Automatic Object and Event Detection in Field Hockey Videos Using Deep Learning Techniques

A Synopsis Submitted in Partial Fulfilment
for the Award of Degree of

Doctor of Philosophy

In

ELECTRONICS AND COMMUNICATION ENGINEERING

By

Suhasbhai Haribhai Patel

Enrollment No: 189999915029

Under the Supervision of

Dr. Dipesh G.Kamdar

Asst. Professor (E.C. Dept.)

V.V.P Engineering College, Rajkot, Gujarat.

GUJARAT TECHNOLOGICAL UNIVERSITY, AHMEDABAD

September 2023

Contents

List of Figures

List of Tables

Title of the thesis and abstract

Title of the thesis

Automatic Object and Event Detection In Field Hockey Videos Using Deep Learning Techniques

Abstract

Field hockey, commonly known as hockey, is an outdoor team sport played between two opposing teams, each consisting of 11 players. The players use sticks that are curved at the striking end to hit a small, hard ball with the objective of scoring goals in their opponent's net. The term "field hockey" is used to distinguish this version of the game from a similar sport played on ice.

The analysis of field hockey videos is significant as it plays a crucial role in enhancing the overall understanding of the game and providing valuable insights for performance assessment. However, manual video analysis is a time-consuming and subjective process, heavily dependent on human observers. To overcome these challenges and enhance the analysis process, this research aims to develop an automated system for object and event detection in field hockey videos using deep learning techniques. The proposed research harnesses the power of deep learning, specifically convolutional neural networks (CNNs), to automatically detect and recognize key objects and events within field hockey videos. The system will be trained to identify players, the ball, and the umpire as essential objects, and events such as field goals, penalty corners, and penalty strokes. The research will consist of several stages, starting with data collection and annotation. A comprehensive dataset of field hockey videos will be gathered, and human annotators will meticulously label the relevant objects and events in each frame to create a ground truth for model training and evaluation. Next, deep learning models, particularly CNNs, will be employed to process the annotated data. State-of-the-art training techniques will be utilized to optimize the performance of these models, aiming for high accuracy and generalization in detecting the specified objects and recognizing the various events characteristic of field hockey gameplay.

This study introduces a deep learning-based transfer learning model, YOLOv3, specifically designed for object detection in field hockey scenarios. Key elements, namely AUS (Team 1), BEL (Team 2), Hockey ball, and Umpire, are successfully identified using this pre-trained model on the hockey dataset. The YOLOv3 model achieves an impressive accuracy of 91.3%, while the YOLOv8 model attains an even higher accuracy of 94.0% in detecting hockey objects within dataset_1.

Moreover, the application extends to recognizing significant activities such as goals, penalty corners, and penalties. Remarkably, VGG16 and Densenet models, employing a transformer-based approach, achieve a high accuracy of 99.47% in this activity recognition task. Additionally, when

applied to video datasets, the ConvLSTM-based model achieves an accuracy of up to 67%. These exceptional accuracy levels attained in both object and event recognition highlight the innovative potential of this approach within the realm of field hockey video analysis.

2. Brief description on the state of the art of the research topic

The emergence of machine learning technology has brought about a critical need to identify and track objects in sports videos. The sports industry is actively exploring automated systems to enhance productivity within organizations[1]. Figure 1 (a) and (b) present a sports event example, with a normal scene and an event scene depicting changes occurring in a short period. To recognize events and behaviors, observing the movement of objects and players' reactions, such as identifying goals in soccer, is common. However, the temporal aspect of the video is often overlooked during event analysis [2]. Sports video analysis is widely used to extract quick highlights, especially with advancements in video capture systems and analysis tools. Numerous applications have been developed for sports analysis, including video replay, statistics collection, and video archiving [3]. Due to the complexity of sports videos, different techniques are employed for low-level feature extraction. High-speed games pose challenges due to the ball's motion, making feature extraction complicated. The vast amount of data generated by sports channels and recorded events creates difficulties in producing comprehensive highlights packages [4]. Nonetheless, the growth of hardware technologies and video processing power has transformed sports video analysis into a significant research area with various applications, such as video annotation, referee decision-making, automatic play detection, and customized advertisement insertion [5].Sports video analysis encompasses object detection, highlight detection, and text analysis[6]. Object detection involves identifying the ball or players in sports videos, achieved through techniques like R-CNN, YOLO, and Mask R-CNN, based on convolutional neural networks. Highlight detection focuses on identifying video scenes that depict critical events by detecting changes in visual content [7]. Text analysis involves extracting context, including event results and match summaries, using the scoreboard template to interpret real-time scoreboard data [8]. The integration of machine learning and automated systems in sports video analysis has significantly enhanced event recognition, providing valuable insights and efficient processing. The advancements in this research area hold promise for revolutionizing sports industry practices and enriching the viewing experience for fans worldwide.

(a) (b)

Figure 1: Examples of Sport Event Detection (a) Normal scene (b) Event scene.

As of the current state of the art, research on "Automatic Object and Event Detection in Field Hockey Videos Using Deep Learning Techniques" has made significant strides in addressing the challenges of video analysis in the field of sports. Object detection in field hockey videos involves identifying and localizing essential elements, including players, the ball, and umpires Event recognition, another critical aspect of the research, aims to identify specific actions and occurrences during gameplay, such as field goals, penalty corners, and penalty strokes. Deep learning models have shown promise in recognizing these events from video sequences, offering valuable insights for coaches, players, and sports analysts. To train the deep learning models effectively, researchers have collected large datasets of annotated field hockey videos. Human annotators label the relevant objects and events in the video frames, serving as ground truth for model training and evaluation. The applications of the developed system extend beyond object and event detection. Sports video processing has various practical applications, including performance analysis, augmented reality presentation of sports events, faster and accurate video analysis, real-time feedback and decision support, improved video highlights and summaries, and advancement in sports analytics.

Overall, the state-of-the-art research in automatic object and event detection in field hockey videos using deep learning techniques has paved the way for advancements in sports video analysis. The integration of sophisticated deep learning models, coupled with annotated datasets and real-time capabilities, promises to provide comprehensive insights into field hockey gameplay, benefiting players, coaches, broadcasters, and sports enthusiasts alike.

3. Definition of the problem

The problem of "Automatic Object and Event Detection in Field Hockey Videos Using Deep Learning Techniques" involves creating an advanced computational system capable of autonomously identifying objects and events within field hockey videos through deep learning. Key objects encompass players (grouped into teams), the umpire, hockey ball, and goalies. The challenge is developing a model that accurately recognizes these objects within the dynamic

context of field hockey. Additionally, the problem entails automated recognition of significant events, including goals, penalty corners, and penalty strokes. The overarching goal is to replace labor-intensive manual video analysis by harnessing deep learning potential. By automating object and event detection, the system offers real-time insights, benefiting coaches, players, analysts, and audiences. This research contributes to sports video analysis, advancing understanding, performance assessment, and strategic decision-making in field hockey.

4. Objective and scope of work

Automatic object and event detection in field hockey videos using deep learning techniques stems from the lack of specialized models, scarcity of annotated datasets, insufficient incorporation of temporal information, and the need for practical real-time implementation. The primary objective of this research is to develop specialized deep learning models tailored for the precise identification and detection of objects and events within field hockey videos.

A crucial aspect of achieving this objective is the creation of a comprehensive and diverse annotated dataset of field hockey videos. This dataset will serve as the foundation for training and fine-tuning the deep learning models. Each video in the dataset will be meticulously annotated to accurately label key objects, such as players, the ball, and the umpire, as well as significant events, including goals, penalty corners, and penalty strokes.

In summary, the scope of this project revolves around the creation of specialized deep learning models, the establishment of a comprehensive annotated dataset, and the optimization of the detection system's practical implementation.

By accomplishing these objectives, this research aims to significantly enhance the automated analysis of field hockey videos, contributing to improved understanding, performance assessment, and decision-making in the sport.

5. Original contributions by the thesis

This research makes significant contributions to the domain of automatic object and event detection in field hockey videos through the utilization of deep learning methodologies. The primary accomplishments and findings of this study are outlined below:

YOLO Model for Object Detection: The work introduces the utilization of various iterations of the YOLO model for robust object detection within field hockey videos. These models effectively identify key objects including the teams (AUS and BEL), the hockey ball, and the umpire, sourced from the collected hockey dataset (Dataset-A). Notably, the YOLOv8 model attains the highest

accuracy within the range of 88.60% to 94.00%.

Event Detection via Image Classification: Event detection through image classification, Model-I utilizing the VGG16 architecture and Model-IX based on the Densenet with Transformer framework emerge as the standout performers among the assessed models. Both Model-I and Model-IX showcase outstanding performance, achieving precision, recall, F1 score, and accuracy levels hovering around 99.47% on Dataset-1, which encompasses a collection of 3035 images. Furthermore, the YOLOv8 model, pretrained on ImageNet, presents promising outcomes in terms of image classification and event detection within hockey videos on Dataset-2, which comprises a total of 7195 images.

ConvLSTM and LRCN Models for Event Recognition: The research widens its scope to encompass video classification, assessing the performance of both ConvLSTM and LRCN models. Remarkably, the ConvLSTM model outperforms the LRCN model in terms of accuracy on Dataset-3, which shares a similar volume of videos and input setups. Noteworthy for its complex architecture featuring a larger number of trainable parameters, the ConvLSTM model exhibits promising capabilities for automated event detection.

In summary, this research significantly contributes to the realm of automatic object and event detection in field hockey videos, underscoring the power and versatility of deep learning techniques. The findings and achievements outlined herein hold promise for future developments and innovations within this specialized domain.

6. Methodology of Research, Results / Comparisons

After establishing the problem and study objectives, the focus shifted to outlining the procedural framework. This involved identifying necessary materials and selecting the operational method. Each of these components is elaborated upon in this section.

This work utilizes self-prepared datasets for both object detection and event detection in field hockey. The image datasets employed are openly accessible through the Roboflow universe. Python serves as the primary programming language for implementing deep learning networks. The implementation of these models was executed utilizing libraries such as PyTorch, Keras, TensorFlow and Ultralytics.

The project was carried out utilizing the Google Colaboratory Pro platform. The implementation encompassed both the Colab Free version, equipped with the T4 GPU, and the Colab Pro version, featuring advanced GPUs such as NVIDIA P100 and V100.

Figure 2 illustrates the comprehensive block diagram outlining the methods employed in the pursuit of Automatic Object and Event Detection in Field Hockey Videos. The subsequent section will explore a detailed description of these methods, focusing on those that exhibit superior performance.

