components
CNN	:	Convolutional Neural Network
GB	:	Giga Byte
KNN	:	K-Nearest Neighbor
ML	:	Machine Learning
RGB	:	Red Green Blue
SVM	:	Support Vector Machine

## CHAPTER 1

### INTRODUCTION

Gestures are known as expressive, meaningful body movement which involves physical motion of the fingers, hands, arms, head, face, or body with the intent of assigning meaningful information or interacting with the environment. In recent years gestures are widely used by humans to interact with computers and machines. Many common everyday equipment like TV, Smartphone, Car dashboard etc. can now be controlled by simple hand gestures. Gestures are also applied in many fields like developing aids for the hearing impaired, enabling very young children to interact with computers, recognizing sign language, medically monitoring patients emotional states or stress levels, lie detection, monitoring automobile drivers alertness/drowsiness levels, etc. Involvement of gesture recognition technology in sports will make the gameplay fairer and proficient. The gestures performed by the sports officials which indicate to what is going on in the game. Which also can provide something meaningful about a player, team, or the entire game. If the gestures of these officials are able to be recognized, meaningful information can be derived. We refer to a gesture as an intentional action by a person whereby part of the body is moved in a predefined way to indicate a specific event. Detecting these events enables automatic generation of highlights, contextual labeling of video and more importantly helps in decision making and automatic score update.

Cricket is the most popular sport in India. It is the 2<sup>nd</sup> most famous sports in Asia, 4<sup>th</sup> in Europe and also 2<sup>nd</sup> most famous sport in the whole world. But from the 19<sup>th</sup> century the same old manual method is being use to update the scoreboard, which is a great burden for the scorekeeper. The way of viewing the score has changed a lot in time, but the basic score updating process is still the same and performed by a person. So in the 21<sup>st</sup> century an automatic system is very much needed at this sector. Automatic systems are taking places of many boring manual task which was performed by humans 5 or 10 years ago. So developing a fully operational automated system like this is in dire need. Using modern equipment's and machine learning algorithms we can increase the accuracy of the decision and provide flawless result what the naked eye misses.

This motivates us to design a system which will recognize gesture of cricket umpire in real time. Besides the reason for choosing cricket is because it is the most popular sports in Bangladesh and unlike other popular sports it doesn't have enough technological application to make it fairer and more accurate. Gesture recognition has been explored in both vision and sensor based. In vision-based system some image processing or computer vision-based method

is used to recognize the gesture. On the other hand, in sensor-based system the person who perform the gesture wears sensor, and when the gesture is performed the sensor data taken and used to identify the gesture. In the area of sports and other sector, several attempts have been made for gesture recognition using both sensor-based and vision-based technique.[11]

## 1.1 Scope

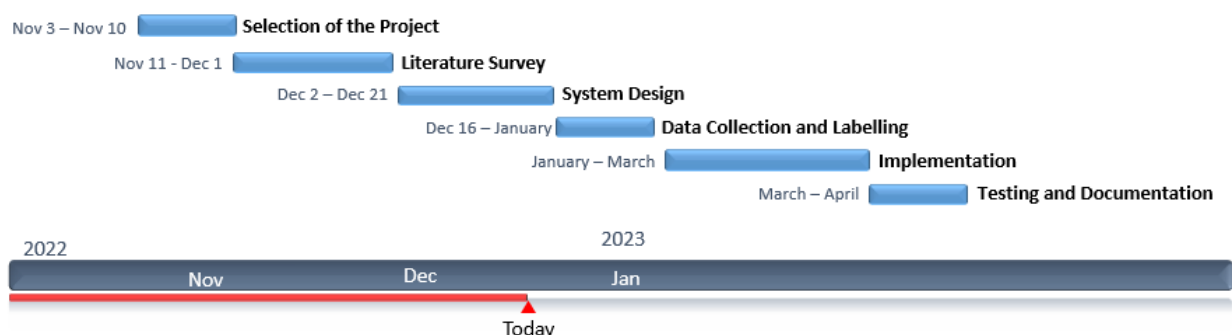
Technology can act as a flexible medium for hearing and speech impaired people to communicate amongst themselves and with other individual to enhance their education. Another expectation is to use the research outcome as learning tool of sign language where learner can practice. It can be implemented in all types of sports like football, tennis, baseball, hockey, kabaddi etc.

## 1.2 Objective

Main objective of the proposed system is to build a system to uniquely detect the Umpire using his dress code and then detect the gestures and postures done by him to Update the score. All the meaningful gestures by umpire can be detected and displayed in the score board.

### 1.2.1 Plan of action for the project

The below Figure 1.1 shows the completion of both the proposed and current plan of action for the project. On November 3<sup>rd</sup>, the project title and domain of the project were decided. Once domain of the project and title of the project were approved by the project guide and the project coordinator, literature survey papers regarding the project with similar motives were studied. The survey papers required for the project were shortlisted. These survey papers which were shortlisted suggested of what modules can be used for the project and the modules that are suitable were selected.



**Fig. 1.1: Timeline of the project work**

### 1.2.2 Current status of project

We have so far finished detailed literature survey on the given topic and currently working on understanding python language. Currently we are working on finding modules that can be used in this project to improve accuracy.

### 1.2.3 Proposed plan for completion

1. In the month of November, the domain of the project and title of the project was discussed and selected.
2. In the month of December, the literature survey was done based on title of our project was decided and we also decided the modules and the components to be used in our project.
3. In the month of January and February we will start learning advanced python.
4. In the month of March, we will start learning and implementing the modules.
5. In the month of April, we will Finalize code and final testing.

### 1.2.4 Outline of the chapters

**Chapter 2:** This chapter explains about literature survey which gives information which we gain from each papers.

**Chapter 3:** This chapter gives information about Hardware and software requirement, Functional and non-functional requirements.

**Chapter 4:** This chapter explains design methodology and work flow of the project.

**Chapter 5:** This chapter tells us about each modules to be used in the project.

**Chapter 6:** This chapter explains summary, conclusion and future work of the project.

## CHAPTER 2

# LITERATURE SURVEY

### LITERATURE SURVEY-1

**Title:** Low cost approach for Real Time Sign Language Recognition

**Author:** Matheesha Fernando, Janaka Wijayanayaka

**Methodology:**

Identify the signs and convert them into text and speech using appearance based approach with a low cost web camera. A series of image processing techniques with Hub-moment classification was identified as the best approach. Uses fast Convex Hull Algorithm for Binary image for pattern recognition, Histogram based classification. During the research available human computer interaction approaches in posture recognition were tested and evaluated. A series of image processing techniques with Hu-moment classification was identified as the best approach. The system is able to recognize selected Sign Language signs with the accuracy of 76% without a controlled background with small light adjustments. With the development of information technology new areas of human computer-interaction are emerging. There, human gesture plays a major role in the field of human computer interaction. As sign language is a collection of gestures and postures any effort in sign language recognition is in the field of human computer interaction. there are two types of approaches commonly used to interpret gestures for human computer interaction. First category, Data Glove based approach relies on electromechanical devices attached to a glove for digitizing hand and finger motions into multiparametric data. The major problem with that approach is it requires wearing the devices and will cause less natural behaviors. And also these devices are quite expensive. Appearance based method was selected for the research as main objective of the research is to identify a low cost method for sign language recognition. In appearance based method feature extraction and classification are the major components.

Feature extraction methods are used to reduce the number of dimensions of an image. A descriptor can be used for that. A descriptor describes an image and if properly used, image can be represented less with dimensions than the image itself. Also it can introduce some useful properties like scale and rotation invariance. Classification includes a broad range of decision theoretic approaches to the identification of images. It analyses the numerical

properties of various image features and organizes data in to categories. The classification represents the task of assigning a feature vector or a set of features to some predefined classes in order to recognize the hand gesture.

