

DEEP LEARNING APPROACH FOR EVENT RECOGNITION IN FIELD HOCKEY VIDEOS

●
Suhas H. Patel^{1*}, Dr. Dipesh Kamdar^{2*}, Dr. D. D. Vyas³, Dr. Prakash P. Patel⁴

¹Research Scholar, Gujarat Technological University, Gujarat, India
E-mail: suhas048@gmail.com

²Assistant Professor, Electronics and Communication Engg. Department, V.V.P. Engineering College, Rajkot, Gujarat, India
E-mail: kamdardipesh@gmail.com

³Dean, Transformative Academics, Atmiya University, Rajkot, Gujarat, India
E-mail: vyasdd@yahoo.com

⁴Principal, V.P.M.P Polytechnic, Gandhinagar, Gujarat, India
E-mail: ecprakashpatel@gmail.com

Abstract

The objectives of this research are to develop a deep learning approach for event recognition in field hockey videos, construct a dataset that includes important activities in field hockey such as goals, penalty corners, and penalty, and evaluate the performance of the approach using the constructed dataset. By achieving these objectives, the research aims to improve the accuracy and effectiveness of event recognition in the fast-paced and complex domain of field hockey videos. 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. To facilitate this process, a dataset is constructed, consisting of labeled instances of key activities in field hockey, namely goals, penalty corners, and penalty. The performance of the approach is then evaluated using this carefully prepared 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. Furthermore, the results demonstrate the potential of this approach in various field hockey applications, including performance analysis, coaching, and video replay. The accurate recognition of events opens new possibilities for leveraging field hockey videos for enhanced analysis, coaching strategies, and engaging video presentations. The novelty of this research lies in the introduction of a deep learning approach specifically designed for event recognition in field hockey videos. Unlike traditional methods, this approach leverages the power of deep learning, particularly a pretrained CNN, to improve the accuracy of event recognition. Additionally, the construction of a domain-specific dataset addresses the limitation of existing field hockey datasets and enhances the effectiveness of the approach. The remarkable accuracy achieved in event recognition further emphasizes the novelty and potential of this approach in the field of field hockey video analysis.

Keywords: Event recognition, field hockey videos, deep learning, convolutional neural network (CNN), VGG16.

1. INTRODUCTION

Field hockey is a fast-paced and dynamic sport, requiring players to showcase their skills in a highly competitive environment. Event recognition in field hockey videos is a crucial task for extracting valuable insights from the gameplay, enabling performance analysis, coaching, and video replay. Latest computer vision techniques applied in sports video analysis, encompassing player and ball

tracking, trajectory prediction, skill analysis, team strategy assessment, and object detection and classification in sports [1]. Traditional methods for event recognition, such as rule-based systems and motion analysis, often fall short in accurately identifying events due to the sport's complex nature and rapid movements. Some research has prioritized semantic event detection for its capacity to generate insightful outcomes, including pattern recognition and team strategy analysis, while combining video processing, computer vision, and machine learning[2]. This integration offers substantial potential in the sports entertainment domain, enhancing referee decision-making and providing sports fans with improved systems for match analysis. Deep learning techniques have shown significant advancements in event recognition across different sports, offering promising results. One crucial aspect for effective event recognition is the ability of the network to learn high-level features that capture human actions and the contextual scene information[3]. This requires two key factors: Adequate input image size and Network Depth. The input image size should be sufficiently large to enable the network to capture fine-grained details and extract meaningful features related to the sports events. Deep neural networks are essential for learning complex and abstract representations from the input data. A deeper network architecture allows for the extraction of hierarchical features, leading to improved event recognition performance. By incorporating large input image sizes and deep network architectures, deep learning models can effectively learn high-level features that facilitate accurate event recognition in sports. These advancements have contributed to significant improvements in event recognition across various sports domains. Combining convolutional and recurrent neural networks enables the analysis of sports video sequences and yields experimental results[4]. However, sports analytics face challenges and unresolved issues, especially in data collection and labeling, as well as the complexity of recognizing fast actions and analyzing multiple players' involvement in team sports like football and basketball[5]. In this research paper, we present a deep learning-based approach for event recognition in field hockey videos. Our approach leverages the power of pretrained convolutional neural networks (CNNs) to train a classifier capable of accurately identifying different events. By exploiting the knowledge learned from large-scale datasets in related domains, the pretrained CNN captures rich visual representations that are crucial for distinguishing between various field hockey events. One of the challenges we encountered was the lack of existing field hockey datasets suitable for event recognition. To address this, we created our own dataset, specifically designed for field hockey, consisting of three primary activities: goals, penalty corners, and penalty. This dataset enables us to train and evaluate the performance of our deep learning approach in a realistic field hockey scenario. Through extensive evaluations, we demonstrate the effectiveness of our approach in accurately recognizing events in field hockey videos.

Our approach achieves an impressive accuracy of 99.47% on our self-prepared dataset, highlighting its potential for real-world applications in field hockey analysis and coaching. The contributions of this research paper extend beyond event recognition in field hockey. The remaining sections of the paper are structured as follows. Section 2 presents an overview of related work, focusing on event recognition and the application of deep learning in sports. Section 3 outlines the methodology employed for event recognition in field hockey videos. The results and analysis are presented in Section 4. Lastly, Section 5 concludes the paper by summarizing our contributions and highlighting future research directions.

2. RELATED WORK

Event recognition in sports videos has been a topic of extensive research in recent years. Various approaches have been explored to tackle the challenges associated with accurately identifying events in dynamic sporting environments. In this section, we provide an overview of related work in the field of event recognition and highlight the contributions of deep learning methods in sports

analysis.

2.1 Traditional Approaches for Event Recognition

Traditional approaches for event recognition in sports videos often rely on handcrafted features and rule-based systems. These methods involve manually designing features based on domain knowledge and utilizing predefined rules to detect specific events. For field hockey, these rules might include analyzing the positions and movements of players, the trajectory of the ball, or specific gameplay patterns. While these approaches have been effective to some extent, they often struggle to handle the complexity and variability of events in dynamic sports like field hockey. The manual design of features and rules limits their adaptability to different scenarios and may result in suboptimal performance.