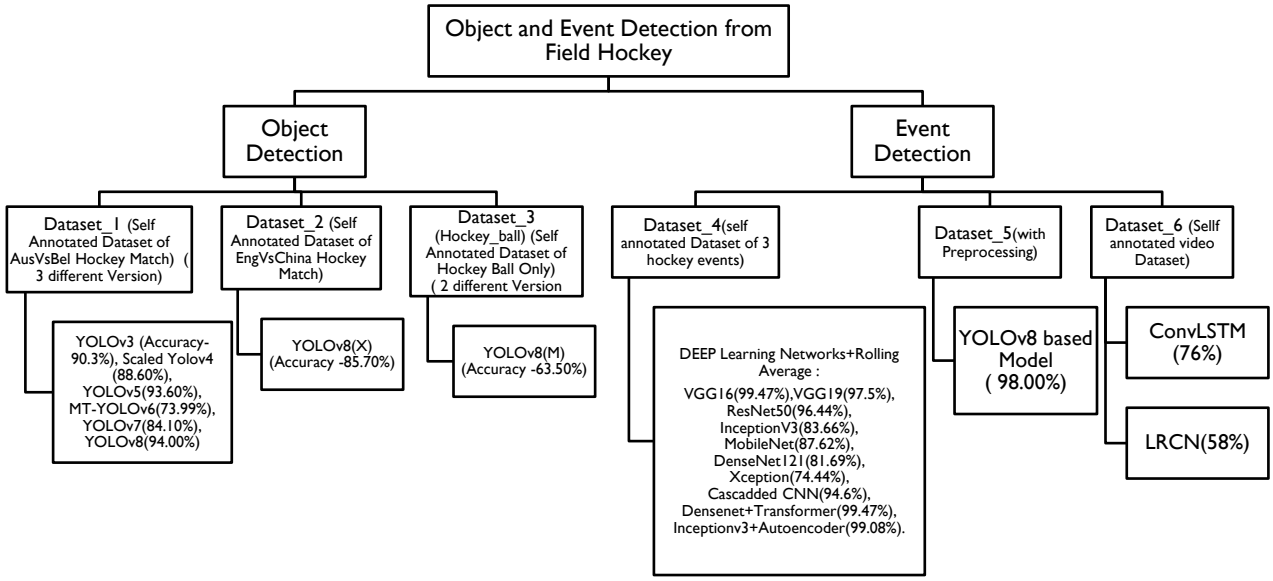


Figure 2: Comprehensive block diagram of the study's framework.

6.1 Automatic Object Detection Model

The surge in sports analysis research stems from the availability of vast internet datasets and the effective application of Convolutional Neural Network (CNN) techniques for object detection and image classification [9] . Traditional strategies for object detection in sports videos typically revolve around player identification, utilizing methods such as connected component analysis[10], shallow convolutional neural networks[11], histogram of oriented gradients and support vector machines (HOG-SVM) [12], and deformable part model (DPM) [13]. These methods aimed to locate basic attributes, like the position of a primary object, such as a soccer ball [14]. In baseball, pitcher style detection employed techniques like object segmentation algorithms [15]. Figure 3 (a) is an example of a typical image where each object in the frames is separate from the others. Traditional model of object detection can identify the objects from these types of images, while it can barely detect objects in case of occlusion, self-occlude and situation as in figure 3 (b)-(d). The landscape of object detection has undergone a significant transformation through the ascent of neural networks, spurred by the rapid advancement of computer vision [16]. The evolution of YOLO (You Only Look Once) object identification algorithms, spanning versions 1 through 8, has notably surpassed

traditional methods in both capacity and performance.

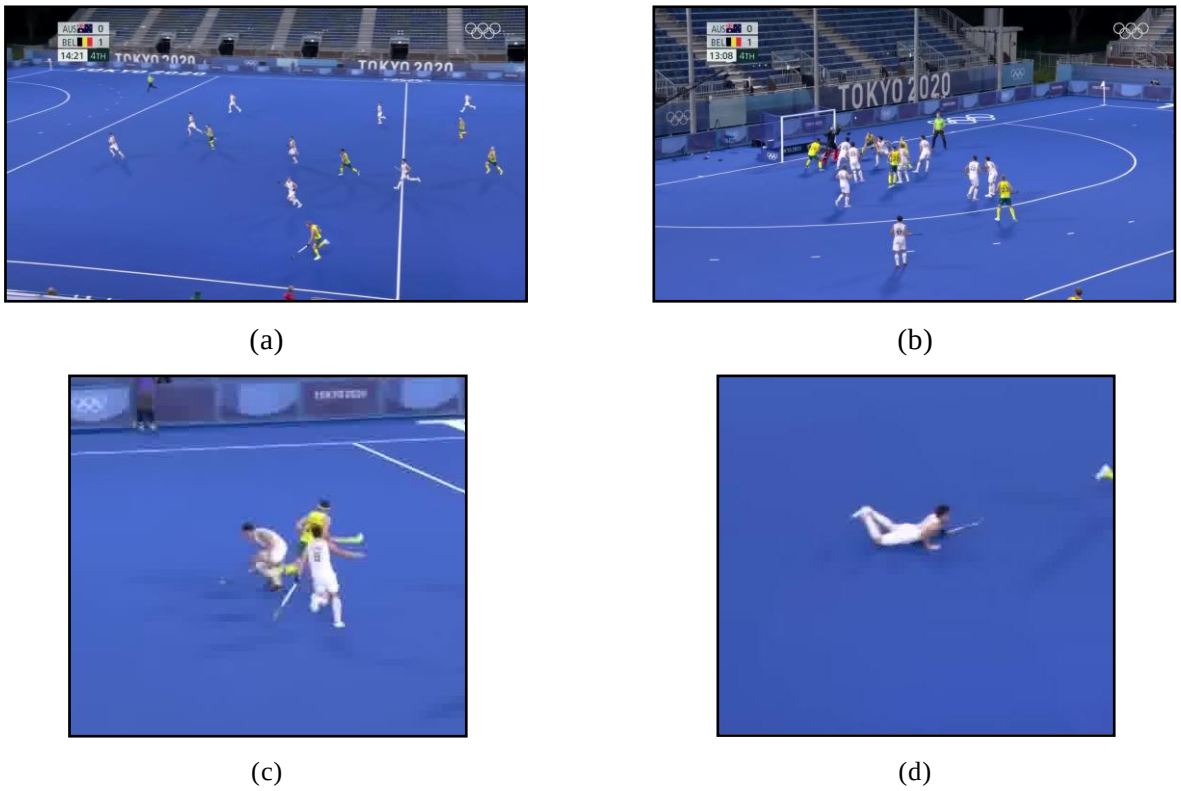


Figure 3: (a) a normal hockey match condition (b) Players are close to each other (c) players position causes occlusion (d) one player lay down on ground.

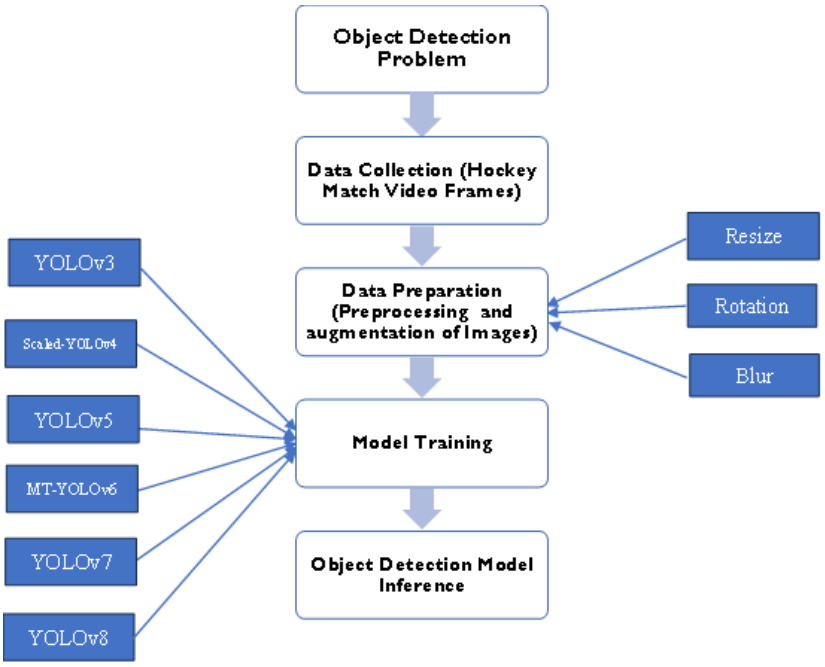


Figure 4: Flow Chart of Object detection in field hockey

6.1.1 Dataset Preparation

As, there is no publically accessible object detection dataset of hockey match. We use a self-

prepared dataset1 for object detection from a YouTube video of a field hockey match between Australia and Belgium (Tokyo Olympics 2020 gold medal match), where four object names are AUS(Team 1), BEL(Team 2), Hockey ball, and Umpire. The video resolution was 1920X1080, and it splits into shorter duration video where the camera angle is in wide mode. These video clips are converted into frames that are manually annotated.

Total 1119 frames with 1920x1080 resolutions were manually labeled with four labels AUS (Team 1), BEL (Team 2), Hockey ball, and Umpire. Results are below optimum when model training is started right after data collection, in this case, images are preprocessed by image resize(640x640) and auto orientation, also augmented by rotation(-15 degree to +15 degree) and blur effect(up to 10 pix) ,after Preprocessing and augmentation total of 2683 images as per details in Table 1[17].

	Without Pre-processing (DATASET_1A)	**With Pre-processing and Augmentation (DATASET_1B)**	**With Pre-processing and Augmentation (Image Size Remain Same) (DATASET_1C)**
Total Images	1119	2683	2683
Classes	4	4	4
Unannotated	0	0	0
Training Set	783 (70%)	2347 (87%)	2347 (87%)
Validation Set	224 (20%)	224 (8%)	224 (8%)
Testing Set	112 (10%)	112 (4%)	112 (4%)
Annotation	12,937 (11.6 per Image (Average)	30939 (11.53 per Image (Average)	30939 (11.53 per Image (Average)
Average Image Size	2.07 mp	33.02 k	2.07 mp
Median Image Ration	1920 x1080	640x640	1920x1080
Class Instances			
1. AUS	5559(42.96%)	13293(42.96%)	13293(42.96%)
2. BEL	5973(46.16%)	14258(46.08%)	14258(46.08%)
3.Hockey_Ball	865(6.68%)	2075(6.70%)	2075(6.70%)
4.Umpire	540(4.17%)	1313(4.24%)	1313(4.24%)

Table 1: Object Detection Dataset_1

Total 2532 frames with 1280x720 resolutions were manually labeled with four labels China, (Team 1), England (Team 2), Hockey ball, Umpire and Goalies which represented here as DATASET_2 as shown in Table 2.

	Without Pre-processing (DATASET_2)
Total Images	2532
Classes	5
Unannotated	0
Training Set	1791 (70%)
Validation Set	511 (20%)
Testing Set	252 (10%)

Annotation	24918 (9.8 per Image(Average)
Average Image Size	0.92 mp
Median Image Ration	1280x720
Class Instances	
1.China	11558(46.38%)
2. England	11087(44.49%)
3.Hockey_Ball	1259(5.05%)
4.Umpire	681(2.73%)
4.Goalies	331(1.32%)

Table 2: Object Detection Dataset_2

A comprehensive dataset comprising a total of 2955 frames, each possessing a resolution of 1280x720, was meticulously annotated to identify the presence of the Hockey ball. To further enhance the dataset's diversity and richness, an array of pre-processing and image augmentation techniques were systematically employed, culminating in the creation of an expanded dataset comprising a total of 7087 images as shown in table 3**.**

The augmentation procedures encompassed various transformative actions, including auto-orientation, flips (horizontal, vertical), rotations (90° clockwise, counter-clockwise, upside down), cropping (ranging from 0% minimum zoom to 20% maximum zoom), and rotations within a range of -15° to +15°. Additionally, shear effects were introduced, both horizontally and vertically, within a tolerance of ±15°. The technique of mosaic application was also integrated into the augmentation process.