The research is focused on an application that can convert a video signal (processed as sequence of images) into a sequence of written words (text) and speech in real time. In the real-world, visual information could be very rich, noisy, and incomplete, due to changing illumination and dynamic backgrounds and obstacles, etc. Vision-based systems should be user independent and robust against all these factors. The suggested solution for the communication problem requires real-time facility, with effective as well as cost efficient techniques/algorithms. Therefore robustness, computational efficiency and uses Tolerance were the important challenges need to be considered here. In this approach the movement of the hand is recorded by a camera and the input video is decomposed into a set of features by taking individual frames into account. The video frames contain background pixels other than the hand, as the hand will never fill a perfect square. These pixels have to be removed as they contain arbitrary values that introduce noise into the process. The basic idea is to get the real time image and then extract the predefined features and then compare the feature vectors against the features extracted from stored template sign images. In the feature extraction phase what is most important is to get possible precise features as output. A very simple method is to compare each pixel location with each other and sum all the differences. This method is not realistic as it is not going to work on images that are not the same size or orientation. And also for the same hand posture there will be different images with small variations. Also it is computationally expensive for larger images. For an image of 100 by 100 pixels there are already 10,000 dimensions. Therefore features selected for classification are hand contour, orientation histogram, convex hull, convexity defects and moments.

**Advantage:** Helps in identifying a low cost, affordable method that can facilitate hearing and speech impaired people to communicate with the world in more comfortable way where they can easily get what they need from the society and also can contribute to the well-being of the society.

**Limitation:** This project only looks at the hand postures not on hand gestures.

## LITERATURE SURVEY-2

**Title:** Static Hand Gesture Recognition Based on Convolutional Neural Networks[2]

**Author:** Raimundo F. Pinto Jr. , Carlos D. B. Borges, Antonio M. A. Almeida ,and Ialis C. Paula Jr.

**Methodology:**

Proposes a gesture recognition method using convolutional neural networks. The procedure involves the application of morphological filters, contour generation, polygonal approximation, and segmentation during preprocessing, in which they contribute to a better feature extraction. Segmentation algorithms can be implemented to separate, colors, textures, points, lines, discontinuities, borders, among others. Training and testing are performed with different convolutional neural networks, compared with architectures known in the literature and with other known methodologies. All calculated metrics and convergence graphs obtained during training are analyzed and discussed to validate the robustness of the proposed method. The images are obtained from the database. The images go through an image processing stage, in which the following operations occur: color segmentation using an MLP network, morphological operations of erosion and closing, contour generation, and polygonal approximation, to remove image noise. After segmentation, binary images are obtained, so a logical AND operation is performed between these images and the originals, in order to preserve the information contained in the fingers and the surface of the hand. After these steps, the images are used to train a CNN and assess the performance of the technique with cross validation. Finally, the validation results are analyzed. Proposed architectures presented good results, with average rates of success of 96%. For CNN 1, with only two layers of convolution, it presented an accuracy rate of 94.7%, and for CNN 2, 3 and 4 accuracies remained above 96%. This possible due to the proposed image processing methodology, in which unnecessary information is removed, allowing improved feature extraction by the CNN. Proposed methodology and CNN architecture open the door to a future implementation of gesture recognition in embedded devices with hardware limitations.

One of the problems in gesture recognition is dealing with the image background and the noise often present in the regions of interest, such as the hand region. Use of neural networks for colour segmentation, followed by morphological operations and a polygonal approximation, presented excellent results as a way to separate the hand region from the background and to remove noise. This step is important because it removes image objects that are not relevant to the classification method, allowing the convolutional neural network to extract the most relevant gesture features through their convolution and pooling layers and,

therefore, to increase network accuracy. Proposal to make a logical AND operation with the segmentation masks and the original images provided the relevant information of the palms and fingers.

The proposed CNN architectures achieved high success rates at a relatively low computational cost. In addition, the proposed architectures reached accuracies very similar to the architectures already defined in the literature, although they are much simpler and have a lower computational cost. Is possible due to the proposed image processing methodology, in which unnecessary information is removed, allowing improved feature extraction by the CNN. It is clear that starting from 3 layers of convolution, together with pooling layers, modifying this type of neural network architecture does not increase the feature extraction and classification capacities of the network. Therefore, there is no significant increase in accuracy, but only in the computational cost of the network. After analysing the network convergence times during training, it is seen that increasing the number of convolution layers reduces the number of epochs necessary for the network to converge, thus extracting the data faster. CNN 1 converges at epoch 16, while CNN 3 converges at epoch 7. The architectures already defined in the literature presented accuracy rates with values up to 99%. However, they are more complex than the proposed architectures, some of them are more than 200 layers deep.

Thus, these architectures are also capable of extracting characteristics of the images more quickly, presenting the smallest numbers of epochs for convergence, as is the case of InceptionV3 and ResNet50, which converged at epochs 5 and 7. Results were also generated using the individual image datasets. It is possible to conclude that the CNNs are able to extract the features and classify the patterns well enough to reach correct answers close to 100%. Demonstrating robustness of the methodology independent of the base used.

**Advantage:** Proposed methodology are much simpler and have a lower computational cost.

**Limitation:** The proposed methodology approaches only cases of gestures present in static images.



### LITERATURE SURVEY-3

**Title:** Hand Gesture Recognition Systems with the Wearable Myo Armband [3]

**Author:** Engin Kaya, Tufan Kumbasar

**Methodology:**

The hand gesture recognition systems deal with identifying a given gesture performed by the hand. Utilized machine learning techniques to recognize the hand gestures. Seven different time domain features are extracted from the raw EMG signals using sliding window approach to get distinctive information. The performance of KNN, SVM and ANN algorithm will be compared. Then, the dimension of the feature matrix is reduced by using the principal component analysis to reduce the complexity of the deployed machine learning methods. The presented study includes the design, deployment and comparison of the machine learning algorithms that are k-nearest neighbor, support vector machines and artificial neural network. The results of the comparative comparison show that the support vector machines classifier based system results with the highest recognition rate.

The first step of the process is to extract meaningful and distinctive features from the raw data. Due to the complex and noisy characteristics of EMG, proper selection of features is essential for classification. In machine learning, working with directly raw data increases the complexity of implementation and processing cost. Also, some hidden patterns which cannot be visible in a single data point can be obtained in features. In this study, Authors have performed a sliding window method to record the data with windows length of 40 and sliding length of 20 as a pre-processing step. Then, the seven time domain features are calculated which are the Mean Absolute Value (MAV), variance, waveform length, Root Mean Square (RMS), Willison amplitude, Zero Crossing (ZC) and Slope Sign Change (SSC) of the signals. It can be firstly observed that increasing the number of neurons in the hidden layer has not significantly improved the recognition accuracy.

For the handled data set, an optimal number of hidden of neurons can be set as 12. In comparison of the training methods, the *trained* method resulted with lowest accuracy; whereas, *trainrp* and *trainscg* provided significantly better performances. Structure recognition systems with the wearable armband to classify the numbers from 0 to 9 in Turkish Sign Language (TSL). First step is to introduce the armband that has 8 electrodes/channel and handled data set. Then, to utilize machine learning techniques to recognize the hand gestures. In this context, a sliding window approach will be applied for signals of each channel of the armband and seven time domain features will be extracted from each obtained window.

Hence, the feature matrix will have 56 dimensions that will be then reduced by applying Principal Component Analysis (PCA) algorithm to 15 dimensions. Then, using the processed feature set, design and deploy the machine learning algorithms KNN, SVM and ANN as classifiers and compare their recognition performances.

In the comparative comparisons, the performance of the KNN algorithm will be examined for various distance metrics and number of neighbors while the SVM classifier will be tested for linear and radial basis function kernel methods. For the SVM, the performances of One Versus One (OVO) and One Versus All (OVA) binary classification methods will also be examined. The performance of the ANN for various numbers of neurons in hidden layer is analysed and three different training algorithm. In the light of these analyses, the tuning parameters of the classification methods is adjusted to optimal values to perform overall performance comparison in hand gesture recognition. The results will show that highest recognition rate will be obtained when the SVM classifier is deployed.

The first step of the process is to extract meaningful and distinctive features from the raw data. Due to the complex and noisy characteristics of EMG, proper selection of features is essential for classification. In machine learning, working with directly raw data increases the complexity of implementation and processing cost. Also, some hidden patterns which cannot be visible in a single data point can be obtained in features. In this study, Authors have performed a sliding window method to record the data with windows length of 40 and sliding length of 20 as a pre-processing step. Then, the seven time domain features are calculated which are the Mean Absolute Value (MAV), variance, waveform length, Root Mean Square (RMS), Willison amplitude, Zero Crossing (ZC) and Slope Sign Change (SSC) of the signals.