In a series of research papers, traditional approaches have been proposed for activity recognition and detection in sport videos. One study introduces Histograms of Oriented Gradients (HOG) for player representation and combines it with a probabilistic framework and multi-class sparse classifier for action recognition [6]. Another paper focuses on evaluating action recognition approaches for fight detection, presenting a new fight dataset, and achieving high accuracy in detecting fights[7]. Hierarchical poselets are introduced in another study, enabling human pose modeling and serving as an intermediate representation for action recognition[8]. A violence detection method utilizing the MoSIFT algorithm and sparse coding is proposed in a different research, addressing the limitations of traditional descriptors and achieving promising results on challenging datasets[9]. Additionally, a novel approach using the Markov Game formalism is presented to value player actions in ice hockey, considering context and lookahead [10]. Lagrangian measures are employed in another paper for violent video detection, outperforming other local features in detecting violence [11]. Lastly, a technique using histogram of oriented gradients and local binary pattern features is presented for accurate recognition of basketball referees' signals in game videos[12]. These studies collectively contribute to the field of video-based activity recognition, offering insights and advancements in various aspects such as player representation, violence detection, pose modeling, action valuation, and gesture recognition in sports videos. Table 1 provides a comprehensive list of traditional sport event detection models.

Table 1: *Traditional event detection models for various sport categories.*

Reference	Problem statement	Proposed method	Sports
[6]	Track and identify the actions of multiple hockey players.	Histograms of Oriented Gradients (HOG), boosted particle filter (BPF)	Ice Hockey
[7]	detection of fights or aggressive behaviors in ice hockey sport videos	Space-Time Interest Points (STIP), Motion Scale-Invariant Feature Transform (MoSIFT)	Ice Hockey
[8]	human parsing and action recognition from static images	hierarchical Poselets	Multiple sports
[9]	detect violence in videos with crowded and non-crowded scenes	MoSIFT and sparse coding.	Ice Hockey
[10]	Assessing player actions in ice hockey.	Markov Game formalism	Ice Hockey

[11]	to detect violent scenes in videos	Lagrangian Scale Invariant Feature Transform (LaSIFT)	Ice Hockey
[12]	recognize the signals of basketball referees from recorded game videos.	histogram of oriented gradients +SVM, local binary pattern features +SVM	Basketball

2.2 Deep Learning in Event Recognition in Sports

Deep learning has emerged as a powerful paradigm for event recognition in sports videos. Convolutional Neural Networks (CNNs) have demonstrated remarkable success in extracting meaningful representations from visual data. CNNs can automatically learn hierarchical features by employing multiple layers of convolutions and nonlinear activations, enabling them to capture complex patterns and spatial dependencies.

In the context of sports event recognition, deep learning models have shown superior performance by leveraging large-scale annotated datasets and pretraining on related tasks. By utilizing pretrained CNNs, such as those trained on ImageNet, the models can capture generic visual representations that are transferable to sports-specific tasks. Fine-tuning or retraining the pretrained models on specific sports datasets further enhances their ability to recognize events accurately.

One paper proposes a novel framework for soccer video event detection, utilizing 3D convolutional networks and shot boundary detection, and introducing temporal action localization and play-break rules [13]. Another study focuses on action recognition in hockey, introducing the ARHN architecture and achieving high accuracy by leveraging pose information[14]. A methodology for fight scene detection in hockey videos is proposed in a different paper, using blur, radon transform, and convolutional neural networks [15]. Furthermore, a 3D CNN-based multilabel deep HAR system is presented for hockey video action recognition, outperforming existing solutions[16]. Another paper introduces a two-stream architecture for hockey action recognition, combining pose estimation and optical flow[17]. A modified 3D ConvNet is proposed for violent video detection, achieving competitive results with improved strategies[18]. An automated activity recognition model for hockey matches using deep learning is presented, achieving a high accuracy of 98% [19]. In cricket, a hybrid deep-neural-network architecture is proposed for shot classification [20]. Puck localization in hockey videos is addressed using a network that incorporates expert annotations and temporal context [21]. Lastly, a deep learning method for event detection in football videos achieves superior precision and recall [22]. These papers collectively contribute to advancing activity recognition in sports videos, offering novel frameworks, architectures, and methodologies for accurate and efficient detection and recognition of various actions and events. Table 2 lists various sport event detection models based on deep learning.

Table 2: *Deep learning-based event detection model for various sport categories.*

Reference	Problem statement	Proposed method	Sports
[13]	Soccer video event detection	3D Convolutional Networks +Deep Feature Distance	Soccer
[14]	Interpreting player actions in ice hockey videos	Action Recognition Hourglass Network (ARHN)	Ice Hockey
[15]	Detecting fight scenes in hockey sport videos	Feed forward neural network and VGG16-Net	Ice Hockey
[16]	Multi-label class-imbalanced action recognition in hockey videos.	3D CNN based multilabel deep HAR system	Ice Hockey
[17]	Action recognition in ice hockey	two-stream architecture	Ice

			Hockey
[18]	violent video detection in ice hockey	Modified 3D ConvNet	Ice Hockey
[19]	Field hockey activity recognition	VGG-16	Field Hockey
[20]	The task involved classifying 10 different cricket batting shots from offline videos.	CNN+GRU	Cricket
[21]	Puck localization and event recognition in broadcast hockey videos,	CNN	Ice Hockey
[22]	Event detection in football videos	InceptionV2, 3DCNN	Soccer

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.

In summary, while traditional approaches for event recognition in sports videos have been explored, deep learning methods have shown significant promise in improving event recognition accuracy. The utilization of pretrained CNNs and transfer learning techniques enables these models to learn rich visual representations and adapt to specific sports domains. In the case of field hockey, there is a need for further research and development of deep learning-based approaches tailored to the unique characteristics of the sport. Our proposed approach aims to address this gap and contribute to the advancement of event recognition in field hockey videos.