In conjunction with these procedures, bounding boxes received dedicated attention. They underwent analogous transformations such as flips (horizontal, vertical), rotations (90° clockwise, counter-clockwise, upside down), cropping (ranging from 0% minimum zoom to 20% maximum zoom), rotations within the aforementioned -15° to +15° spectrum, and shear effects (±15° horizontally and vertically). Beyond geometric variations, bounding boxes were subjected to alterations in brightness (ranging from -25% to +25%), exposure (within the same -25% to +25% range), blur (up to 2.5 pixels), and noise (up to 5% of pixels). These meticulous processes collectively contributed to the augmentation of the dataset, encapsulating a broad spectrum of potential scenarios and variations.

	Without Pre-processing (Dataset_3A)	**With Pre-processing and augmentation (Dataset_3B)**
Total Images	2955	7087
Classes	01	01
Unannotated	00	98
Training Set	2066	6198
Validation Set	577	577
Testing Set	312	312

Annotation	4740 (1.0 per Image (Average))	12542(1.76 per Image (Average))
Average Image Size	0.92 mp	0.92 mp
Median Image Ratio	1280x720	1280x720
Class Instances		
Hockey_Ball	2955(100%)	12542(176%)

Table 3: Object Detection Hockey ball Dataset_3

6.1.2 Object Detection using YOLOv3

In this object detection system, YOLOv3-based model Darknet-53 has been pre-trained by COCO Dataset is receives images of 640x640 pixels as inputs; the batch size is set to 16. The model trained for 100, 200, and 300 epochs. The Scaled weight_decay is 0.0005, and Stochastic Gradient Descent (SGD) optimizer with parameter groups 72 weight, 75 weights (no decay) and 75 biases are set for this model. The video frames of hockey game are the input of the model. The inputs were passed on to the yolov3 model, fine-tuned for this hockey object detection task. The output was obtained from the highest confidence score of the bounding box after non-maximum suppression. This model takes the complete visual frame and extracts features at the frame level. The YOLOv3 model passes through successive convolution layers from the first input layer to the last layer, learning patterns from the entire frame and extracting low-level features to high-level features[18].

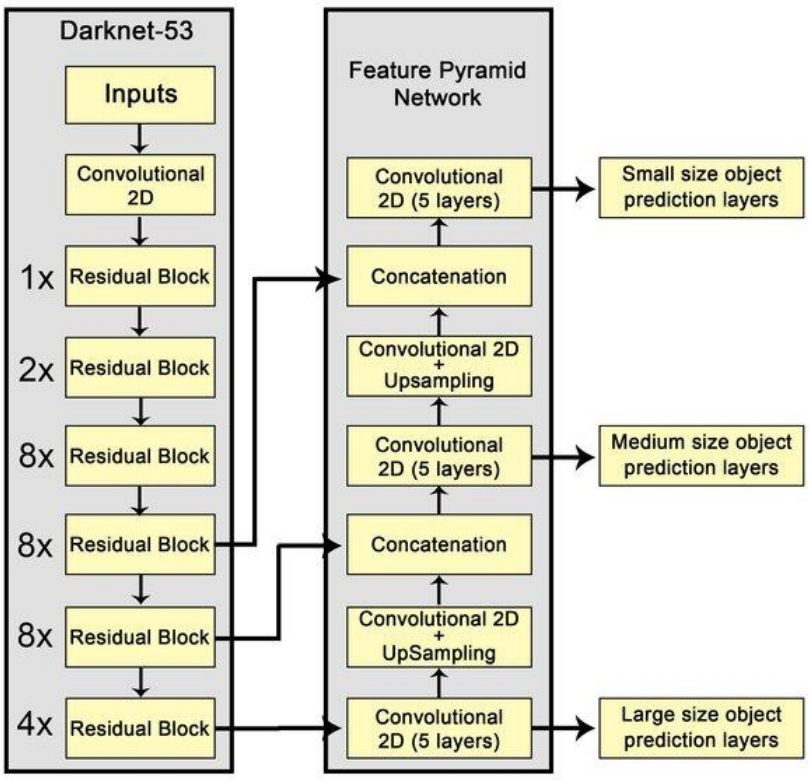


Figure 5: Yolov3 Model Architecure

No. of epochs	CLASS	PRECISION	RECALL	F-1 SCORE	mAP@ 0.5	Overall Accuracy (Map@0.5)
100	AUS	0.982	0.988	0.985	98.90%	**92.30%**
	BEL	0.985	0.988	0.986	99%	
	HOCKEY BALL	0.829	0.684	0.75	71.90%	
	UMPIRE	0.985	1	0.992	99.50%	
200	AUS	0.979	0.99	0.984	99.10%	92.00%
	BEL	0.984	0.99	0.987	99.30%	
	HOCKEY BALL	0.83	0.684	0.75	70.10%	
	UMPIRE	0.984	1	0.992	99.50%	
300	AUS	0.982	0.989	0.985	99%	92.20%
	BEL	0.988	0.993	0.99	99.40%	
	HOCKEY BALL	0.836	0.684	0.752	71%	
	UMPIRE	0.993	1	0.996	99.50%	

Table 4: Accuracy of Model : Yolov3, Dataset_1A (Epochs 100,200,300 ; Image Size 640; Patience 100 ; Device GPU; Batch Size 16 , Optimizer: SGD)

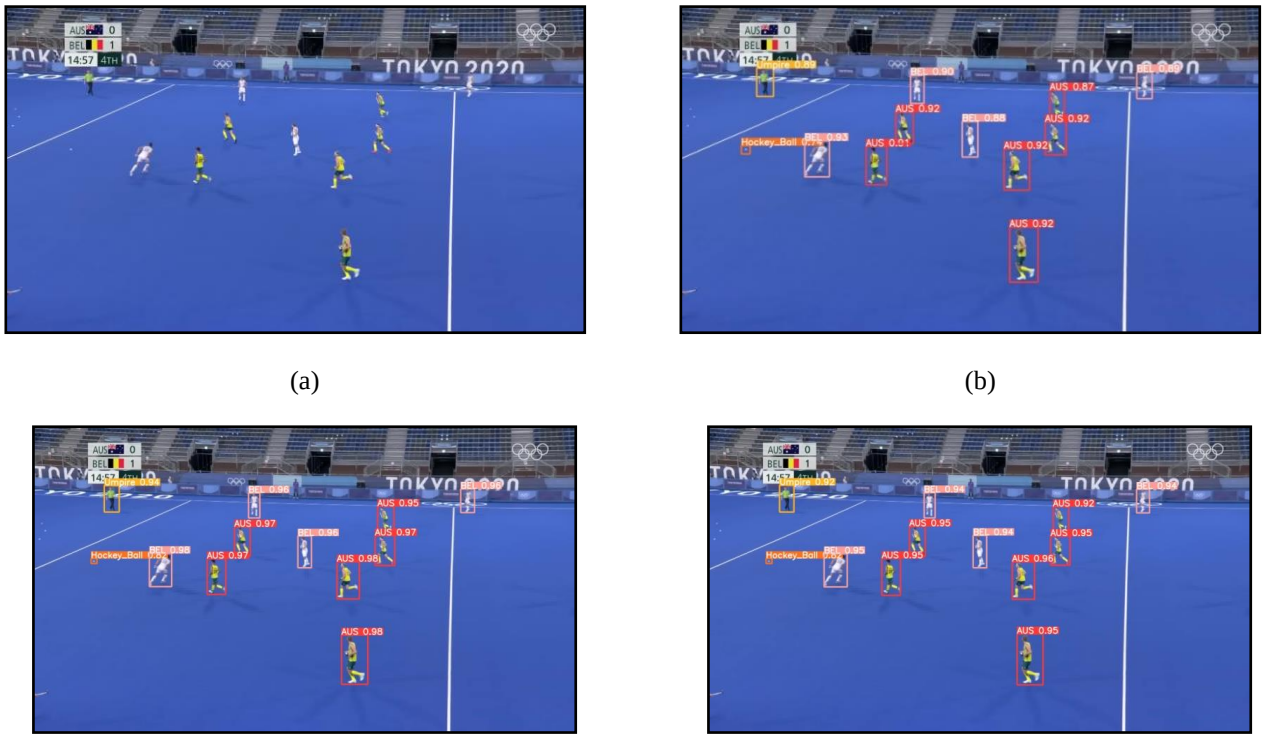
No. of epochs	CLASS	PRECISION	RECALL	F-1 SCORE	mAP@ 0.5	Overall Accuracy (Map@0.5)
100	AUS	0.976	0.99	0.983	99.1%	88.9%
	BEL	0.98	0.99	0.985	99.3%	
	HOCKEY BALL	0.779	0.596	0.675	58.3%	
	UMPIRE	0.974	0.989	0.981	98.9%	
200	AUS	0.97	0.98	0.975	99%	91.2%
	BEL	0.99	0.994	0.992	99.4%	
	HOCKEY BALL	0.86	0.637	0.732	66.7%	
	UMPIRE	0.98	1	0.990	99.5%	
300	AUS	0.972	0.988	0.980	99%	**91.3%**
	BEL	0.988	0.992	0.990	99.4%	
	HOCKEY BALL	0.803	0.69	0.742	67.5%	
	UMPIRE	0.983	1	0.991	99.5%	

Table 5: Accuracy of Model : Yolov3, Dataset_1B (Epochs 100,200,300 ; Image Size 640; Patience 100 ; Device GPU; Batch Size 16 ,Optimizer: SGD)

No. of epochs	CLASS	PRECISION	RECALL	F-1 SCORE	mAP@ 0.5	Overall Accuracy (Map@0.5)
100	AUS	0.976	0.988	0.982	98.80%	**93.30%**
	BEL	0.987	0.993	0.99	99.30%	
	HOCKEY BALL	0.865	0.749	0.803	75.80%	
	UMPIRE	0.985	1	0.992	99.50%	
200*	AUS	0.983	0.988	0.985	99.10%	92.80%
	BEL	0.987	0.991	0.989	99.40%	
	HOCKEY BALL	0.865	0.713	0.782	73.20%	
	UMPIRE	0.988	1	0.994	99.50%	

* The training process has been stopped early as there was no observed improvement in the last 100 epochs

Table 6: Accuracy of Model: Yolov3, Dataset_1C (Epochs 100,200 ; Image Size 640; Patience 100 ; Device GPU; Batch Size 16 ,Optimizer: SGD)



(a) (b)

(c) (d)

Figure 6: Output of object Detection model for DATASET_1B, MODEL: YOLOv3 (a) Input Image, Object Detection output for (a) 100 epoch (b) 200 epochs (c) 300 epochs

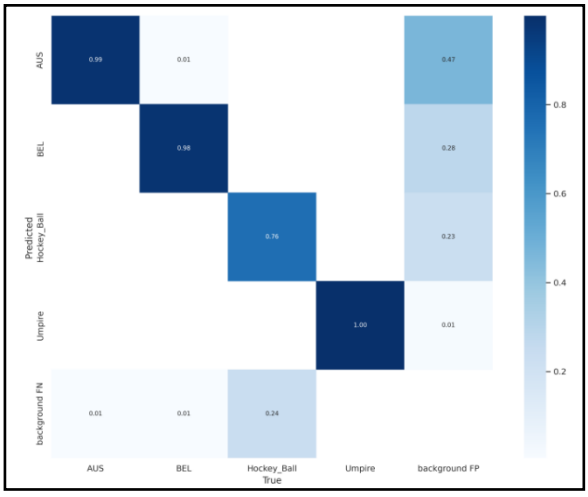


Figure 7: Confusion matrix for DATASET_1B, MODEL: YOLOv3, Epochs: 300

6.1.3 Object Detection using Scaled YOLOv4

Scaled-YOLOv4 is a variant of the YOLO (You Only Look Once) object detection model that has been enhanced and optimized for improved performance. Scaled-YOLOv4 builds upon the original YOLOv4 model by incorporating various enhancements and optimizations that allow it

to achieve better accuracy and efficiency in object detection tasks[19].