Dimension reduction is a process of reducing the number of variables in the feature vector set. In the classification problems, choosing the number of inputs has great importance to determine the time and space complexity of the classifier; therefore, working with less dimensional data provides simpler solution. In this study, the well-known technique PCA, has been employed to reduce the dimension of the feature matrix. It is observed that the first 15 principle components preserve about 96 percent of the variance; and thus reduced the feature matrix dimension.

**Advantage:** Achieve good accuracy

**Limitation:** Need to testing the proposed method with recording signals from different people and for more complicated hand gestures.

## LITERATURE SURVEY-4

**Title:**Gesture Recognition to Make Umpire Decisions [4]

**Author:** Lesha Bhansali ,Meera Narvekar

**Methodology:**

The Umpire gesture Recognition System aims squarely to introduce a more robust technology to show Umpire choices with the assistance of Gesture Recognition and trailing of hand movement of the Umpire. This technology helps to alleviate the burden of the scorekeepers. Authors tested the subsequent six gestures particularly OUT, SIX, NEWBALL, NO-BALL, DEAD\_BALL, FOUR. Edge detection algorithms which emphasize edges and transitions. The Umpire gesture Recognition System aims squarely to introduce a more robust technology to show Umpire choices with the assistance of Gesture Recognition and trailing of hand movement of the Umpire. This technology helps to alleviate the burden of the scorekeepers. It conjointly minimizes errors in displaying Umpire choices therefore adding to a more robust viewing expertise. The steps concerned during this method are as follows: Authors Implemented the gradient magnitude calculation. The aim is to outline wherever within the image the most important gradient magnitudes are present. Then, it'll be used to determine a threshold within the gradients so as to filter the concerned area of interest (hands, palms and legs) and to discard all the background.

They Created a gradient magnitude threshold that had to erase the lower levels gradients so as to keep the higher ones. This cut all the noise and regularized the background. Then, succeeding step was to calculate the geometric distance between the vectors of the various pictures analysed. This half was formed to check the various photos, by scrutiny of histograms. With this, area of unit is calculated to acknowledge the various gestures. This technique is termed Eigenfaces. It's a helpful applied mathematics technique that has found application in numerous fields (such as face recognition and image compression). Authors have conferred results on the recognition of Umpire's gestures during a cricket match. The novelty of their work is that they've had an inclination to use the recognition of gesture to display the output directly on screens. Gesture recognition is then performed among the device domain that avoids the problems associated with correct image segmentation. They have an inclination to use this approach to the popularity of various sports. Results show that this recognition system is capable of recognising a group of six umpire gestures from the game of cricket and performs best once using a feature set. Also in addition authors want to show the performance of segmenting gestures from a stream of continuous gestures by

selecting candidate gestures by the existence of movement. Then, to use mathematician filter to blur the image and have a homogeneous image. It allowed to get higher ends up in the gradient magnitude. The goal of this filter is to erase the background defects.

It's very vital to obtain a regular background to avoid noise. Authors created a gradient magnitude threshold that had to erase the lower levels gradients so as to keep the higher ones. This cut all the noise and regularized the background. Then, succeeding step was to calculate the geometric distance between the vectors of the various pictures analysed. This half was formed to check the various photos, by scrutiny of histograms. With this, they calculated area of unit to acknowledge the various gestures. The first side to put into thought is the style of underlying structure design. An important reason for choosing this framework is that it is user-friendly.

With the assistance of MATLAB authors have done it in a simple manner. The very last step was to check by calculating the geometric distance between the coefficients that area unit before every eigenvector. As the game of cricket has evolved over the decades, so has the technology behind the game. The system "Umpire Hand Gesture Recognition System"(UHGRS) created has been instrumental in increasing the accuracy of the decisions, provide flawless replays to help see what the naked eye misses.

The technology does not only benefit the players but is equally instrumental in providing a better viewing experience to the audience. Often while watching a cricket match live at a stadium, it is difficult to discern the umpire decisions due to poor visibility. The score-keepers have to be constantly alert as to when there is a fall of a wicket or a wide ball, dead ball, no ball etc. "To err is human". As humans are prone to error this technology come to their rescue.. If the scorekeeper accidentally misses to display the correct signals on the flat screen, the audience will definitely disapprove it. Thus, the "UHGRS" is a clever way to use technology and minimize the errors associated with displaying the correct umpire decisions.

**Advantage:** Capable of recognising a group of six umpire gestures from the game of cricket.

**Limitation:** No performance of segmenting gestures.

## LITERATURE SURVEY-5

**Title:** Automatic Labeling of Sports Video Using Umpire Gesture Recognition [5]

**Author:** Graeme S. Chambers, Svetha Venkatesh, and Geoff A.W. West.

**Methodology:**

Annotating sports videos Data from accelerometers is used to augment sports video. Umpires in the game wear wrist bands. A hierarchical hidden Markov model to solve the problem of automatic segmentation and robust gesture classification. The algorithm proceeds as follows: for each period of movement ahead in time (up to 10sec) of the start of a candidate gesture, calculate the likelihood of each model for that region. Gestures can be considered to exist at multiple levels in a hierarchy, where simple movements are grouped into more complex movements and complex movements are grouped into ordered sequences. The advantage of hierarchical modelling is this temporal decomposition of gestures.

The classification stage becomes more manageable for increasingly complex gestures as the dynamics of the gesture are explicitly encoded. Not only does grouping allow for segmenting a gesture into its subparts, hierarchical modelling allows new gestures to be learnt on-line by reusing subparts from already known gestures. Sports officials perform many gestures which are indicative of what is going on in the game. Their gestures can provide something meaningful about a player, a team, or the entire game. If the gestures of these officials are able to be recognised, meaningful information can be derived. Authors refer to a gesture as an intentional action where by part of the body is moved in a predefined way to indicate a specific event. Detecting these events enables automatic generation of highlights and more importantly, rich, contextual labelling of video.

To solve this problem Authors addressed the issues of segmenting continuous gesture data and performing robust gesture classification. Potentially more than one model will exceed the threshold for the filler model, however in this work, they simply took the maximum. The difference between the most likely model and the second most likely model is not taken into account.

A Gaussian distribution modelling the magnitude of gravity is used to detect periods of movement over a sliding window. The likelihood of the Gaussian is used to make a binary decision on whether the window contains movement. In some cases, when there is little acceleration, spurious responses to the Gaussian distribution can occur. For example, the sequence 1,1,0,1 (where 1 is movement and 0 is no movement), may result, however this is

not consistent labelling since there is a large degree of overlap (48/144). Thus, this sequence is replaced by the sequence 1,1,1,1. Similarly, if the sequence 0,0,1,0 occurs, it is replaced with the sequence 0,0,0,0. These two filters are ensuring that contiguous regions of data have consistent labelling over a sliding window. Once the regions of movement indicating candidate gestures are detected, the regions must be classified.

Since many sport umpire gestures contain pauses such as the cricket, there needs to be a method for grouping adjacent regions of identified movement including the pause periods. The problem is that a pause can be either a valid sub-gesture (as part of a gesture) or a pause between gestures. Authors approach to overcoming this is by using a conservative estimate for the maximum length of a gesture (10sec) and grouping all detected gestures and pauses in the window. This window is then iteratively reduced in length removing the last region of movement at each step. For example, the first gesture starting after time 50 would have four candidate regions for grouping (times 69 to 106, 69 to 86, 69 to 79, and 69 to 75). The algorithm proceeds as follows: for each period of movement ahead in time (up to 10sec) of the start of a candidate gesture, calculate the likelihood of each model for that region. After all candidate regions have been identified and their corresponding model likelihoods calculated, they are normalised by the filler ratio. The region and model corresponding to the maximum of the calculated ratios is considered the gesture for that region.

Using a filler model to compare different length observations for accurately finding segmentation endpoints is novel and has worked well on data and example domain. Without a filler model, different length observations can not be compared across models as the HMM likelihood function is non-linear.

**Advantage:** The system performs well overall with the exception of handling unknown movements which have similarities to known movements.

**Limitation:** Filler ratio requires further investigation for deciding when a known gesture occurs.

## LITERATURE SURVEY-6

**Title:** Vision-Based Sign Language Translation Device [6]

**Author:** Yellapu Madhuri, Anitha.G, Anburajan.M

**Methodology:**

This report presents a mobile Vision-Based Sign Language Translation Device for automatic translation of Indian sign language into speech in English to assist the hearing and/or speech impaired people to communicate with hearing people.