3. METHODOLOGY

3.1 Field Hockey Dataset

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 28 highlights videos from the tournaments of the hockey pro league for the years 2021-22 and 2022-23. These videos showcase the most remarkable moments and thrilling gameplay sequences from various field hockey games played between two teams. By carefully analyzing these highlights videos, we were able to identify and extract important events such as goals, penalty corners, and penalty. These events represent significant turning points in the matches and provide valuable data for building a comprehensive hockey event detection dataset. The utilization of highlights videos ensures that the dataset captures the most exciting and impactful moments from the field hockey games. This enables the development and evaluation of event detection models on key events that greatly influence the outcome of matches and attract the attention of viewers. By leveraging the information extracted from these videos, we aim to construct a high-quality field hockey event detection dataset that can be used for training and evaluating event recognition models. To create the ground truth annotations, we meticulously watched and manually labeled the videos. Each event of interest, such as goals, penalty corners, penalty, and other relevant actions, was annotated with their respective start and end timestamps. The dataset is designed to be representative of the challenges encountered

in real-world field hockey scenarios. It includes enough positive samples for each event category, ensuring a balanced distribution for training and evaluation purposes. Figure 1 illustrates the sequential video frames used for hockey event recognition.

Table 3: Hockey Event Recognition Dataset

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%)

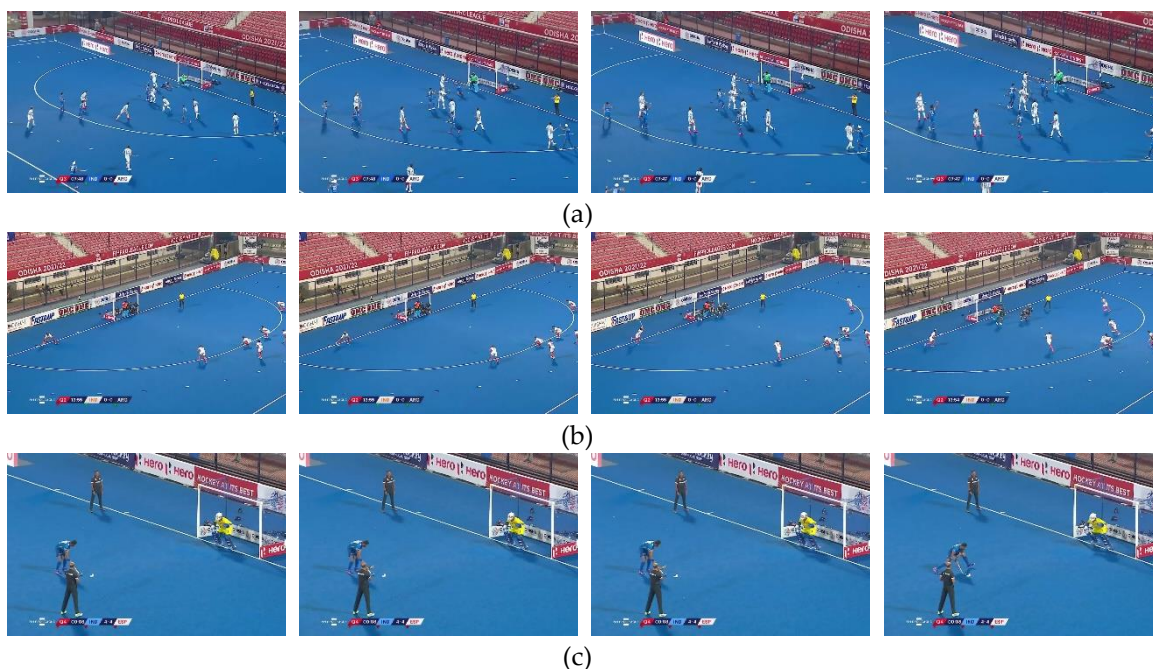


Figure 1: Sequential video frames for (a) Goal, (b) Penalty Corner, (c) Penalty

3.2 Model Architecture

In our approach, we utilize a pretrained convolutional neural network (CNN) for feature extraction, as depicted in Figure 2. The model architecture consists of a CNN that has been pretrained on a large-scale image dataset, such as ImageNet, to learn generic visual representations. This pretrained CNN can capture low-level to high-level visual features, making it well-suited for recognizing complex events in field hockey videos. The process begins with the application of the "flatten" operation, which reshapes the output of the pretrained CNN into a one-dimensional vector. This step enables easier processing and subsequent layers in the model. Following the flattening, a "dense512" layer is introduced. This layer is a fully connected layer with 512 units and applies the

rectified linear unit (ReLU) activation function to introduce non-linearity to the network. To mitigate the risk of overfitting, a "dropout(0.5)" layer is included. During training, this layer randomly sets 50% of the values to 0, effectively disabling certain connections between neurons. By doing so, it helps prevent the model from relying too heavily on specific features and enhances its generalization capability. The subsequent layer in the model is a "dense(3)" layer, which is used for multi-class classification. This layer consists of 3 units, representing the number of classes, and applies the softmax activation function to produce class probabilities. It is through this layer that the model assigns probabilities to each class, indicating the likelihood of the input frame belonging to a particular class. To summarize, the model architecture encompasses a pretrained CNN base, the flattening operation, a dense layer with ReLU activation, dropout regularization, and a dense layer with softmax activation for classification. These operations constitute a deep learning model configuration commonly employed in classification tasks. Moreover, Figure 8 provides an overview of the model architecture and its components. During the inference stage, each video frame is passed through the pretrained CNN, and activations from one of the intermediate layers are extracted. These activations represent the specific visual features learned by the model for that frame. By considering multiple frames within a temporal window, we capture the temporal dynamics of events, enabling a comprehensive understanding of the evolving actions in the field hockey videos.