No. of Epochs	PRECISION	RECALL	F-1 SCORE	mAP@ 0.5
100	0.748	0.862	0.801	83.30%
200	0.791	0.893	0.839	88.40%
300	0.803	0.892	0.845	87.90%
400	0.811	0.897	0.852	88.00%
500	0.81	0.902	0.854	**88.60%**

Table 7: Accuracy of Model: Scaled Yolov4, Dataset_1C (Image Size 416; Device GPU; Batch Size 16)

6.1.4 Object Detection using YOLOv5.

YOLOv5 is a cutting-edge deep learning model for real-time object detection and image classification[20]. It boasts enhanced accuracy, speed, and efficiency compared to its predecessors.

No. of Epochs	YOLOv5 Model	Model Size (MB)	Label	PRECISION	RECALL	F-1 SCORE	Overall Accuracy (Map@0.5)
100	N (Nano)	5.0	2595	0.94	0.834	0.884	85.30%
	S (Small)	17.6	2595	0.941	0.869	0.904	88.80%
	M (Medium)	48.1	2595	0.954	0.882	0.917	90.80%
	L (Large)	101.8	2595	0.944	0.899	0.921	91.40%
	X	556.8	2595	0.946	0.896	0.92	**91.70%**

Table 8: Accuracy of Model: Yolov5, Dataset_1A (Epochs 100 ; Image Size 640; Patience 100 ; Device GPU; Batch Size : Custom)

No. of epochs	YOLOv5 Model	Model Size (MB)	Label	PRECISION	RECALL	F-1 SCORE	Overall Accuracy (Map@0.5)
100	N (Nano)	5.0	2595	0.916	0.895	0.905	90.10%
	S (Small)	17.7	2595	0.947	0.91	0.928	92.20%
	M (Medium)	48.1	2595	0.951	0.916	0.933	93.50%
	L (Large)	101.8	2595	0.946	0.921	0.933	93.30%
	X	556.8	2595	0.948	0.933	0.940	**93.60%**

Table 9: Accuracy of Model: Yolov5, Dataset_1B (Epochs 100 ; Image Size 640; Patience 100 ; Device GPU; Batch Size : Custom)

No. of epochs	YOLOv5 Model	Model Size (MB)	Label	PRECISION	RECALL	F-1 SCORE	Overall Accuracy (Map@0.5)
100	N (Nano)	5.0	2595	0.915	0.86	0.887	87.00%
	S (Small)	17.6	2595	0.94	0.884	0.911	90.60%
	M (Medium)	48.1	2595	0.94	0.896	0.917	91.20%
	L (Large)	101.8	2595	0.945	0.903	0.924	92.70%
	X	185.9	2595	0.965	0.898	0.930	**92.80%**

Table 10: Accuracy of Model: Yolov5, Dataset_1C (Epochs 100 ; Image Size 640 ; Patience 100 ; Device GPU; Batch Size : Custom)

6.1.5 Object Detection using MT-YOLOv6, YOLOv7

Developed by Meituan's Visual Intelligence Department, MT-YOLOv6 is a single-stage object detection framework tailored for industrial applications. Enhanced YOLOv6 incorporates an anchor-free paradigm, SimOTA label assignment, and SIoU bounding box regression loss, resulting in improved speed, detection accuracy, and network learning[21].

YOLOv7 provides notable improvements in real-time object detection accuracy while preserving inference efficiency, boasting a 40% reduction in parameters and 50% computation compared to alternative detectors[22].

No. of epochs	MODEL	CLASS	Images	Label	PRECISION	RECALL	F-1 SCORE	Overall Accuracy (Map@0.5)
100	MT-YOLOv6	ALL	224	2595	0.74	0.62	0.675	73.99%
55	YOLOv7	ALL	224	2595	0.861	0.857	0.859	84.10%

Table 11: Accuracy of Model: MT-YOLOv6 and YOLOv7 Dataset_1C

6.1.6 Object Detection using YOLOv8

YOLOv8 represents a single-stage object detection model, making predictions for object bounding boxes and class labels in one go. Building upon the YOLOv3 foundation, YOLOv8 introduces several enhancements, including a more efficient and accurate backbone network, an anchor-free approach in the head network for object detection, and a robust loss function that handles object occlusion and deformation better[20].

No. of epochs	YOLOv8 Model	Model Size (MB)	CLASS	Images	Label	PRECISION	RECALL	F-1 SCORE	Overall Accuracy (Map@0.5)
100	N (Nano)	5.9	ALL	224	2595	0.942	0.85	0.894	86.00%
	S (Small)	21.4	ALL	224	2595	0.952	0.872	0.910	89.10%
	M (Medium)	49.6	ALL	224	2595	0.939	0.883	0.910	91.10%
	L (Large)	83.6	ALL	224	2595	0.95	0.893	0.921	91.60%
	X	130.4	ALL	224	2595	0.944	0.909	0.926	**92.40%**

Table 12: Accuracy of Model: Yolov8, Dataset_1A (Epochs 100 ; Image Size 640; Patience 100 ; Device GPU; Batch Size : Custom)

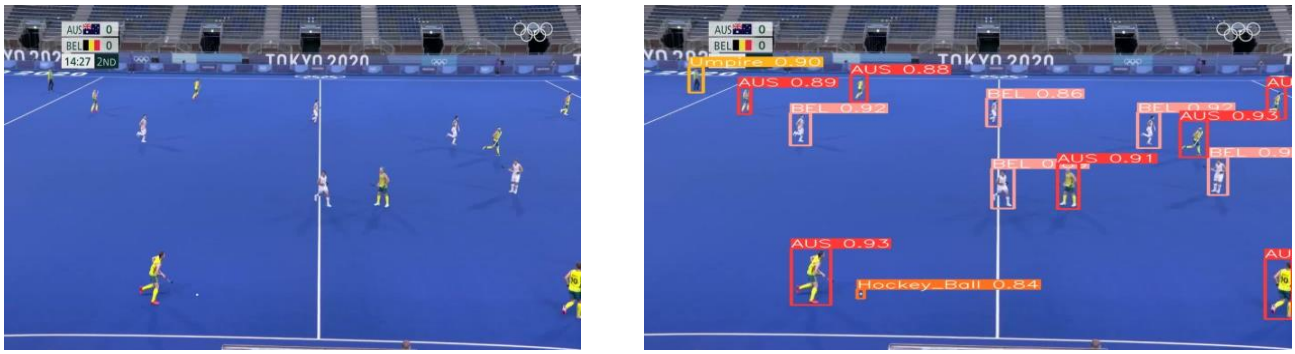
No. of epochs	YOLOv8 Model	CLASS	Model Size (MB)	Images	Label	PRECISION	RECALL	F-1 SCORE	Overall Accuracy (Map@0.5)
100	N (Nano)	ALL	6.0	224	2595	0.915	0.891	0.903	89.40%
	S (Small)	ALL	21.5	224	2595	0.914	0.907	0.910	91.80%

	M (Medium)	ALL	49.6	224	2595	0.957	0.922	0.939	94.00%
	L (Large)	ALL	83.6	224	2595	0.955	0.928	0.941	93.60%
	X	ALL	130.4	224	2595	0.949	0.93	0.939	94.00%

Table 13: Accuracy of Model: Yolov8, Dataset_1B (Epochs 100 ; Image Size 640; Patience 100 ; Device GPU; Batch Size : Custom)

No. of epochs	**YOLOv8 Model**	**Model Size(MB)**	**CLASS**	**Images**	**Label**	**PRECISION**	**RECALL**	**F-1 SCORE**	**Overall Accuracy (Map@0.5)**
100	N (Nano)	5.9	ALL	224	2595	0.91	0.87	0.89	87.60%
	S (Small)	21.4	ALL	224	2595	0.959	0.875	0.915	91.00%
	M (Medium)	49.6	ALL	224	2595	0.964	0.897	0.929	92.70%
	L (Large)	83.6	ALL	224	2595	0.947	0.902	0.924	92.40%
	X	130.4	ALL	224	2595	0.937	0.913	0.925	93.40%

Table 14: Accuracy of Model: Yolov8, Dataset_1C (Epochs 100; Image Size 640; Patience 100 ; Device GPU; Batch Size : Custom)



(a) (b)

Figure 8: Output of object Detection model for DATASET_1, MODEL: YOLOv8x (a) Input Image, Object Detection output for (a) 100 epoch

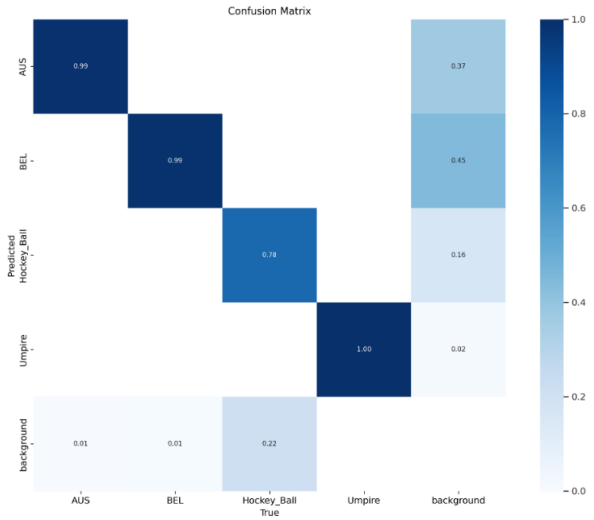


Figure 9: Confusion matrix for DATASET_1, MODEL: YOLOv8x,Epochs: 100

No. of epochs	YOLOv8 Model	Model Size (MB)	CLASS	Images	PRECISION	RECALL	F-1 SCORE	Overall Accuracy (mAP@0.5)
100	N (Nano)	5.9	ALL	511	0.825	0.857	0.841	84.30%
	S (Small)	21.4	ALL	511	0.802	0.875	0.837	84.80%
	M (Medium)	49.6	ALL	511	0.832	0.867	0.849	85.50%
	L (Large)	83.6	ALL	511	0.833	0.861	0.847	85.40%
	X	130.4	ALL	511	0.832	0.861	0.846	85.70%

Table 15: Accuracy of Model: Yolov8, **Dataset_2** (Epochs 100 ; Image Size 640; Patience 100 ; Device GPU; Batch Size : Custom)

No. of epochs	YOLOv8 Model	Model Size (MB)	CLASS	Images	PRECISION	RECALL	F-1 SCORE	Overall Accuracy (mAP@0.5)
100	N (Nano)	6.2	ALL	511	0.726	0.622	0.67	63.30%
	S (Small)	22.5	ALL	511	0.726	0.622	0.67	63.40%
	M (Medium)	50.5	ALL	511	0.696	0.634	0.664	65.70%
	L (Large)	87.6	ALL	511	0.724	0.586	0.648	62.80%
	X	136.7	ALL	511	0.682	0.64	0.660	63.30%

Table 16: Accuracy of Model: Yolov8, **Dataset_3A** (Epochs 100 ; Image Size 640; Patience 100 ; Device GPU; Batch Size : Custom)

No. of epochs	YOLOv8 Model	Model Size (MB)	CLASS	Images	PRECISION	RECALL	F-1 SCORE	Overall Accuracy (mAP@0.5)
100	N (Nano)	5.9	ALL	577	0.682	0.628	0.654	63.30%
	S (Small)	21.5	ALL	577	0.704	0.657	0.680	65.60%
	M (Medium)	49.6	ALL	577	0.693	0.648	0.670	63.50%
	L (Large)	83.6	ALL	577	0.669	0.666	0.667	64.60%
	X	130.4	ALL	577	0.706	0.631	0.666	62.80%

Table 17: Accuracy of Model: Yolov8, **Dataset_3B** (Epochs 100 ; Image Size 640; Patience 100 ; Device GPU; Batch Size : Custom)

6.2 Automatic Event Detection

The utilization of deep learning techniques for automatic event detection has brought about a transformative shift in the realm of sports video analysis. This advancement facilitates the effortless identification and classification of events within video footage, eliminating the necessity for human intervention. This automated procedure is not only characterized by its

rapidity but also its remarkable accuracy, culminating in substantial reductions in both time and effort. By swiftly and precisely identifying events such as goals, fouls, and crucial moments, this technology significantly amplifies our comprehension of the intricate dynamics at play within the videos.