This system is broken down into three main parts

- a) Image acquisition
- b) Image processing to extract features for recognition
- c) Recognition stage where signs are identified and audio output is given.

Sign language is recognized using Lab View Software. The experienced lag time between the sign language and the translation is little because of parallel processing. This allows for almost instantaneous recognition from finger and hand movements to translation. This is able to recognize one handed sign representations of alphabets (A-Z) and numbers (0-9). The results are found to be highly consistent, reproducible, with fairly high precision and accuracy. The gestures of sign language are captured by the inbuilt camera to detect the movement of the hand. Capturing thirty frames per second (fps) is found to be sufficient. Higher fps will only lead to higher computation time of the computer as more input data to be processed. As the acquisition process runs at real time, this part of the process has to be efficient. Thus, previous frame that has been processed will be automatically deleted to free the limited memory space in the buffer. Image acquisition process is subjected to many environmental concerns such as the position of the camera, lighting sensitivity and background condition.

The camera is placed to focus on an area that can capture the maximum possible movement of the hand and take into account the difference in height of individual signers. Sufficient lighting is required to ensure that it is bright enough to be seen and analysed. In this work, a vision based sign language recognition system using LABVIEW for automatic sign language translation has been presented. This approach uses the feature vectors which include whole image frames containing all the aspects of the sign. This project has investigated the different issues of this new approach to sign language recognition to recognize on the hand sign language alphabets and numbers using appearance based features which are extracted directly from a video stream recorded with a conventional camera making recognition system



more practical. Although sign language contains many different aspects from manual and non-manual cues, the position, the orientation and the configuration or shape of the dominant hand of the signer conveys a large portion of the information of the signs.

Therefore, the geometric features which are extracted from the signers' dominant hand, improve the accuracy of the system to a great degree. Facial expressions are not focused, although it is well known that facial expressions convey important part of sign-languages. A wearable IOS phone system provides the greatest utility for automatic Sign Language to spoken English translator. It can be worn by the signer whenever communication with a non-signer might be necessary, such as for business or on vacation. Providing the signer with a self-contained and unobtrusive first-person view translation system is more feasible than trying to provide second-person translation systems for everyone whom the signer might encounter during the day. To increase the performance and accuracy of the Automatic Sign Language Translator (ASLT), the quality of the training database used should be enhanced to ensure that the ASLT picks up correct and significant characteristics in each individual sign and further improve the performance more efficiently. A larger dataset will also allow experimenting further on performance in different environments.

Such a comparison will allow to tangibly measuring the robustness of the system in changing environments and provide training examples for a wider variety of situations. Adaptive colour models and improved tracking could boost performance of the vision system. Current collaboration with Assistive Technology researchers and members of the Deaf community for continued design work is under progress. The gesture recognition technology is only one component of a larger system that one day be an active tool for the Deaf community.

This project did not focus on facial expressions although it is well known that facial expressions convey important part of sign-languages. The facial expressions can e.g. be extracted by tracking the signers' face. Then, the most discriminative features can be selected by employing a dimensionality reduction method and this cue could also be fused into the recognition system. This system can be implemented in many application areas examples include accessing government offices for filling out forms whereby no interpreter may be present to help.

**Advantage:** Translator between deaf and people who do not understand sign language.

**Limitation:** Doesn't focus on facial expressions.



## LITERATURE SURVEY-7

**Title:** RGB-H-CbCr Skin Color Model for Human Face Detection [7]

**Author:** Nusirwan Anwar bin Abdul Rahman, Kit Chong Wei and John See

**Methodology:**

This paper presents a novel skin color model, RGB-H-CbCr for the detection of human faces. Skin regions are extracted using a set of founding rules based on the skin color distribution obtained from a training set. The proposed scheme was also compared with the well-known AdaBoost face detector/classifier by Viola and Jones. The proposed scheme is able to reach comparable standards to that achieve by the AdaBoost algorithm (90.17%) on the similar data set. Face detection in colour images has also gained much attention in recent years. Colour is known to be a useful cue to extract skin regions, and it is only available in colour images. This allows easy face localisation of potential facial regions without any consideration of its texture and geometrical properties. In this paper, Authors have presented a novel skin colour model, RGB-H-CbCr to detect human faces. Skin region segmentation was performed using a combination of RGB, H and CbCr subspaces, which demonstrated evident discrimination between skin and non-skin regions.

The experimental results showed that new approach in modelling skin colour was able to achieve a good detection success rate. On a similar test data set, the performance of their approach was comparable to that of the AdaBoost face classifier. The RGB-H-CbCr skin color model is able to deal with various brightness and illumination conditions, but it remains susceptible to detection of non-skin objects that possess similar chrominance levels as skin colour.

The eccentricity property measures the ratio of the minor axis to major axis of a bounding ellipse. Eccentricity values of between 0.3 and 0.9 are estimated to be of good range for classifying face regions. Though this property works in a similar way as box ratio, it is more sensitive to the region shape and is able to consider various face rotations and poses. The proposed scheme was also compared with the well-known AdaBoost face detector/classifier by Viola and Jones and results showed that the proposed scheme (with the right configuration of morphological operators) is able to reach comparable standards to that achieve by the AdaBoost algorithm (90.17%) on the similar data set. To evaluate the effectiveness of the RGB-H-CbCr skin colour model, the face detection system was tested with various combination of colour models, each represented by its own set of bounding rules. The combination of all 3 subspaces resulted in the best DSR and lowest FDR values.

The proposed method sometimes failed to detect a face correctly, as seen from the high FDR of 28.29%. This could be attributed to the usage of morphological operators. Though these operators are used parallelly to improve the likelihood of detecting faces, it may sometimes cause “over-detection” of faces. In this paper, Authors have presented a novel skin colour model, RGB-H-CbCr to detect human faces. Skin region segmentation was performed using a combination of RGB, H and CbCr subspaces, which demonstrated evident discrimination between skin and non-skin regions. The experimental results showed that approach in modelling skin colour was able to achieve a good detection success rate. On a similar test data set, the performance of the approach was comparable to that of the AdaBoost face classifier.

Authors intend to refine the use of morphological operations in the post-processing of the extracted skin regions. An adaptive training (incremental learning) of the skin colour model can be used to improve the overall classification of skin regions. Primarily, the elimination of false detections and false dismissals is crucial to the success of a robust face detector. The next step of the face detection system involves the use of morphological operations to refine the skin regions extracted from the segmentation step. Firstly, fragmented sub-regions can be easily grouped together by applying simple dilation on the large regions. Hole and gaps within each region can also be closed by a flood fill operation. The problem of occlusion often occurs in the detection of faces in large groups of people. Even faces of close proximity may result in the detection of one single region due to the nature of pixel-based methods.

Hence, they used a morphological opening to “open up” or pull apart narrow, connected regions. Additional measures are also introduced to determine the likelihood of a skin region being a face region. Two region properties – box ratio and eccentricity are used to examine and classify the shape of each skin region. The box ratio property is simply defined as the width to height ratio of the region bounding box. By trial and error, the good range of values lie between 1.0 and 0.4. Ratio values above 1.0 would not suggest a face since human faces are oriented vertically with a longer height than width. Meanwhile, ratio values below 0.4 are found to misclassify arms, legs or other elongated objects as faces.

**Advantage:** Brightness and illumination conditions can be effectively dealt.

**Limitation:** Low success of a robust face detector.

## LITERATURE SURVEY-8

**Title:** A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition [8]

**Author:** Ali Moin, Andy Zhou, Abbas Rahimi , Alisha Menon.

**Methodology:**

Wearable devices that monitor muscle activity based on surface electromyography could be of use in the development of hand gesture recognition applications. Most devices with local processing cannot offer training and updating of the machine-learning model during use, resulting in suboptimal performance under practical conditions. The system can classify 13 hand gestures with 97.12% accuracy. A high accuracy (92.87%) is preserved on expanding to 21 gestures. A HD computing algorithm<sup>31</sup> for training and inference of hand gestures is used to provide a real-time analysis of physiological signals, wearable biosensors can implement machine-learning models for signal processing. Local (in-sensor) processing of signals from biosensors has advantages over wirelessly streaming raw data to an external computational device, including reduced communication link bandwidth and radio power requirements. To provide a real-time analysis of physiological signals, wearable biosensors can implement machine-learning models for signal processing. Local (in-sensor) processing of signals from biosensors has advantages over wirelessly streaming raw data to an external computational device, including reduced communication link bandwidth and radio power requirements.