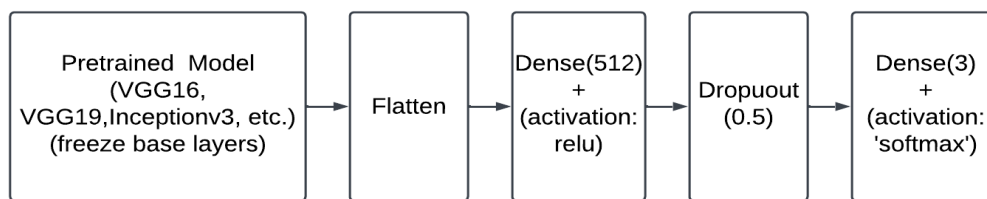


Figure 2: Overview of different pretrained model architectures

3.3 Model Training and Evaluation

To train our deep learning-based event recognition system, we split the annotated dataset into training and test sets. The dataset used in this study, as shown in Table 3, consists of a total of 3,035 images that are classified into three distinct classes. All images in the dataset have been fully annotated, ensuring that there are no unannotated instances. The training set comprises 2,276 images, which corresponds to 75% of the total dataset, while the remaining 759 images form the testing set, accounting for 25% of the dataset. On average, the images have a size of approximately 2.07 megapixels. The median image resolution is 1920x1080 pixels, indicating a consistent aspect ratio among the images. The class distribution within the dataset is as follows: The "Goal" class consists of 1,000 instances, representing 32.95% of the dataset. The "Penalty Corner" class comprises 1,017 instances, accounting for 33.51% of the dataset. Lastly, the "Penalty" class contains 1,018 instances, making up 33.54% of the dataset. The images are resized to a dimension of 224x224 pixels. The VGG16 model is pretrained on a dataset like ImageNet to learn features and predict labels. To adapt it for a new task, the top layers are replaced, and the base layers are frozen as a feature extractor. New layers are added on top, such as fully connected and dropout layers. The modified VGG16 model is then trained on the new dataset, optimizing its parameters using the training set. Cross-validation techniques can be used for robustness. The system's performance is evaluated on the test set, measuring event recognition accuracy and other metrics. Results are compared with baseline methods or alternative architectures to assess the effectiveness of the approach. The following pretrained models from Keras were utilized in this study, as shown in Table 4.

The event detection system is implemented on Google Colaboratory, which is a Python 3 environment, utilizing the GPU support provided by Google Compute Engine backend. In this

study, a pre-trained VGG16 based model-1 was utilized for event detection. The input images were resized to 224x224 pixels, and a batch size of 32 was used during training.

Table 4 : *Overview of Pretrained Networks*

Reference	Model	Description
[23]	VGG16	A deep convolutional neural network (CNN) with 16 layers, known for its simplicity and effectiveness.
[23]	VGG19	Like VGG16 but with 19 layers, providing a slightly deeper architecture.
[24]	ResNet50	A deep residual network with 50 layers, designed to address the vanishing gradient problem and enable training of very deep networks.
[25]	InceptionV3	A deep CNN architecture with multiple parallel branches, allowing for efficient feature extraction at different scales
[26]	MobileNet	A lightweight CNN architecture designed for mobile and embedded devices, balancing model size and accuracy.
[27]	DenseNet121	A densely connected CNN architecture that facilitates feature reuse and enables deeper networks without sacrificing performance.
[28]	Xception	An extension of the Inception architecture that replaces standard convolutions with depthwise separable convolutions, resulting in improved performance.

The model is trained for 100 epochs using Stochastic Gradient Descent (SGD) optimizer with specific parameters including a learning rate of 1e-4, momentum of 0.9, and a decay of 1e-4/100. The video frames representing hockey events are used as inputs to the fine-tuned VGG16 based model-1 that was specifically tailored for event detection in this study. Figure 3 illustrates the architecture of the proposed pretrained model framework.

Table 5 displays the modified model details, including the modules, output dimensions, and trainable parameters. The flowchart in Figure 5 illustrates the process of hockey event recognition using a deep learning model. It encompasses preprocessing the input video, training the model, evaluating its performance, and utilizing the model to predict events in a video clip. The output is a labeled video where events are assigned specific labels based on the model's predictions. We evaluate our event recognition system using standard evaluation metrics, including accuracy, precision, recall, and F1 score, to measure its performance. Accuracy represents the proportion of correctly classified events, while precision, recall, and F1 score assess the system's performance in identifying specific event categories. After training, we assess the system's accuracy, precision, recall, and F1 score on the test set and compare our results with alternative deep learning architectures. Table 6 presents the results of seven different architectures for hockey event detection on dataset, demonstrating the effectiveness and performance improvements of our proposed approach. The experimental results are visualized in Figure 6.

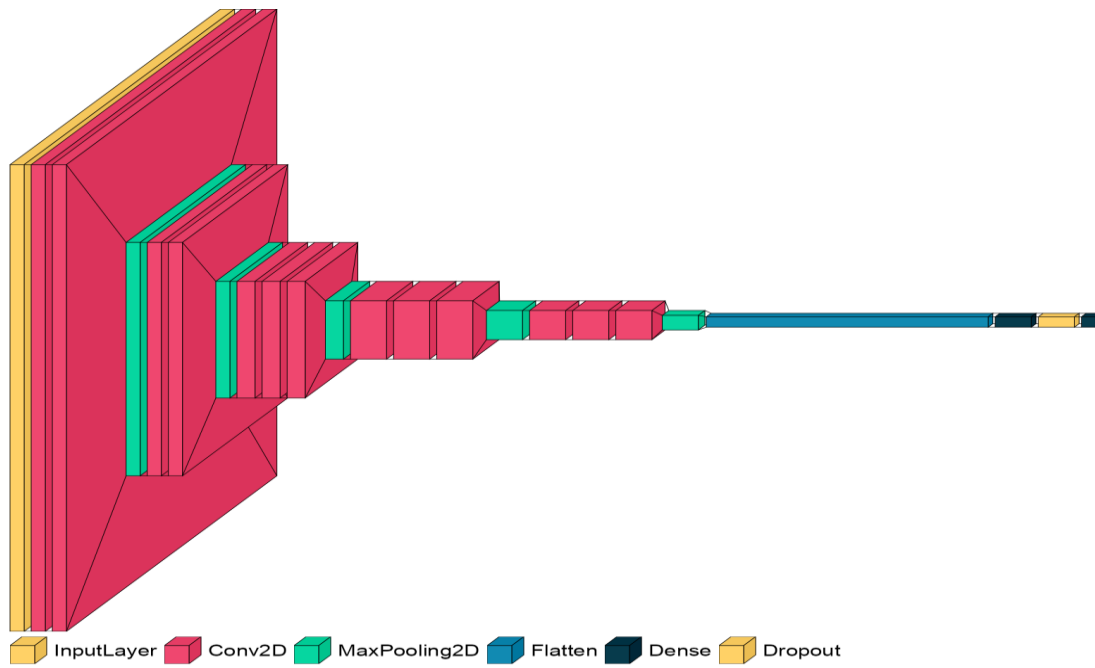


Figure 3: Model-1 Architecture.