Deep learning methods have been successfully applied to event recognition in various sports, including soccer, basketball, tennis, and cricket. These approaches often involve preprocessing video frames, extracting visual features using pretrained CNNs, and employing classifiers to recognize specific events. In the domain of field hockey, however, there is limited research on event recognition using deep learning methods. Our work aims to bridge this gap by proposing a deep learning-based approach specifically designed for field hockey event recognition. By leveraging the power of pretrained CNNs, we aim to overcome the challenges associated with accurately identifying events in the fast-paced and complex nature of field hockey gameplay.

For this research on automatic event detection, the following field hockey events are identified for analysis:

1. Goal: A goal is an event that signifies the successful scoring of a point by a team when they hit the ball into the opponent's goal post.

2. Penalty Corner: Penalty corner situations are identified when the defending team commits a foul inside their own circle, leading to the attacking team being awarded a set play opportunity to score a goal.

3. Penalty Stroke: A penalty stroke involves a one-on-one situation where the attacking player takes a shot from a specified distance against the goalkeeper.

6.2.1 Dataset of Field Hockey Events

As there is a lack of publicly available field hockey datasets for event recognition, we constructed our own dataset specifically tailored to the sport. We analyzed a collection of more than 28 highlights videos from the tournaments of the hockey pro league for the years 2021-22 and 2022-23. By carefully analyzing these highlights videos, we were able to identify and extract important events such as goals, penalty corners, and penalty.

Total Images	3035
Classes	3
Unannotated	0
Training Set	2276 (75%)
Testing Set	759 (25%)
Average Image Size	2.07 mp
Median Image Ration	1920x1080

Class Instances	
Goal	1000(32.95%)
Penalty Corner	1017(33.51%)
Penalty	1018(33.54%)

Table 18: Hockey Event Recognition Dataset_3

6.2.2 Proposed Deep Learning Networks for Event Detection

The methods employed in this research involve utilizing a pretrained convolutional neural network (CNN) to train a classifier specifically designed for event recognition in field hockey videos. The performance of the approach is then evaluated using this carefully prepared Field Hockey Event Dataset, providing insights into the effectiveness and accuracy of the proposed method for event recognition in the context of field hockey videos. The findings of this research reveal that the proposed deep learning approach for event recognition in field hockey videos achieves a remarkable accuracy of 99.47%. This high level of accuracy highlights the effectiveness of the approach in accurately identifying and classifying events in field hockey.

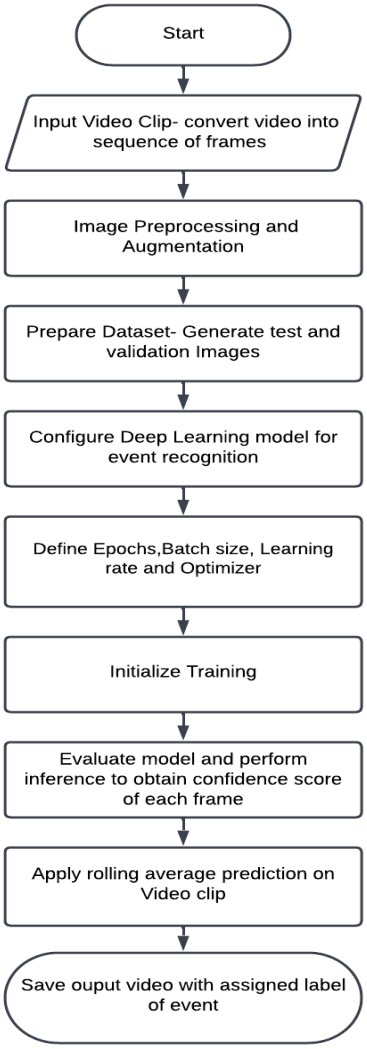


Figure 10: Process of hockey event recognition using a deep learning model

Sr no.	Model	Pretrained Network/ Deep Learning Network	Trainable Parameters	Precision (%)	Recall (%)	F1 Score (%)	Accuracy (%)
1	Model-1	VGG16	12,847,107	99.33	99.33	99.33	99.47
2	Model-II	VGG19	264,195	97.67	97.67	97.33	97.50
3	Model-III	ResNet50	1,050,627	96.33	96.33	96.33	96.44
4	Model-IV	InceptionV3	4,196,355	84.67	83.67	84.00	83.66
5	Model-V	MobileNet	526,339	88.33	87.67	87.67	87.62
6	Model-VI	DenseNet121	526,339	86.00	81.67	81.67	81.69
7	Model-VII	Xception	1,050,627	76.33	74.00	74.00	74.44
8	Model-VIII	Cascadded CNN	44,528,195	94.67	94.67	94.67	94.60
9	Model-IX	Densenet+Transformer	11,168,647	99.67	99.67	99.33	99.47
10	Model-X	Inceptionv3+Autoencoder	26,816,035	99.00	99.00	99.00	99.08

Table 19: Fine-tuned Deep Learning model results.

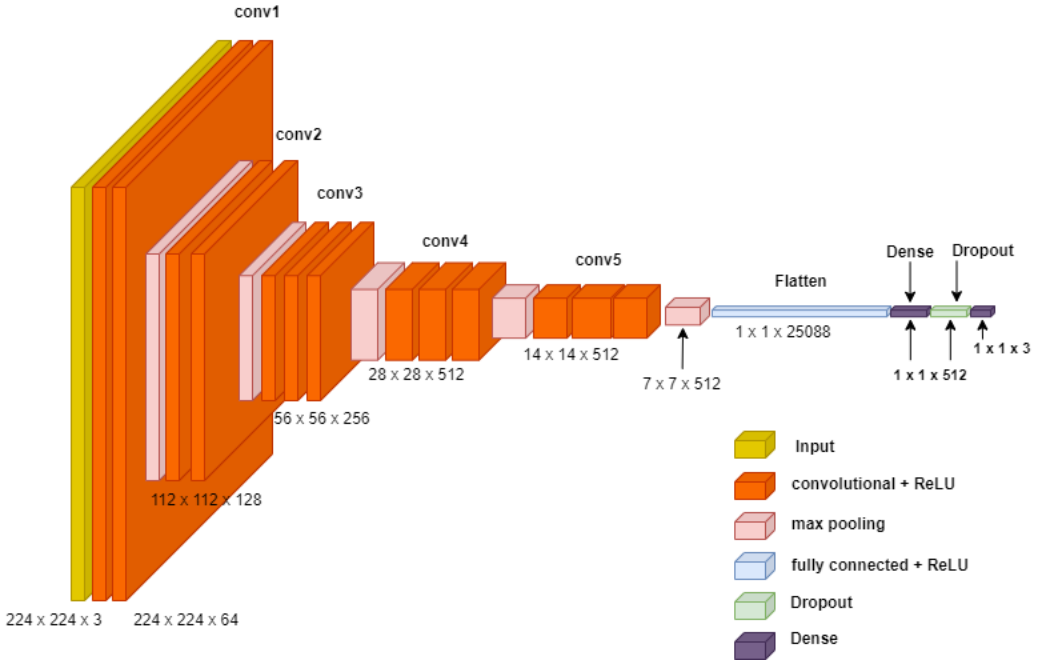


Figure 11: Proposed Model-1 Architecture.

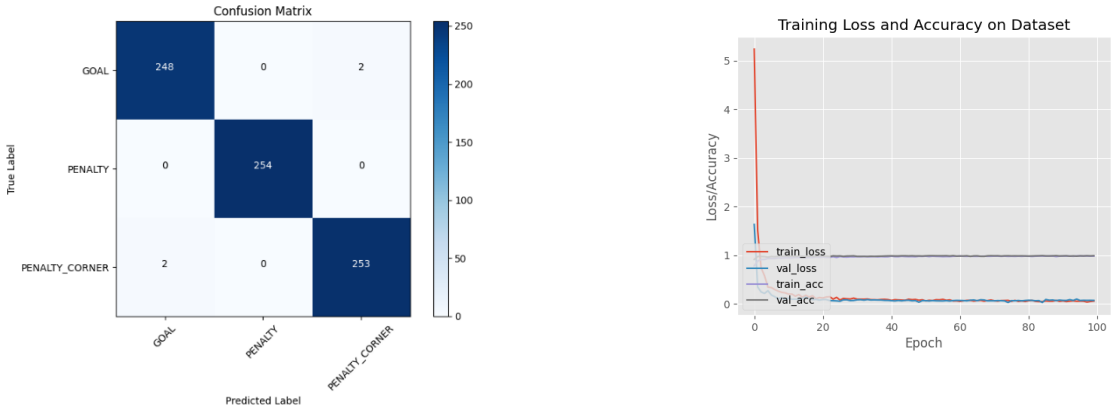


Figure 12: Proposed VGG-16 based Model-1 confusion matrix and Training loss and accuracy graph.

6.2.3 Rolling average prediction for Event detection.

To enhance the stability and reliability of the predictions, we apply a rolling average technique. This involves averaging the recent predictions over a certain period or a specific number of frames. By incorporating information from multiple frames, we can mitigate the impact of temporary variations or noise in individual frame predictions, resulting in a more robust and consistent prediction for the event happening in the video. The rolling average prediction approach helps to smooth out any fluctuations or inconsistencies in the frame-level predictions, providing a more accurate estimation of the event occurring in the video at any given time [23].