Processing the signals locally can also offer improved latency and security. Machine learning models for in-sensor processing are, however, typically trained offline before they are implemented in low-power embedded processors. The 1,000-dimensional item memory elements were generated sequentially using a cellular automaton with a hardcoded seed for a smaller memory footprint. Subsequent processing steps were exactly as described in the previous section on HD classification architecture, replacing algebraic operations on bipolar hyper vectors with Boolean operations on binary vectors<sup>52</sup>. A shift register consisting of 21 hyper vectors was used as the AM to store trained prototype hyper vectors and search for the closest class during inference. Training examples for creating or updating a model or as queries for inference using a trained model . A prototype hyper vector for each class is formed by computing the class centroid. For binary and bipolar hyper vectors, this amounts to finding the majority of each element across all training examples. These prototype hyper vectors are then stored in an associative memory (AM), an entirely feedforward operation with a single pass over training data.

This is in contrast to other neuro-inspired approaches in which training often employs sophisticated, iterative frameworks and is much more computationally demanding than classification (for example, gradient descent with backpropagation<sup>46</sup>). Adding new classes to the model simply involves adding new prototype hyper vectors to the AM, again differentiating HD computing from other algorithms that may require full retraining or modifications to the architecture. Once all prototypes have been computed and stored, classification involves finding the nearest-neighbouring prototype to a query hyper vector.

Authors tuned model hyperparameters and validated the HD algorithm using the offline dataset before optimizing it for efficient implementation using hardware description language (HDL) for synthesis on the device's FPGA. Raw 15-bit analogue-to-digital converter codes were used as the input for feature extraction, with incremental calculations of the 50-sample MAV performed with each new sample. The implemented arithmetic operations consisted only of addition, two's complement inversion, and arithmetic right shift for division. Features were quantized and saturated to 6-bit integers based on analysis of the offline dataset, optimizing for the dynamic range (range divided by step size) given the arithmetic requirements. The 1,000-dimensional item memory elements were generated sequentially using a cellular automaton with a hardcoded seed for a smaller memory footprint.

Subsequent processing steps were exactly as described in the previous section on HD classification architecture, replacing algebraic operations on bipolar hyper vectors with Boolean operations on binary vectors<sup>52</sup>. A shift register consisting of 21 hyper vectors was used as the AM to store trained prototype hyper vectors and search for the closest class during inference. A single contextual update with 50% weighting was enabled for each AM entry by merging a predetermined set of 500 bits from a newly trained prototype into the stored one. The frequency contents of the different recordings were qualitatively similar. Spot SNR was calculated by comparing the power spectrum during a gesture performance to the power spectrum during rest.

Overall SNR (calculated as the integral of spot SNR) varied for different channels and different gestures. The Cometa and CapgMyo systems exhibited better peak SNR for the best channels and associated best gestures, probably because those systems consist of differential, bipolar recording configurations with improved common-mode noise rejection.

**Advantage:** Low-cost and low-complexity

**Limitation:** Classification accuracy is less.

## LITERATURE SURVEY-9

**Title:** A Dataset and Preliminary Results for Umpire Pose Detection Using SVM Classification of Deep Features [9]

**Author:** Aravind Ravi, Harshwin Venugopal, Sruthy Paul, Hamid R. Tizhoosh.

**Methodology:**

The umpire in cricket has the authority to make critical decisions about on-field events. The umpire communicates important events through distinct hand signals and gestures. They Determine four such events for classification: SIX, NO BALL, OUT and WIDE based on detecting the umpire's pose from the video frames from a cricket game. CNN algorithm is used, the early layers learn more generic features such as shapes, edges, and colour blobs, the deeper layers learn features more specific to the original dataset. Automatic video summarization has gained increased attention in the recent past. Sports highlights generation, movie trailer generation, automatic headlines generation for news are some examples of video summarization. Convolutional neural networks (CNNs) have outperformed most of the traditional computer vision algorithms for tasks such as image classification and object detection. A CNN is a combination of a feature extractor and a classifier. The convolutional layers of the CNN are the feature extractors. They learn the representations automatically from the input data. The early layers in the CNN learn more generic features such as shapes, edges, and colour blobs, while the deeper layers learn features more specific to that contained in the original dataset.

This work extends on the idea of identifying events in cricket videos based on detecting the pose of the umpire. In addition to this, they propose the SNOW dataset which comprises of five classes of umpire actions each corresponding to four events such as Six, No Ball, Out, Wide (SNOW) and a no action class is included. They have set the benchmark evaluation for image classification based on a linear SVM classifier trained on features extracted from pretrained networks. Automatic video summarization has gained increased attention in the recent past. Sports highlights generation, movie trailer generation, automatic headlines generation for news are some examples of video summarization. The focus of the present work is sports video summarization in the form of highlights.

The highlights of a game provide the summary of important events of that game such as a goal in soccer or a wicket in cricket. It is a challenging task to summarize the highlights from sports videos as these videos are unscripted in nature. An efficient approach can be based on identifying key events from the sports video and use them to automatically generate the highlights. Among sports, cricket is the most popular game in the world after soccer and

has the highest viewership rating. A detailed explanation of the game of cricket can be found .

In the game of cricket, the umpire is the person with the authority to make important decisions about events on the field. The umpire signals these events using hand signals, poses and gestures. This innate characteristic of the cricket video can be leveraged as one approach for solving the problem of cricket highlight generation. Therefore, a system can be developed to detect the unique signals and poses shown by the umpire to automatically generate cricket highlights. A method for umpire pose detection for generating cricket highlights based on transfer learning is proposed in this work. Authors explore the use of features extracted from the pre-trained networks such as Inception V3 and VGG19 networks pre-trained on ImageNet dataset. A linear support vector machine (SVM) classifier is trained on the extracted features for detecting the pose of the umpire.

A new dataset, SNOW, is introduced in this work and all experiments are performed on this dataset. The system built using this dataset is evaluated on cricket videos for highlights generation. The paper is organized as follows: In Section II, the back ground work covering existing techniques will be discussed. Section III introduces the proposed dataset. Section IV outlines the overall system design and methodology for evaluating the proposed dataset as a benchmark for umpire pose detection and highlight generation. The results of the experiments and their analysis are discussed in Section V. Section VI concludes the paper by providing directions for future work.

A benchmark database such as TREC Video Retrieval Evaluation (TRECVID) for general video indexing, summarization and retrieval has been used in many studies. In the domain of sports video annotation and summarization, several works. Automatic summarization of soccer videos has been proposed. Similar studies for sports such as basketball, baseball, and tennis have also been reported. More recent work, in the domain of sports video summarization, has been reported in for the game of cricket.

Prior studies have used event detection in cricket videos as the basis for highlight generation. Hari et al. have proposed a method based on intensity projection profile of umpire gestures for detecting events. A technique based on Bayesian belief networks for indexing broadcast sports video has been proposed. The use of sequential pattern mining to segment cricket videos into shots and identify the visual content is presented.

**Advantage:** The proposed system is an effective solution for the application of cricket highlights generation.

**Limitation:** Time complexity is high.

## LITERATURE SURVEY-10

**Title:** An Approach to Automate the Scorecard in Cricket with Computer Vision and Machine Learning [10]

**Author:** Md. Asif Shahjalal, Zubaer Ahmad, Rushrukh Rayan, Lamia Alam.

**Methodology:**

This form of sports is widely played in more than 125 countries recognized by the International Cricket Council. One of the most challenging issues that first initiates the discussion on its prosperity is the duration of the game. The on-field umpire has to authorize decisions almost after each delivery. Haar-Cascade classifier algorithm is used to classify a specific type of object. This process would eliminate the manual updating of scorecards and thereby reduce the game duration notably. In addition, it excludes the prerequisite of wearing special gloves involving sensors. The efficiency of the algorithm is then cross-checked with the training and test data. This proved to be a very simple but efficient algorithm for umpires gesture detection.