Table 5: Model -1 details in terms of module, output dimension and trainable parameters.

Model: "model-1"		
Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 512)	12845568
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 3)	1539
Total params: 27,561,795		
Trainable params: 12,847,107		
Non-trainable params: 14,714,688		

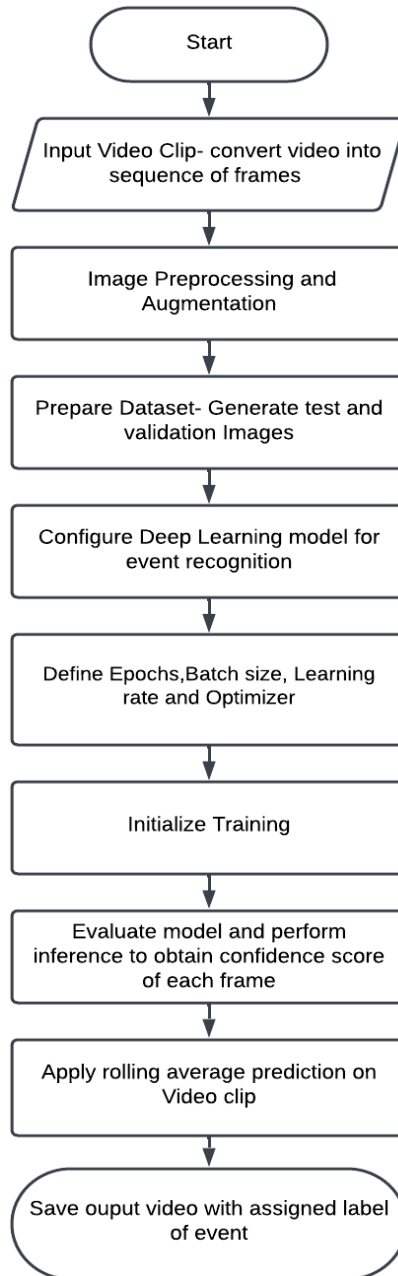


Figure 5 : Process of hockey event recognition using a deep learning mode

Table 6 : Fine-tuned Deep Learning model results.

Sr no.	Model Name	Pre-trained 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-2	VGG19	264,195	97.67	97.67	97.33	97.50
3	Model-3	ResNet50	1,050,627	96.33	96.33	96.33	96.44
4	Model-4	InceptionV3	4,196,355	84.67	83.67	84.00	83.66
5	Model-5	MobileNet	526,339	88.33	87.67	87.67	87.62
6	Model-6	DenseNet12	526,339	86.00	81.67	81.67	81.69
7	Model-7	Xception	1,050,627	76.33	74.00	74.00	74.44

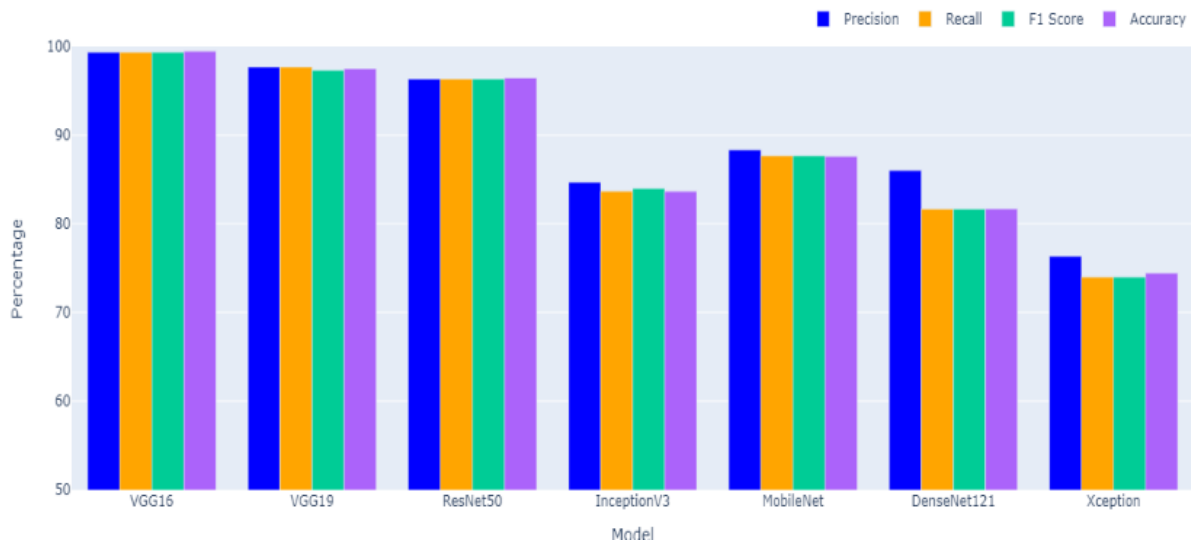


Figure 6: Comparison of Model-1 with other six models

Figure 7 illustrates the flowcharts that enable us to utilize the power of convolutional neural networks (CNNs) in analyzing video frames and making predictions for each frame. The flowchart starts with inputting the video frames into the CNN model. As the frames are processed through the model, predictions are generated for each frame, indicating the likelihood of a particular event occurring at that specific moment.

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[29]. This can be particularly beneficial when dealing with real-world scenarios where videos may contain motion blur, camera movement, or other factors that can introduce uncertainties in individual frame predictions. Overall, the flowcharts in figure 7, combined with the rolling average prediction technique, enable us to leverage the CNN's analytical capabilities to make reliable and stable predictions for the events taking place in the video, improving the overall accuracy and effectiveness of our event recognition system.