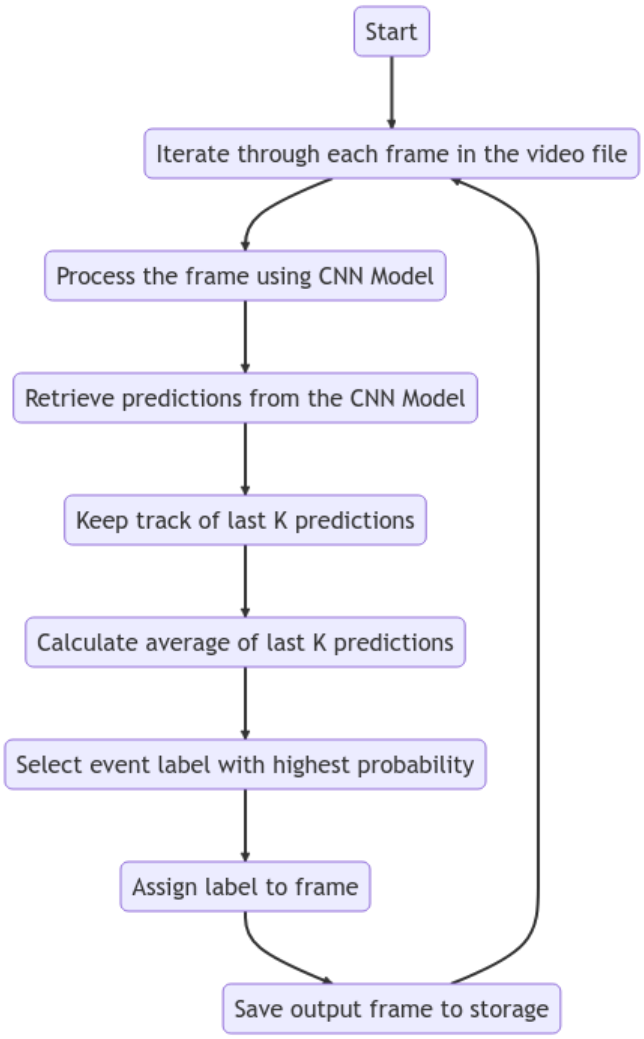
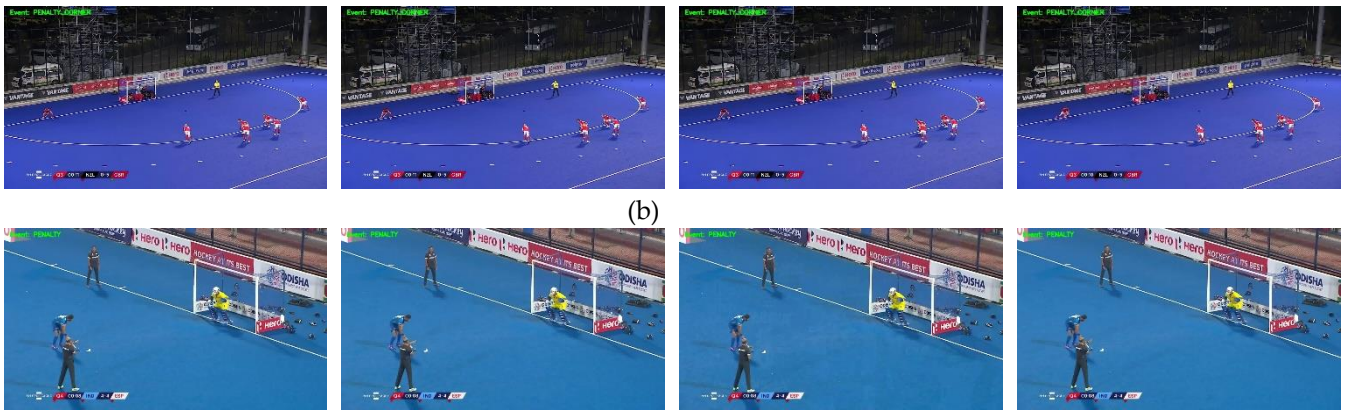


Figure 13: Process of rolling average prediction for Event detection.

(a)



(b)

(c)

Figure 14. Hockey event recognition output for Proposed model-1 (a) Goal, (b) Penalty Corner, (c) Penalty

6.2.4 Proposed YOLOv8 Based Model

In this phase, the focus shifts to applying advanced deep learning image classification models to dataset_5, encompassing a collection of 7195 images. One such model under consideration is YOLOv8, an acclaimed creation by Ultralytics, the same developers behind YOLOv5 [24]. Distinguished as a cutting-edge model for both object detection and image segmentation, YOLOv8 uniquely offers built-in support for image classification tasks as well.

	No. of Images
Input Images	3000
pre-processing	Auto Orient: Applied
	Resize: Stretch to 640x640
Augmentations	Flip: Horizontal
	Rotation: Between -15° and +15°
	Grayscale: Apply to 25% of images
	Brightness: Between -25% and +25%
Training Set	6294(87.47%)
Validation Set	597 (8.29%)
Testing Set	304 (4.22%)
Total Images with pre-processing and augmentations	7195

Table 20: Event Detection Dataset_5

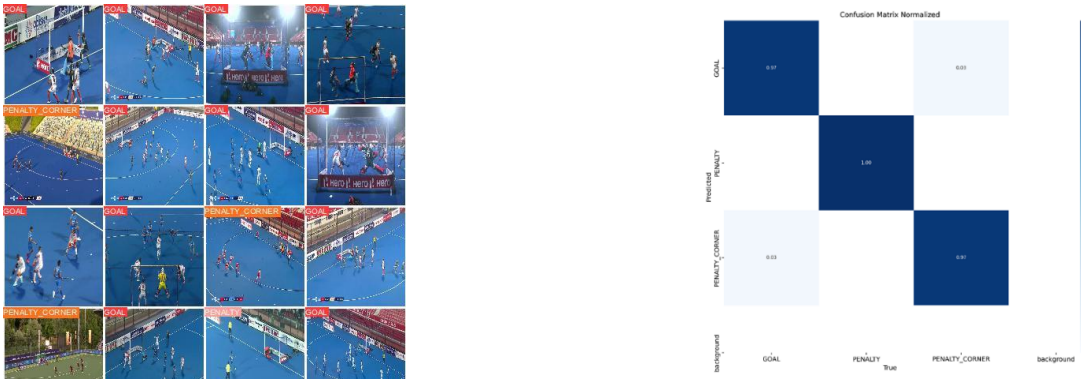


Figure 15: Confusion matrix and Predicted output of proposed YOLOv8

6.2.5 Proposed ConvLSTM based Event Detection Model for Video Dataset

ConvLSTM (Convolutional LSTM) is a type of recurrent neural network (RNN) that extends the capabilities of regular LSTMs to process spatio-temporal data, such as video sequences[25]. It combines convolutional layers and LSTM layers to learn both spatial and temporal features from video data.

Class	No. of Video
Field_Goal	111 (38.54%)
Penalty_Corner	111 (38.54%)
Penalty_Stroke	66 (22.91%)
Total Video	288
Pre-processing	Resize video from 1920x1280 to 320x180 resolution

Table 21: Dataset_6 consists of video files.

Figure 16: Hockey Events images

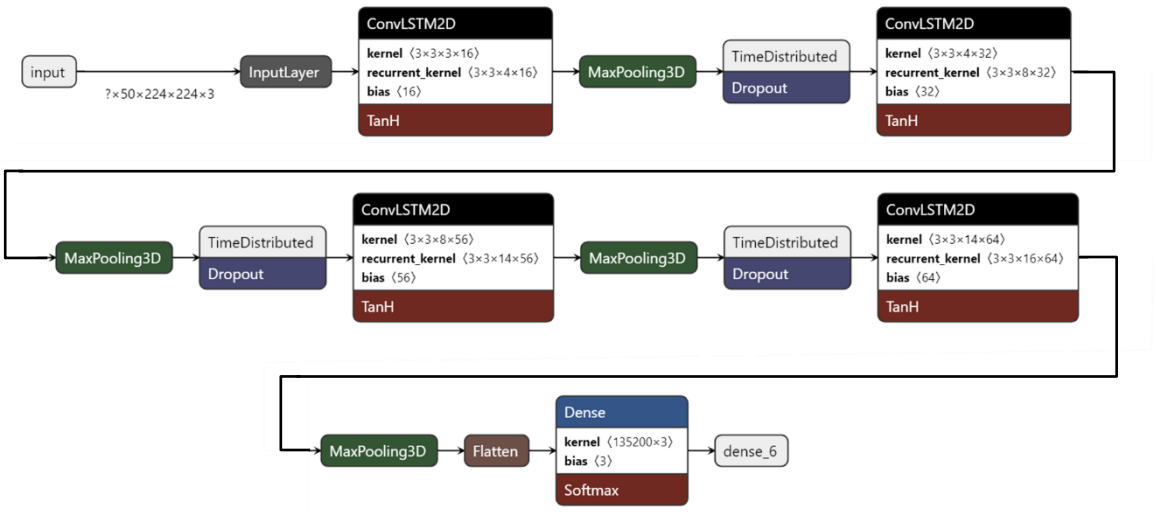


Figure 17: Proposed ConvLSTM based model for Hockey event detection.

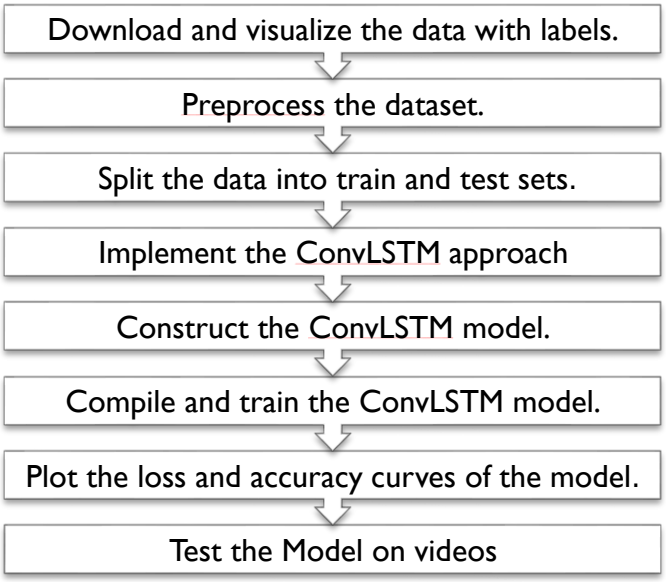


Figure 18 : Flow chart for Event Detection

Class	Precision	Recall	F1-score	Support	Accuracy
Penalty_Corner	0.67	0.61	0.64	23	67.00%
Penalty_Stroke	0.83	0.56	0.67	18	
Field_Goal	0.62	0.77	0.69	31	

Table 22: Accuracy of Proposed Model : ConvLSTM (Epoch = 28 (Early stopping))

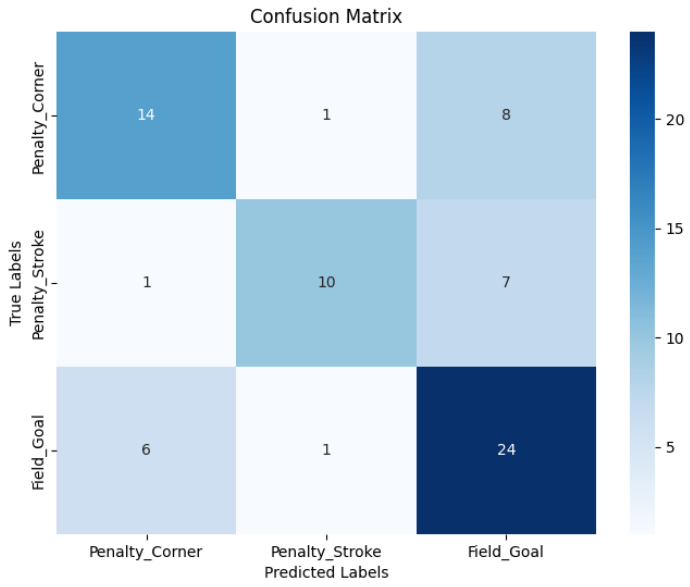


Figure 19: Confusion Matrix of Proposed Model : ConvLSTM

6.2.6 Proposed LRCN based Event Detection Model for Video Dataset

LRCN (Long-term Recurrent Convolutional Networks) is an architecture that combines convolutional neural networks (CNNs) and long short-term memory (LSTM) networks for video

classification tasks. It extends the capabilities of CNNs by incorporating temporal dependencies through the use of LSTMs[26].