Involvement of gesture recognition technology in sports will make the gameplay fairer and proficient. The gestures performed by the sports officials which indicate to what is going on in the game. Which also can provide something meaningful about a player, a team, or the entire game. If the gestures of these officials are able to be recognized, meaningful information can be derived. Authors refer to a gesture as an intentional action by a person whereby part of the body is moved in a predefined way to indicate a specific event. Detecting these events enables automatic generation of highlights, contextual labelling of video and more importantly helps in decision making and automatic score update. In contrast to these, authors intended to design a vision-based system using logistic regression machine learning technique which can detect both static and dynamic hand gestures of a cricket umpire. Main goal of authors is trying to recognize the gestures an umpire performs in a match and update the scoreboard for the corresponding gesture accurately.

They trained a haar-cascade classifier to detect human wrists from the video stream from a static camera. Then the selected region of interest is continuously checked through the multiclass logistic regression model if a gesture is matched. Then accordingly the score is updated. Sensor-based hand gesture recognition started with the invention of glove-based system, which was further divided into two distinct categories- active data glove and passive data glove over the years. The first glove-based systems were designed as a part of a camera-based LED system to track body and limb position for real-time computer graphics animation



in the 1970s, and since then, a number of different designs have been proposed.

A low-cost version named the Power Glove, was commercialized by Mattel Intellivision as a control device for the Nintendo video game console in 1989 and became well known among video games players. Chambers et. al proposed a probabilistic hierarchical framework to extract gestures and significant events of Kung Fu martial art movements acted out by an instructor in a simulated training video using accelerometers and the Hidden Markov Model. In another work, Chambers et. al proposed the use of the Hierarchical Hidden Markov Model (HHMM) in conjunction with a filler model for segmenting and classifying gestures at differing levels of detail. In this work, sports video was augmented with accelerometer data from wrist band worn by umpires in the game. The gestures they recognized are: Dead Ball, Four, Last Hour, Leg-Bye, No Ball, One Short, Out, Penalty Runs, TV Replay and Wide.

Another popular way to recognize hand gesture is to use kineect. Gesture detection was successfully recognized using a feature vector and a real-time histogram based algorithm by Gomez et. al. Wang et. al recently developed a simple and inexpensive for 3D articulated user-input using the hands. Their approach uses a single camera to track a hand, wearing an ordinary cloth glove that is imprinted with a custom pattern. Bhansali et. al proposed a method of Gesture recognition to Make Umpire Decisions by using subtraction method and gradient method. They tested the subsequent six gestures particularly OUT, SIX, NEWBALL, NO-BALL, DEADBALL and FOUR.

Google is recently developing a system called project soli where they use special radar sensor to detect the fingers movement. Soli sensor technology works by emitting electromagnetic waves in a broad beam. Objects within the beam scatter this energy, reflecting some portion back towards the radar antenna. Properties of the reflected signal, such as energy, time delay, and frequency shift capture rich information about the objects characteristics and dynamics, including size, shape, orientation, material, distance, and velocity. In project soli there are some sensors are used but those are not attached with the gesture performers hand or body. So these can be called as a hybrid system where sensor and radar vision both are used simultaneously. Sensor-based systems had limited accuracy and were tethered to computers using cumbersome wiring. They were meant for very specific applications and were never commercialized.

**Advantage:** This proved to be a very simple but efficient algorithm for umpires gesture detection.

**Limitation:** Multiple classifiers were need to be trained in order to make it work.



**Table 2.1 Comparative Analysis**

Reference	Algorithm/ Technique	Platform used	Performance Metrics	Advantage	Drawback
[1]	Convex Hull Algorithm	Linux	Hand outline, Hand contour, SIFT, SURF	low cost, affordable method	Only looks at the hand posture not hand gesture
[2]	CNN Algorithm	Windows, Linux	Precision, Recall, F1 score	Much simpler and lower computational cost.	Cases gestures present in static images.
[3]	KNN,SVM and ANN Algorithm	Linux	Precision, Recall, F1 score	Achieve good accuracy	More complicate hand gestures.
[4]	Edge detection algorithms	Windows	Gesture Recognition	Recognizing a group of six umpire gestures.	No performance segmenting gestures.
[5]	Segmentation algorithm	Linux	Known, Unknown, Recall	System performs well.	Ratio requires further investigation
[6]	Lab View Software.	Windows, Linux,IOS ,Android.	Geometric Features of Hand	Translator between deaf and people who don't understand sign language.	Doesn't focus on facial expressions.
[7]	RGB-H-CbCr skin colour model, AdaBoost Algorithm	Windows	FDR, DSR	Brightness & illumination can be effecttively dealt.	Low success of a robust face detector.

[8]	Neuro-Inspired Hyperdimensional compute Algorithm	Linux	SNR	Low complexity	Mapping are not sparse.
[9]	CNN algorithm	Windows	Accuracy of Inception V3, VGG19-FC1	High detection accuracy.	Complex dataset.
[10]	Haar-Cascade classifier algorithm	Windows	GD,GR,SR	Efficiency enhanced to greater.	Devices expensive, cumbersome experience.

## CHAPTER 3

# SYSTEM REQUIREMENTS

### 3.1 Functional Requirements:

- A Collection of Dataset consisting of various gestures and sign of umpire.
- A layer that can Down sample the input along its spatial dimensions and flattens the input.
- A deep learning framework for processing and extracting the features of the dataset.
- A deep learning architecture for training and saving our model.
- A method for region proposal.
- A deep learning architecture for correctly detecting the input gesture of umpire.
- A display board for correctly displaying the recognized gestures.

### 3.2 Non-Functional Requirements:

- **Usability**

Usability here means the degree to which something is able or fit to be used. This Project is useful in Detecting the Umpire's Gesture and correctly guess the output of the gesture.

- **Reliability**

Reliability here means the probability of performing a specified function without failure under given conditions for a specified period of time. So the project should be reliable as to not cause error which can lead to false outcome.

- **Performance**

Performance here should be high i.e it should not take more time to compute the results of umpire's gestures. If the performance is low it may lead to loss of interest and excitement of the audience.

- **Portability**

This Project should be portable so that it can be used in various places without worrying about the shifting cost. Portability is a great functionality to have because in cricket the matches can be organized at different places.

**3.3 Hardware Requirements:**

- Processor : Intel i5
- Speed : 2.4 GHz
- RAM : 8 GB(minimum)
- Hard Disk : 160 GB