4. RESULTS AND ANALYSIS

4.1 Results

The model-1, with 12,847,107 trainable parameters, demonstrates exceptional performance in image classification. It achieves high precision, recall, F1 score, and accuracy, all at approximately 99.47%. These metrics indicate that the model excels in accurately classifying images with a high level of precision and recall.

Upon analyzing the results, it is evident that the model-1 outperforms other models across all evaluated metrics. It achieves the highest precision, recall, F1 score, and accuracy among the models considered. While model-2 and model-3 also show commendable performance, models such as model-3, model-4, model-5, and model-6 exhibit relatively lower performance in comparison. These findings highlight the effectiveness of the VGG16 based model-1 for image classification tasks, showcasing its superiority in accurately classifying images. Figure 8 illustrates the training and loss

accuracy of the VGG16 based model-1 up to 100 epochs, showcasing the model's learning progression over the training process. Figure 9 displays the confusion matrix of the proposed model-1 for the given dataset, providing insights into the model's performance in terms of classification accuracy. Figure 10 visually presents the output video frames of hockey event recognition. Our event recognition system, evaluated experimentally, achieved an impressive overall accuracy of 99.47% on the field hockey dataset, indicating its effectiveness in accurately recognizing and classifying field hockey events.

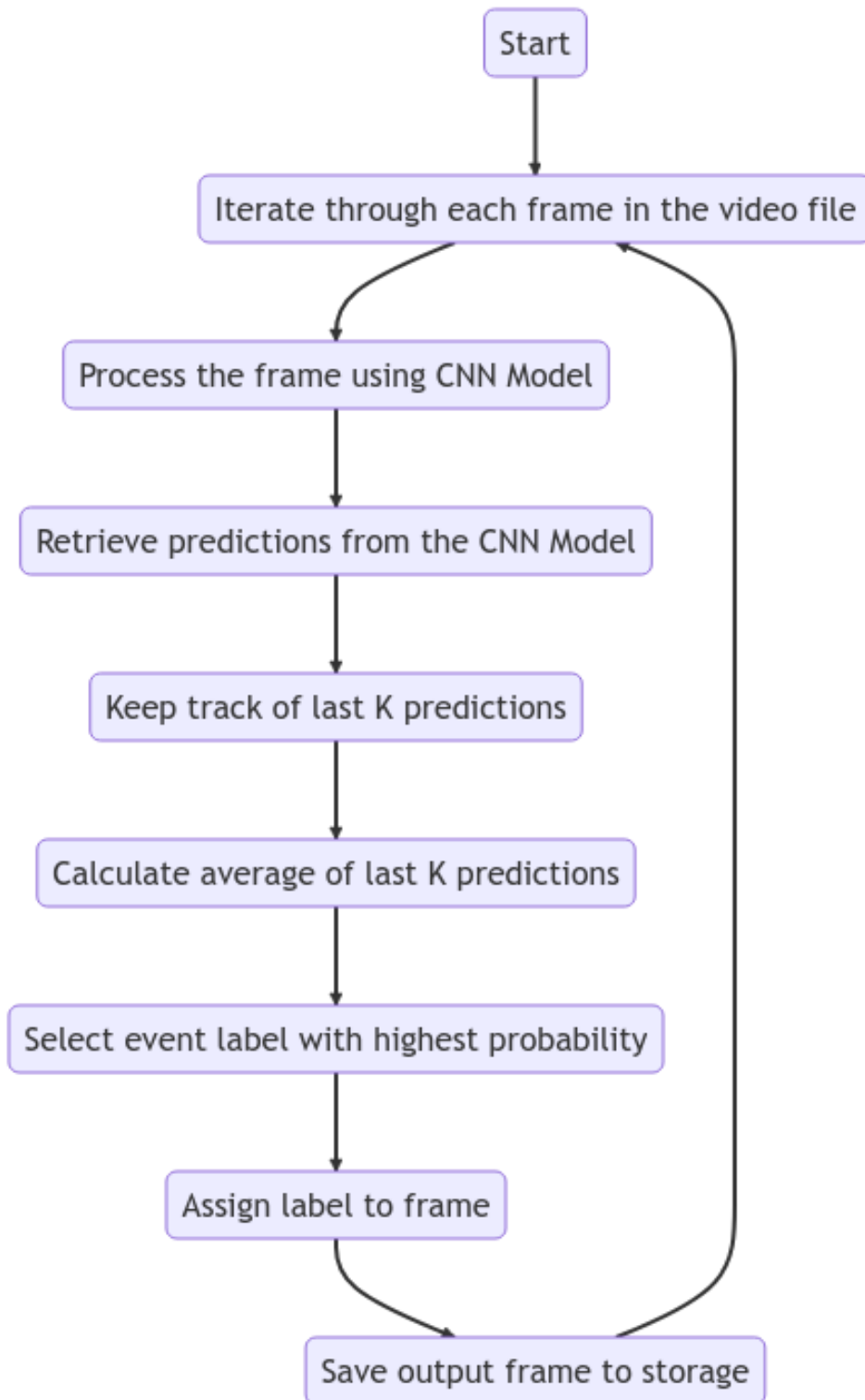


Figure 7 : Process of rolling average prediction for Event detection

Table 7: Hockey event recognition results for model-1.