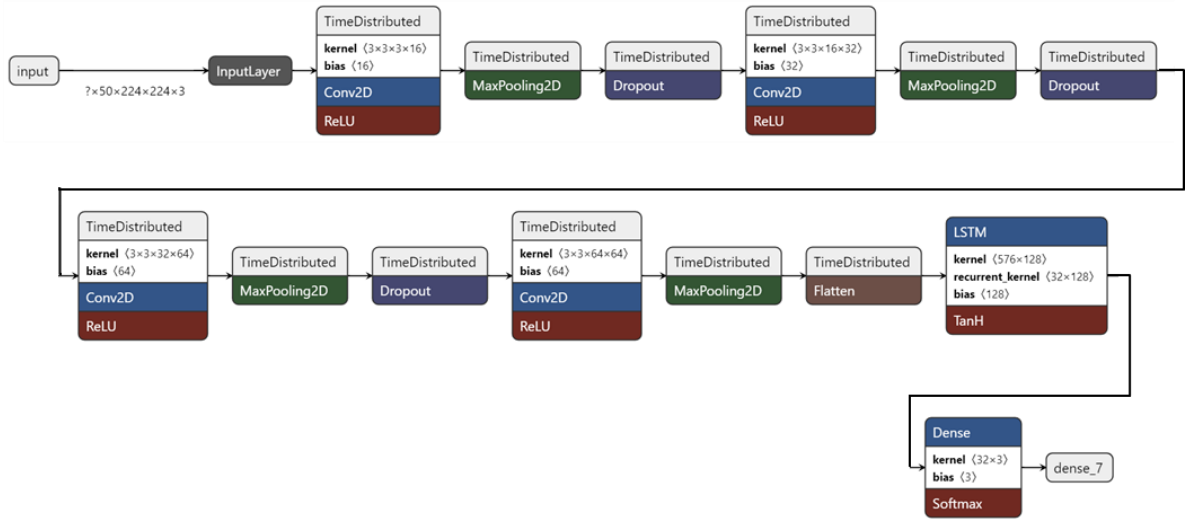


Figure 20: Proposed LRCN based model for Hockey event detection.

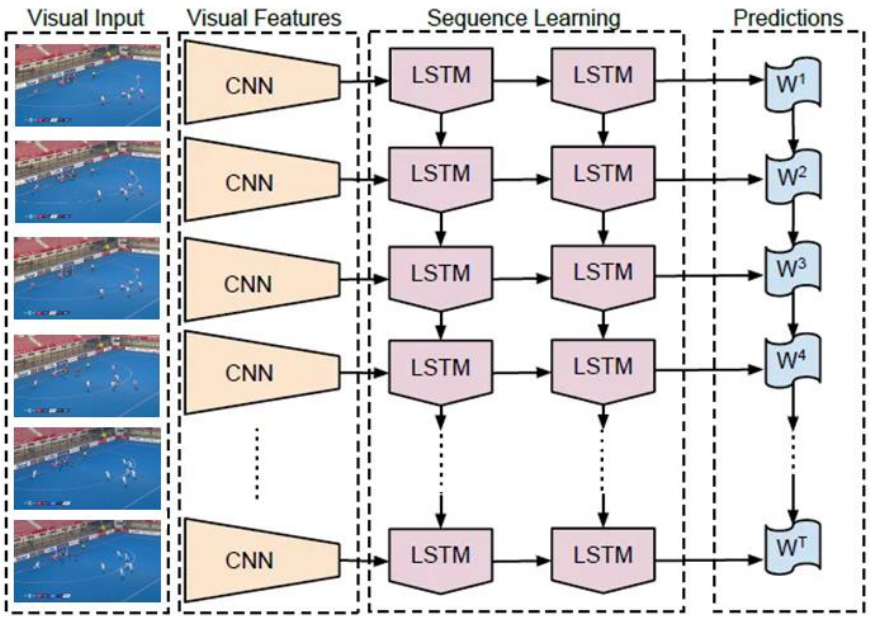


Figure 21 : LRCNs uses CNN and LSTM jointly for image description and video description [26].

	Precision	Recall	F1-score	Support	Accuracy
Penalty_Corner	0.59	0.7	0.64	23	58.00%
Penalty_Stroke	0.5	0.5	0.5	18	
Field_Goal	0.63	0.55	0.59	31	

Table 23: Accuracy of Proposed Model : LRCN (Epoch = 38 (Early stopping))

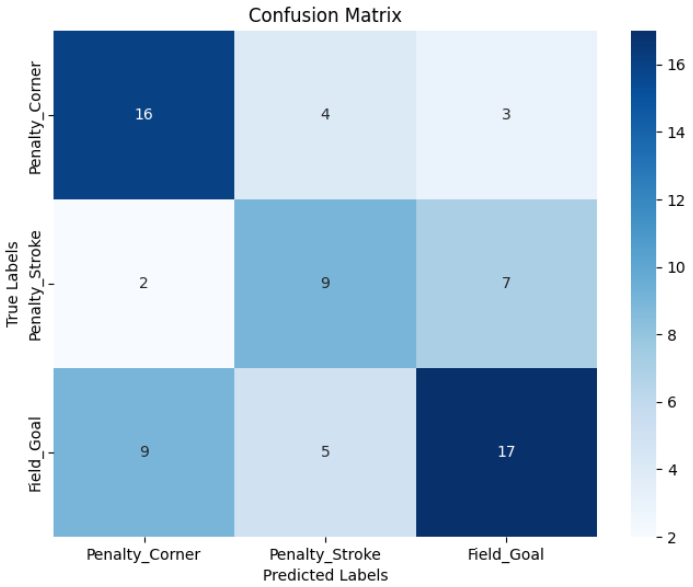


Figure 22: Confusion Matrix of Proposed Model : LRCN

Sr no.	Model	Epoch	Sequence and Image Size	No. of Images	Trainable Parameters	Precision (%)	Recall (%)	F1 Score (%)	Accuracy (%)
1	ConvLSTM	27*	50x224x224	14250	438,603	70.66	64.66	66.66	67%
2		100	50x224x224	14250	138,563	68.66	58	59	58.33%
3	LRCN	38*	50x224x224	14250	138,563	57.33	58.33	57.66	58%
4		100	50x224x224	14250	138,563	52	51.33	51.33	53%

Table 24: Performance comparison of ConvLSTM and LRCN models for (Dataset-3)

*The models were trained using the early stopping technique with a patience of 20. This means that if the model's performance does not improve for 20 consecutive epochs, the training process is stopped early. The maximum accuracy achieved for the ConvLSTM model was obtained at epoch 27, while for the LRCN model, it was achieved at epoch 38.

7. Achievements with respect to objectives

1) Successfully developed efficient object detection models for Dataset_1, Dataset_2 and Dataset_3 using various versions of YOLO based State of the art methods.

Sr No.	Model	Performance			Highest Accuracy
		PRECISION	RECALL	F-1 SCORE	
1	YOLOv3	0.95	0.93	0.94	93.30%
2	Scaled-YOLOv4	0.81	0.902	0.854	88.60%
3	YOLOv5	0.948	0.933	0.94	93.60%
4	MT-YOLOv6	0.74	0.62	0.675	73.99%
5	YOLOv7	0.861	0.857	0.859	84.10%
6	YOLOv8	0.949	0.93	0.939	**94.00%**

Table 25: Performance of object detection models for Dataset_1 (Classes: AUS (Team 1), BEL (Team 2), Hockey ball, and Umpire.)

The models' performance for object detection was evaluated for Dataset_1 based on precision, recall, and F-1 score. Among the models assessed, YOLOv8 exhibited the highest accuracy, achieving an impressive 94.00% accuracy with a precision of 0.949, recall of 0.93, and an F-1 score of 0.939. Other models like YOLOv3, YOLOv5, and Scaled-YOLOv4 also demonstrated robust performances with accuracy ranging from 88.60% to 93.60%. These results offer valuable insights into the strengths and capabilities of each model for the specified detection tasks.

2) Implemented an effective image classification model employing ten distinct deep learning techniques for event detection in hockey videos. The utilization of Rolling Average Prediction further enhanced the accuracy of detection outcomes. For Dataset_3 the Model-1 (VGG16) achieved remarkable performance with 99.33% precision, recall, and F1 score, resulting in 99.47% accuracy. Similarly, Model-II (VGG19) demonstrated high precision and recall of 97.67%, yielding 97.50% accuracy. Model-III (ResNet50) achieved balanced performance with 96.33% precision, recall, and F1 score, leading to 96.44% accuracy. Transitioning to Model-IV and Model-V (InceptionV3 and MobileNet), precision and recall slightly decreased, affecting overall accuracy. Model-VI (DenseNet121) exhibited 86.00% precision and 81.67% recall, resulting in 81.69% accuracy. Model-VII (Xception) achieved 76.33% precision and 74.00% recall, with 74.44% accuracy. Model-VIII, a Cascaded CNN with more trainable parameters, achieved 94.67% precision, recall, and F1 score, showcasing robustness in event detection. Model-IX combined Densenet with Transformer, achieving exceptional performance with 99.67% precision, recall, and 99.47% accuracy. Lastly, Model-X (Inceptionv3 with autoencoder) demonstrated balanced metrics of 99.00%, attaining 99.08% accuracy. The YOLOv8 model, pretrained on ImageNet, performed well for image classification and event detection from hockey videos on Dataset_4.

Overall, this performance analysis offers insights into model strengths and capabilities, aiding selection based on specific event detection needs.

Sr no.	Input Video File	Actual event	Event detected. (Prediction Size = 64)
1	g_m14_slow_part(18).mp4	Goal	Penalty Corner
2	g_m15_slow_part(10).mp4	Goal	Penalty Corner
3	g_m16_slow_part(21).mp4	Goal	Goal
4	G2_m3_slow_part1(3).mp4	Goal	Goal
5	G2_m10_slow_part_(20).mp4	Goal	Goal
6	G2_m12_slow_part_(3).mp4	Goal	Goal
7	G2_m13_slow_part_(20).mp4	Goal	Goal
8	G3_m1_slow_part14.mp4	Goal	Goal
9	G3_m7_slow_part_(30).mp4	Goal	Goal

10	goal_m13_slow_part_(7).mp4	Goal	Goal
11	goal_m16_slow_part(22).mp4	Goal	Goal
12	m27_part3.mp4	Penalty Corner	Penalty Corner
13	m27_part5.mp4	Penalty Corner	Penalty Corner
14	m27_part7.mp4	Penalty	Penalty
15	PC_m6_slow_part(20).mp4	Penalty Corner	Penalty Corner
16	PC_m8_slow_part_(5).mp4	Penalty Corner	Penalty Corner
17	penalty_m3_slow_part1(19).mp4	Penalty	Penalty
18	penalty_m7_slow_part_(44).mp4	Penalty	Penalty

Table 26 : Performance of Proposed VGG-16 based Model-I + Rolling Average Prediction for various input

3) We have successfully created effective event recognition models for Dataset_3 using two different approaches: ConvLSTM and LRCN. The ConvLSTM model demonstrated significant performance, achieving a precision of 70.66%, recall of 64.66%, and an F1 score of 66.66%, resulting in an overall accuracy of 67%. Conversely, the LRCN model exhibited slightly lower but still promising results, with a precision of 57.33%, recall of 58.33%, and an F1 score of 57.66%, leading to an accuracy of 58%. These models contribute to accurate event detection in videos, offering insights into specific actions and occurrences within the dataset.