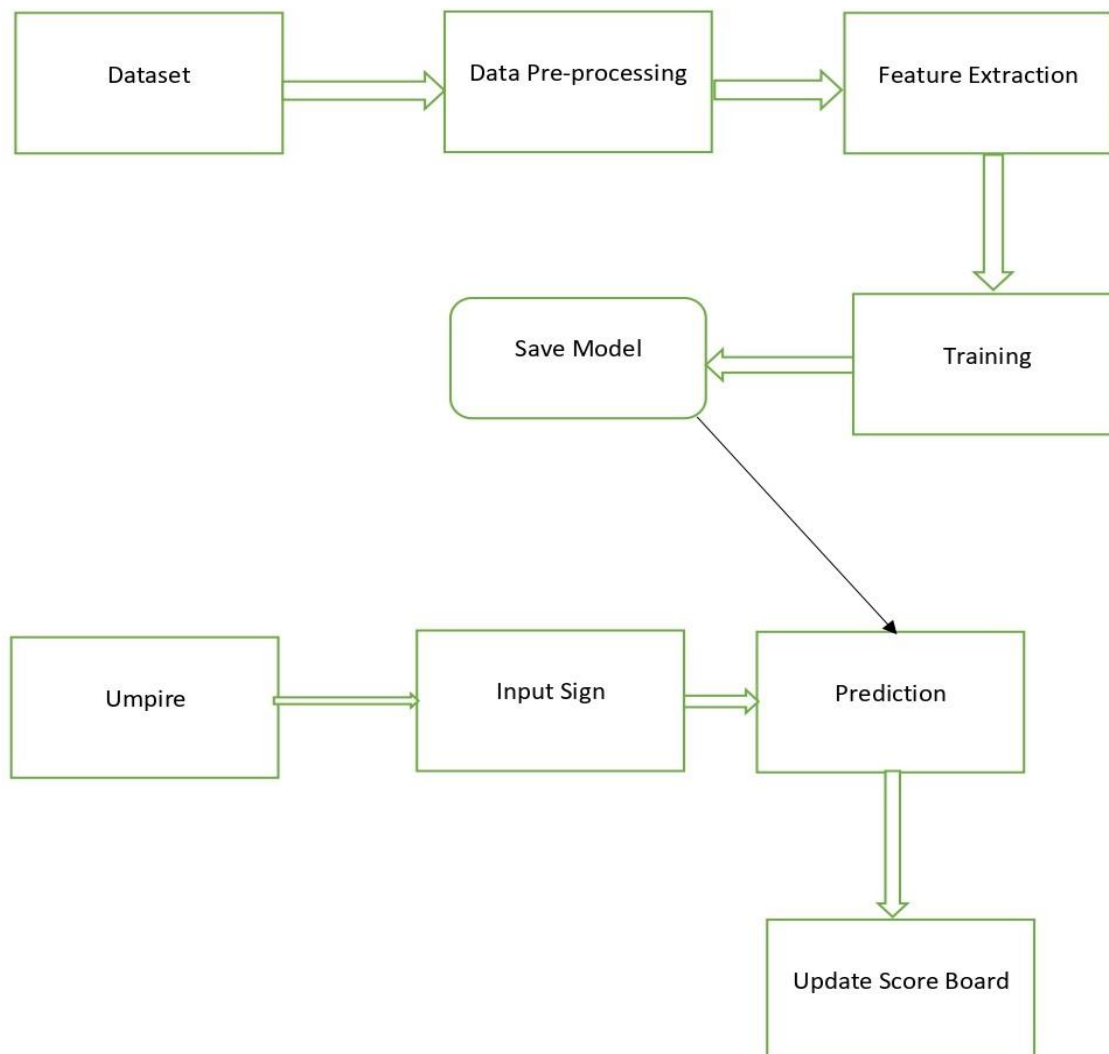
**3.4 Software Requirements:**

- Windows OS : Windows 10
- Language : Python
- Tools : Jupyter Notebook

## CHAPTER 4

# DESIGN METHODOLOGY

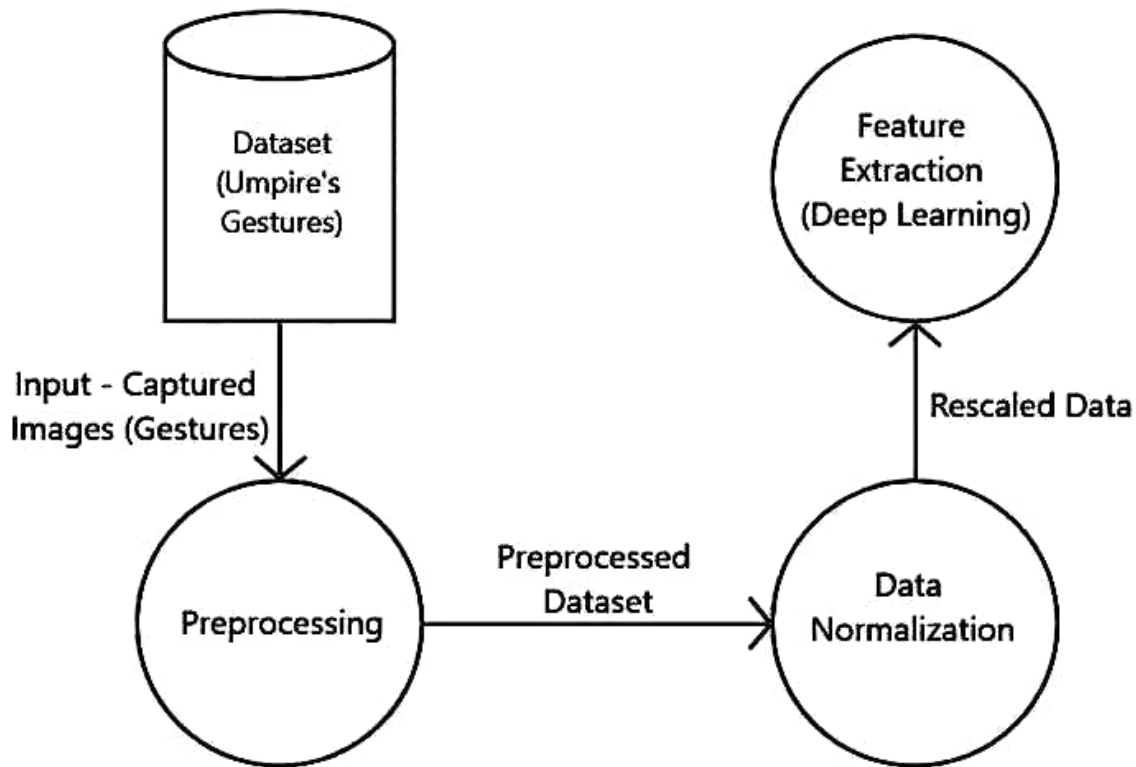
### 4.1 System Architecture



**Fig. 4.1 System Architecture**

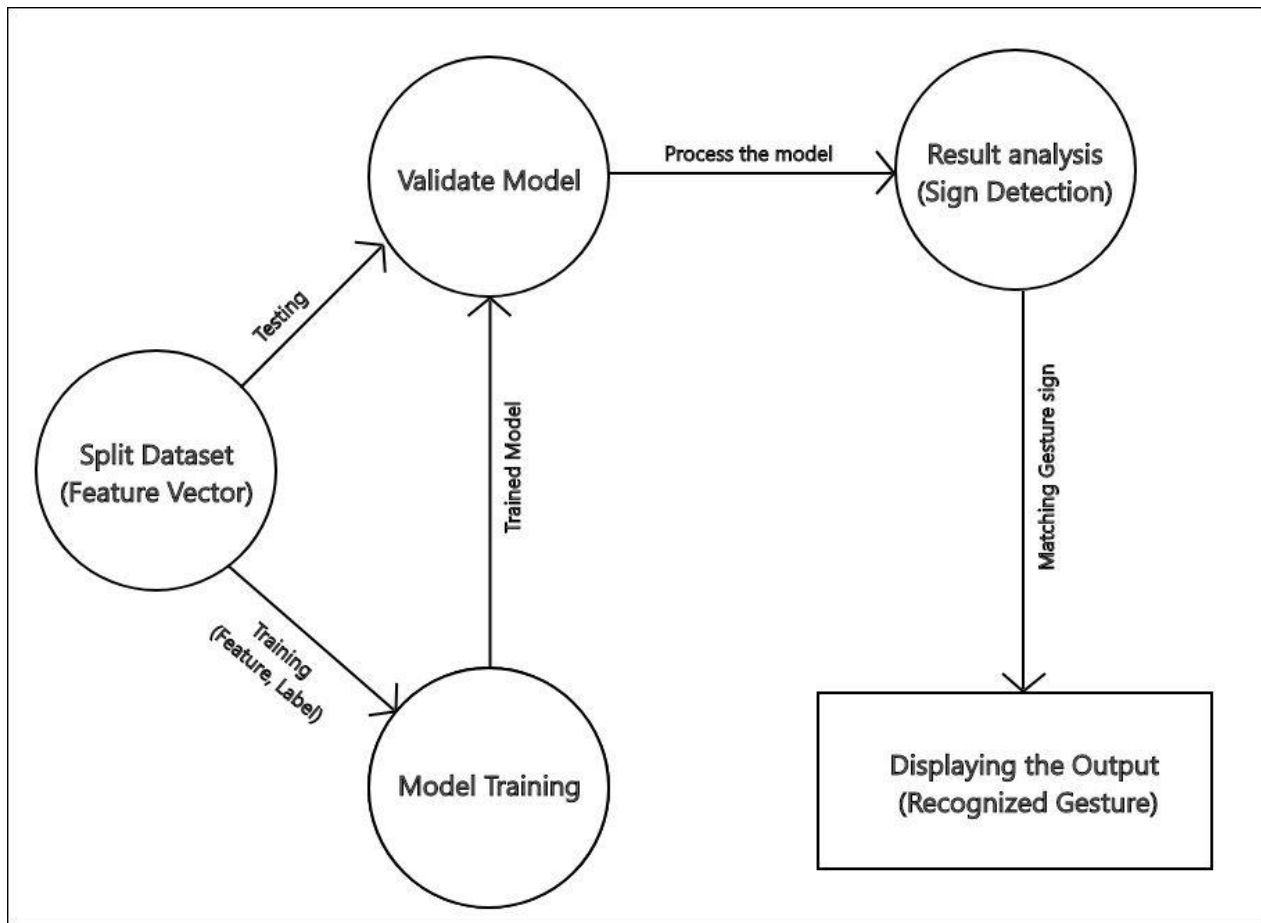
In Figure 4.1, A dataset is there which has umpire's gestures. Open CV is used to save the umpire signs in a particular directory. The captured dataset is now processed for data preprocessing where the noise is removed and passed for features extraction. The extracted features will be the input to the deep learning model to get trained. Once the model is trained, it is saved for future prediction. When the umpire shows the sign, captured image will be passed to the train model for prediction. Based on the predicted signs, the scoreboard will be updated.