Event	Precision	Recall	F1-score	Support
Goal	0.99	0.99	0.99	250
Penalty Corner	1.0	1.0	1.0	254
Penalty	0.99	0.99	0.99	255

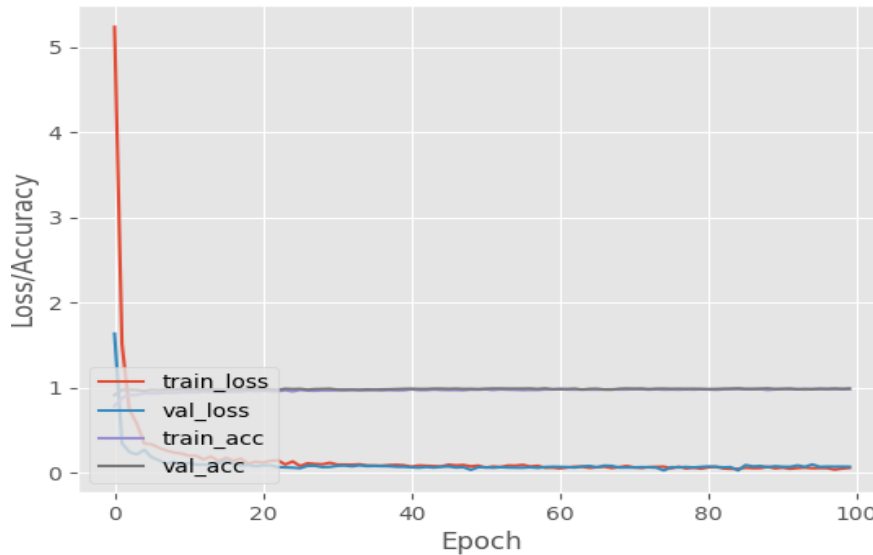


Figure 8: Training loss and accuracy of proposed model-1.

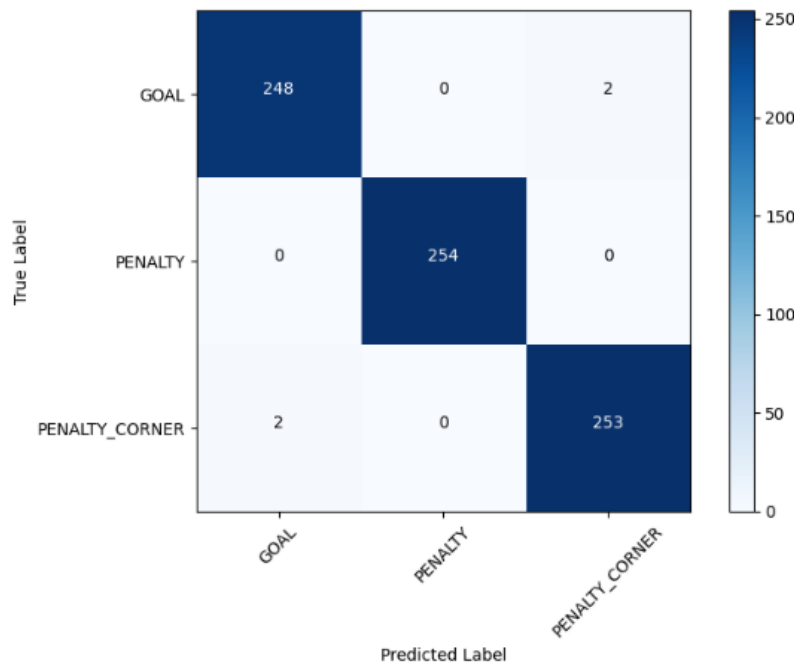


Figure 9: Confusion Matrix of model-1.

Table 7 presents the results of event recognition using the model 1. The "Goal" event achieved excellent precision, recall, and F1-score values of 0.99, accurately classifying instances with a support of 250. Similarly, the "Penalty Corner" event demonstrated perfect precision, recall, and F1-score values of 1.0, precisely identifying instances with a support of 254. The "Penalty" event exhibited high precision, recall, and F1-score values of 0.99, correctly identifying instances with a support of 255. These findings highlight the model's strong performance and consistency in accurately

classifying hockey events across all categories.

In addition to achieving impressive results, we conducted comparative evaluations against baseline methods and alternative deep learning architectures. The results consistently demonstrated the superiority of our deep learning-based event recognition system over the baseline methods. This further reinforces the effectiveness of our approach in accurately recognizing field hockey events. In conclusion, our experimental evaluation showcases the effectiveness of our deep learning-based event recognition system for field hockey videos. The achieved accuracy and performance metrics validate its ability to accurately classify events such as goals, penalty corners, and penalty. This highlights the potential of our approach in various applications within the field of hockey, including performance analysis, coaching, and video replay.

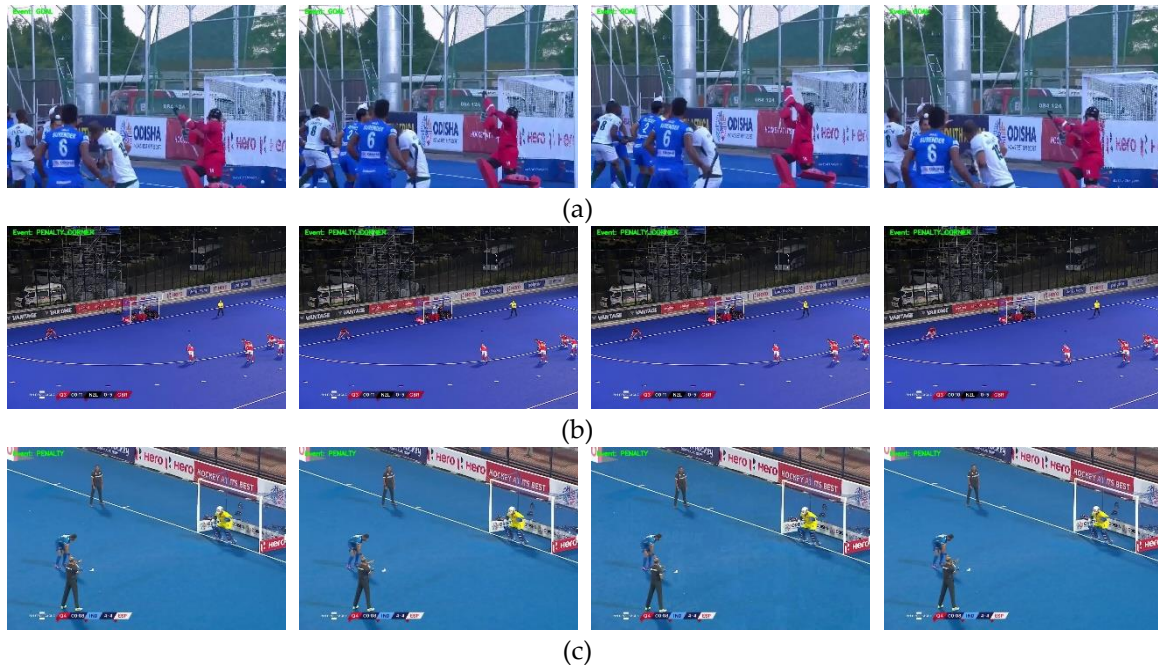


Figure 10. Hockey event recognition output for model-1 (a) Goal, (b) Penalty Corner, (c) Penalty