Sr no.	Video Input	Actual event	ConvLSTM based Model	LRCN based Model
			(Predicted event)	(Predicted event)
1	goal_m13.mp4	Field_Goal	Field_Goal	Field_Goal
2	goal_m16.mp4	Field_Goal	Field_Goal	Field_Goal
3	m27_part3.mp4	Penalty_Corner	Field_Goal	Field_Goal
4	m27_part5.mp4	Penalty_Corner	Field_Goal	Penalty_Corner
5	m27_part7.mp4	Penalty_Stroke	Penalty_Stroke	Field_Goal
6	PC_m8.mp4	Penalty_Corner	Penalty_Corner	Penalty_Corner
7	PC_m8.mp4	Penalty_Corner	Penalty_Corner	Penalty_Corner
8	penalty_m3.mp4	Penalty_Stroke	Penalty_Stroke	Penalty_Stroke
9	Penalty_m7.mp4	Penalty_Stroke	Penalty_Stroke	Penalty_Stroke

Table 27: Performance of ConvLSTM and LRCN models for various input

8. Conclusion

In conclusion, this study introduces the utilization of diverse iterations of the YOLO model for effective object detection within the realm of field hockey. The models aptly identify four principal entities: AUS (Team 1), BEL (Team 2), Hockey ball, and Umpire, from the collected hockey dataset (Dataset_1), achieving accuracy levels ranging from 88.60% to 94.00%, with the YOLOv8 model demonstrating the highest accuracy. Furthermore, we assessed the YOLOv8 model's performance across various classes and model sizes, uncovering that increased model dimensions enhance object detection metrics across all categories. Notably, the "YOLOv8X" model size yielded superior precision, recall, F1 score, and overall accuracy, although the balance

between model performance and size, considering computational resources, is a crucial consideration.

For event detection via image classification, Model-I based on VGG16, along with Model-IX combining Densenet with Transformer, showcased exceptional performance among ten evaluated models. These models achieved remarkable precision, recall, F1 score, and an accuracy of approximately 99.33% on Dataset_3, comprising 3035 images. Moreover, the YOLOv8 model, pre-trained on ImageNet, demonstrated commendable performance in image classification and event detection within hockey videos on Dataset_4, encompassing 7195 images. This emphasizes the effectiveness of deep learning models in capturing and analyzing the visual features required for precise activity recognition in the dynamic sport of hockey. The construction of a domain-specific dataset plays a pivotal role in the success of activity recognition models, and our carefully curated dataset of annotated field hockey videos frames serves as a valuable resource for further advancements in this area. The practical implications of our research hold great significance for stakeholders within the hockey domain.

In the realm of video classification, the ConvLSTM model emerged as a superior performer, surpassing the LRCN model in accuracy on Dataset_5, which shared similar numbers of videos and input configurations. Our research demonstrates the effectiveness of deep learning models for hockey object and event recognition. We have presented a comprehensive evaluation of our approach, achieving exceptional accuracy in classifying hockey object and activities. The construction of a domain-specific dataset further reinforces the reliability and applicability of our findings. Although our research has provided valuable insights and practical implications, there are still avenues for future exploration and improvement. Further work can be conducted to expand the dataset, explore fine-grained activity recognition, enable real-time recognition, and investigate multi-modal fusion approaches.

Overall, our research contributes to the field of hockey object and event recognition and lays the groundwork for further advancements in analyzing and comprehending the intricate dynamics of field hockey. We hope that our work serves as an inspiration for future research and applications in this domain, ultimately benefiting players, coaches, analysts, and hockey enthusiasts.

9. Publications

1. S. H. Patel and D. Kamdar, “OBJECT DETECTION IN HOCKEY SPORT VIDEO VIA PRETRAINED YOLOV3 BASED DEEP LEARNING MODEL,” ICTACT J. IMAGE VIDEO Process., vol. 13, no. FEBRUARY, pp. 2893–2898, 2023, doi: 10.21917/ijivp.2023.0412. (ISSN online: 0976-9102) (**UGC-CARE Indexed**)

2. S. H. Patel and D. Kamdar, “Accurate ball detection in field hockey videos using YOLOV8 algorithm,” Int. J. Adv. Res. Ideas Innov. Technol., vol. 9, no. 2, pp. 411–418, 2023, [ISSN Online: 2454-132X]. Available: https://www.ijariit.com/manuscripts/v9i2/V9I2-1305.pdf

3. Accepted for Publication: Title: “Survey on Sport Video Analysis and Event Detection” Journal: Int. J. of Autonomous and Adaptive Communications Systems (ISSN online:1754-8640) (**Scopus Indexed**)

4. Accepted for Publication: Title: “DEEP LEARNING APPROACH FOR EVENT RECOGNITION IN FIELD HOCKEY VIDEOS”. Journal: Reliability: Theory & Applications (ISSN 1932-2321) (**Scopus Indexed)**

10. References:

[1] S. Moon, J. Lee, D. Nam, W. Yoo, and W. Kim, “A comparative study on preprocessing methods for object tracking in sports events,” in *2018 20th International Conference on Advanced Communication Technology (ICACT)*, IEEE, Feb. 2018, pp. 460–462. doi: 10.23919/ICACT.2018.8323794.

[2] L. Ballan, M. Bertini, A. Del Bimbo, L. Seidenari, and G. Serra, “Event detection and recognition for semantic annotation of video,” *Multimed. Tools Appl.*, vol. 51, no. 1, pp. 279–302, Jan. 2011, doi: 10.1007/s11042-010-0643-7.

[3] Y. Rui, A. Gupta, and A. Acero, “Automatically extracting highlights for TV Baseball programs,” in *Proceedings of the eighth ACM international conference on Multimedia*, New York, NY, USA: ACM, Oct. 2000, pp. 105–115. doi: 10.1145/354384.354443.

[4] A. Kokaram *et al.*, “Browsing sports video: trends in sports-related indexing and retrieval work,” *IEEE Signal Process. Mag.*, vol. 23, no. 2, pp. 47–58, 2006.

[5] T. D’Orazio and M. Leo, “A review of vision-based systems for soccer video analysis,” *Pattern Recognit.*, vol. 43, no. 8, pp. 2911–2926, Aug. 2010, doi: 10.1016/j.patcog.2010.03.009.

[6] M. R. Naphade, T. Kristjansson, B. Frey, and T. S. Huang, “Probabilistic multimedia objects (multijects): A novel approach to video indexing and retrieval in multimedia systems,” in *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No. 98CB36269)*, 1998, pp. 536–540. [Online]. Available: http://oak.cs.ucla.edu/~mjwelch/multimedia/papers/multijects.pdf

[7] C.-H. Liang, W.-T. Chu, J.-H. Kuo, J.-L. Wu, and W.-H. Cheng, “Baseball event detection

using game-specific feature sets and rules," in *2005 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2005, pp. 3829–3832.

[8] Cheolkon Jung and Joongkyu Kim, "Player Information Extraction for Semantic Annotation in Golf Videos," *IEEE Trans. Broadcast.*, vol. 55, no. 1, pp. 79–83, Mar. 2009, doi: 10.1109/TBC.2008.2010377.

[9] G. Yao, T. Lei, and J. Zhong, "A review of Convolutional-Neural-Network-based action recognition," *Pattern Recognit. Lett.*, vol. 118, pp. 14–22, 2019, doi: 10.1016/j.patrec.2018.05.018.

[10] R. G. Abbott and L. R. Williams, "Multiple target tracking with lazy background subtraction and connected components analysis," *Mach. Vis. Appl.*, vol. 20, no. 2, pp. 93–101, Feb. 2009, doi: 10.1007/s00138-007-0109-8.

[11] a Lehuger, "A robust method for automatic player detection in sport videos 2 System Architecture 1 Introduction 3 Training Methodology 4 Player Localization," *Analysis*, 2007, [Online]. Available: https://www.lirmm.fr/coresa2007/PDF/02.pdf

[12] S. Maćkowiak, M. Kurc, J. Konieczny, and P. Maćkowiak, "A complex system for football player detection in broadcasted video," *Int. Conf. Signals Electron. Syst. ICSES'10 - Conf. Proceeding*, pp. 119–122, 2010, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/5595236

[13] D. Zhang, "Vehicle target detection methods based on color fusion deformable part model," *Eurasip J. Wirel. Commun. Netw.*, vol. 2018, no. 1, pp. 0–5, 2018, doi: 10.1186/s13638-018-1111-8.

[14] V. Pallavi, J. Mukherjee, A. K. Majumdar, and S. Sural, "Ball detection from broadcast soccer videos using static and dynamic features," *J. Vis. Commun. Image Represent.*, vol. 19, no. 7, pp. 426–436, 2008, doi: 10.1016/j.jvcir.2008.06.007.

[15] M. Leo, P. L. Mazzeo, M. Nitti, and P. Spagnolo, "Accurate ball detection in soccer images using probabilistic analysis of salient regions," *Mach. Vis. Appl.*, vol. 24, no. 8, pp. 1561–1574, 2013, doi: 10.1007/s00138-013-0518-9.

[16] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Prog. Artif. Intell.*, vol. 9, no. 2, pp. 85–112, 2020, doi: 10.1007/s13748-019-00203-0.

[17] J. Dwyer, B., Nelson, J. (2022), Solawetz, "et. al.," *Roboflow (Version 1.0) [Software]*.

[18] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.02767

[19] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: Scaling Cross Stage Partial Network," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2021, pp. 13024–13033. doi: 10.1109/CVPR46437.2021.01283.

[20] G. Jocher *et al.*, "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation," Nov. 2022, doi: 10.5281/ZENODO.7347926.

[21] C. Li *et al.*, "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications," Sep. 2022, [Online]. Available: http://arxiv.org/abs/2209.02976

[22] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," pp. 1–15, Jul. 2022, [Online]. Available: http://arxiv.org/abs/2207.02696

[23] S. Oprea *et al.*, “A Review on Deep Learning Techniques for Video Prediction,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 2806–2826, Jun. 2022, doi: 10.1109/TPAMI.2020.3045007.

[24] G. Jocher *et al.*, “ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation.” Zenodo, Nov. 2022. doi: 10.5281/zenodo.7347926.

[25] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, and W. Woo, “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting,” *Adv. Neural Inf. Process. Syst.*, vol. 2015-Janua, pp. 802–810, Jun. 2015, [Online]. Available: http://arxiv.org/abs/1506.04214

[26] J. Donahue *et al.*, “Long-Term Recurrent Convolutional Networks for Visual Recognition and Description,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 677–691, Apr. 2017, doi: 10.1109/TPAMI.2016.2599174.