## 4.2 Data Flow Diagram



**Fig 4.2.1 Data Flow Diagram (Level 0)**

In Figure 4.2.1, A Dataset which has images of the umpire hand signal are Preprocessed using OpenCV and after that data normalization is applied. Once the data is normalized feature extraction technique is used.



**Fig 4.2.2 Data Flow Diagram (Level 1)**

In Figure 4.2.2, The dataset is splitted into training and testing. The training dataset is passed to the deep learning algorithm for training. Then the trained model goes for validation where testing data is also sent. Once the model is Validated, it is processed for Detecting the signs. The matched gesture sign which are recognized is then displayed.

## CHAPTER 5

### MODULE DESCRIPTION

The proposed system consists are some modules of following steps to interpret the sign gestures using deep learning method. The working method consists of main stages. These are respectively; loading the data set, the design of the convolutional neural network, configuration of training options, training of the CNN object detector with hand gestures, evaluation of trained detector. These stages and conventional and methods will be discussed in this section.

#### 5.1 Region Proposal:

Various recent studies have provided methods to produce categorical independent zone recommendations. These methods have examples such as the object ness of image windows, selective Search for Object Recognition, category independent object proposals, object segmentation using constrained parametric min-cuts, Multi scale combinatorial grouping and so on. These methods establish cells by implementing convolution neural network with square cuts.

#### 5.2 CNN (Convolutional Neural Network) for Feature extraction:

In this study, a feature vector of size 4096 were extracted from each region proposal with Caffe deep learning framework. Features were calculated by forwarding the average output 227x227 red-green blue image with five convolution layers and two completely connected layers. In order to calculate an attribute in a region proposal, the image data is first converted to a form compatible with CNN. (In this study, fixed entrances of  $227 * 227$  pixels in size are used.). Then, the most simple of the possible transformations of the random-shaped regions was selected. Here, all the pixels in a tight bounding box around the candidate area are resolved unto the required size, regardless of the size or aspect ratio. Before dissolving, the tight bounding box was expanded to provide w pixels skewed picture content around the box at the skewed dimension (w = 16 was used). In addition, a simple bounding box regression was used to expand the localization performance within the application.

#### 5.3 CNN training:

CNN was trained on a large auxiliary data set (gestures classification) using only image-level additional tags. CNN was trained on data set (ImageNET) using only additional tags. This training was carried out using Caffe Deep Learning framework.



### **5.4 Object Category Classifiers:**

Here, binary classifier training was used to perceive gestures. It is a positive example of an image area in which a car is tightly enclosed. In a similar way, a background region that is not interested in cars is a negative example. It is unclear how a partially overlapping region of the car should be labeled. The unclear state is solved by specifying an overlap threshold value. Areas below this threshold value are identified as negative and those above the threshold value as positive. The overlap threshold “0.3” was chosen by conducting a grid search on the verification set. Once the features are removed and the training tags are applied, CNN is applied optimally to all classes.

### **5.5 Result (hand gesture sign detection and recognition):**

The proposed gesture detector with sign has been successfully trained by using Faster CNN deep learning methods on the sample gestures datasets and the sign detection process has been successfully performed by the trained sign detector being tested on the test data set. Different images were tested and found that the new technique of classification was found to show 97% accuracy. Some images tested with other database images are given in the results analysis. In Results analysis are real time detect in gesture sign with functions like alphabet, word, sentence and Indian sign recognize when live camera is start then capture the test images (gestures) that time compare the train model ‘CNN\_MODEL.H5’ and ‘labels.pkl’, if it is matching the dataset after the hand gestures sign process in display the result.

## CHAPTER 6

### SUMMARY

We surveyed total of 10 research papers related to our project and using the modules used in them we are planning to implement a Deep learning algorithm for detecting and extracting the features of a dataset which consists of different signs of umpire. There will be a dataset consisting of images of umpire making sign and we will extract the features of them and use them to train and test our model. We have also started learning about modules that can be used in our project which we found during our literature survey.

## CONCLUSION

We Started with the searching and finalizing of our project name and domain where we found that Machine learning and Artificial intelligence is a blooming industry and it is not yet implemented in Cricket so we finalized “AI Cricket Score” as our project. We intend to implement the detection of sign concept in the field of cricket. To make it happen we will have to capture the umpire’s signs in the field of cricket. We surveyed total of 10 Research Papers of different authors that have done projects in this field. We found that using web camera, we can capture the images and save in the respective directory of the sign. The images will then undergo for pre-processing techniques so that better quality of the images can be used for feature extraction. We also found about deep learning algorithms that can be implemented on the saved model for prediction. We have researched about modules that can be used in our project for high accuracy and we have started learning about them. Our work will also include data gathering using a web camera to increase the dataset size to more than 5000 RGB images making our prediction more robust. Process of real-time prediction using image frames from a web camera with rates of 50 to 100 Hz can also be used. Currently the aim of this project is limited to detecting Umpire’s gesture. Future Enhancement may include more cameras for detecting the umpire in 360 degree to improve gesture recognition and unique gestures for actions that do not have gestures can be added to reduce the workload on the review team. With enough funds this project can overtake and detect even the non gesture based actions like boundary detection for run out, four and six runs.

---

## REFERENCES

- [1] M. Fernando and J. Wijayanayaka, "Low cost approach for real time sign language recognition," *2013 IEEE 8th International Conference on Industrial and Information Systems*, 2013, pp. 637-642, doi: 10.1109/ICIInfS.2013.6732059.
- [2] M. Z. Islam, M. S. Hossain, R. ul Islam and K. Andersson, "Static Hand Gesture Recognition using Convolutional Neural Network with Data Augmentation," 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), 2019, pp. 324-329, doi: 10.1109/ICIEV.2019.8858563.
- [3] E. Kaya and T. Kumbasar, "Hand Gesture Recognition Systems with the Wearable Myo Armband," 2018 6th International Conference on Control Engineering & Information Technology (CEIT), 2018, pp. 1-6, doi: 10.1109/CEIT.2018.8751927.
- [4] Lesha Bhansali and Meera Narvekar. Gesture Recognition to Make Umpire Decisions. *International Journal of Computer Applications* 148(14):26-29, August 2016.
- [5] Suvarna Nandyal and Suvarna Laxmikant Kattimani 2021 *J. Phys.: Conf. Ser.* 2070 012148
- [6] Y. Madhuri, G. Anitha. and M. Anburajan., "Vision-based sign language translation device," 2013 International Conference on Information Communication and Embedded Systems (ICICES), 2013, pp. 565-568, doi: 10.1109/ICICES.2013.6508395.
- [7] Nusirwan Anwar bin Abdul Rahman, Kit Chong Wei and John See Faculty of Information Technology, Multimedia University.
- [8] Moin, A., Zhou, A., Rahimi, A. *et al.* A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition. *Nat Electron* 4, 54–63 (2021). <https://doi.org/10.1038/s41928-020-00510-8>
- [9] A. Ravi, H. Venugopal, S. Paul and H. R. Tizhoosh, "A Dataset and Preliminary Results for Umpire Pose Detection Using SVM Classification of Deep Features," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), 2018, pp. 1396-1402.
- [10] M. A. Shahjalal, Z. Ahmad, R. Rayan and L. Alam, "An approach to automate the scorecard in cricket with computer vision and machine learning," 2017 3rd International Conference on Electrical Information and Communication Technology (EICT), 2017, pp. 1-6, doi: 10.1109/EICT.2017.8275204.
- [11] Suvarna Nandyal and Suvarna Laxmikant Kattimani 2021 *J. Phys.: Conf. Ser.* 2070 012148