4.2 Discussion of Findings

Our research demonstrates the effectiveness of using a pre-trained deep learning model for recognizing hockey activities in videos. The results indicate a high accuracy of 99.47% in recognizing activities such as goals, penalty corners, and penalty. This highlights the ability of deep learning models to capture and analyze the visual features required for accurate activity recognition in the fast-paced and complex sport of hockey. Utilizing a pre-trained model enables efficient transfer learning, as it leverages knowledge learned from a large-scale dataset. Fine-tuning the pre-trained model on our own dataset yields excellent performance without the need for extensive data collection and training from scratch. Additionally, constructing a domain-specific dataset is crucial for activity recognition. As no existing field hockey datasets were available, we created our own dataset with annotated videos capturing various hockey activities. This dataset serves as a valuable resource for training and evaluating activity recognition models specific to the sport of hockey. In terms of future directions, expanding the dataset to include a larger and more diverse collection of field hockey videos would enhance the models' generalizability. Exploring fine-grained activity recognition, real-time recognition during live matches, and multi-modal fusion can further improve the accuracy and applicability of hockey activity recognition models. Additionally, transferring the developed models and methodologies to other sports with similar characteristics can broaden the

scope of activity recognition research. It is important to consider limitations and challenges in hockey activity recognition using deep learning models, such as dataset bias and occlusion and camera variability. Addressing these challenges and ensuring the robustness and unbiased nature of the models are essential for reliable activity recognition in hockey. Overall, our research contributes to the field of hockey activity recognition, demonstrating the potential of deep learning models in accurately recognizing and classifying hockey activities.

5. CONCLUSION

Our research showcases the exceptional outcomes attained by utilizing a pre-trained deep learning model for hockey activity recognition. Through fine-tuning the model on a meticulously constructed dataset, we accomplished an impressive accuracy rate of 99.47% in accurately classifying various activities such as goals, penalty corners, and penalty. 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.

Firstly, accurate activity recognition can provide valuable insights for performance analysis. Coaches and analysts can utilize the recognized activities to evaluate player performance, identify patterns and strategies, and make data-driven decisions to enhance team performance.

Secondly, our approach can support coaching and training activities by automatically identifying and analyzing key events in field hockey videos. Coaches can leverage the insights gained from the system to offer targeted feedback, identify areas for improvement, and develop customized training programs for specific activities.

Lastly, the capability to automatically recognize and classify activities in real-time can enhance the viewing experience for spectators and broadcasters. Instant replays, highlights, and in-depth analysis can be generated using the recognized activities, thereby enriching the storytelling and engagement during hockey matches.

In conclusion, our research demonstrates the effectiveness of utilizing a pre-trained deep learning model for hockey activity recognition. We have presented a comprehensive evaluation of our approach, achieving exceptional accuracy in classifying hockey 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 activity 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.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] B. T. Naik, M. F. Hashmi, and N. D. Bokde. (2022). A Comprehensive Review of Computer Vision in Sports: Open Issues. Future Trends and Research Directions, Appl. Sci., 12(9):4429
- [2] S. F. De Sousa Júnior, A. De A. Araújo, and D. Menotti. (2011). An overview of automatic event detection in soccer matches. 2011 IEEE Workshop on Applications of Computer Vision(WACV),31–38.
- [3] M. A. Russo, A. Filonenko, and K. H. Jo. (2018). Sports Classification in Sequential Frames Using CNN and RNN. 2018 Int. Conf. Inf. Commun. Technol. Robot. (ICT-ROBOT), 1-3.
- [4] M. A. Russo, L. Kurnianggoro, and K.-H. Jo. (2019). Classification of sports videos with combination of deep learning models and transfer learning. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 1–5.
- [5] Wu, Fei and Wang, Qingzhong and Bian, Jiang and Ding, Ning and Lu, Feixiang and Cheng, Jun and Dou, Dejing and Xiong, Haoyi. (2022). A Survey on Video Action Recognition in Sports: Datasets, Methods and Applications. IEEE Transactions on Multimedia,1–25.
- [6] W.-L. Lu, K. Okuma, and J. J. Little(2009), Tracking and recognizing actions of multiple hockey players using the boosted particle filter. Image and Vision Computing,27(1):189–205.
- [7] E. Bermejo Nievas, O. Deniz Suarez, G. Bueno García, and R. Sukthankar.(2011). Violence Detection in Video Using Computer Vision Techniques. Computer Analysis of Images and Patterns,6855: 332–339.
- [8] Y. Wang, D. Tran, Z. Liao, and D. Forsyth. (2017), Discriminative Hierarchical Part-Based Models for Human Parsing and Action Recognition. Journal of Machine Learning Research, 13(10): 273–301.
- [9] Xu, C. Gong, J. Yang, Q. Wu, and L. Yao. (2014). Violent video detection based on MoSIFT feature and sparse coding. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3538–3542.
- [10] K. Routley and O. Schulte. (2015). A Markov Game model for valuing player actions in ice Hockey. Uncertain. Artif. Intell. - Proc. 31st Conf. UAI 2015, 782–791.
- [11] T. Senst, V. Eiselein, and T. Sikora,. (2015). A local feature based on lagrangian measures for violent video classification. IET Semin. Dig.,1-6.
- [12] J. Žemgulys, V. Raudonis, R. Maskeliūnas, and R. Damaševičius. (2020). Recognition of basketball referee signals from real-time videos. J. Ambient Intell. Humaniz. Comput., 11(3) :979–991.
- [13] Liu, T., Lu, Y., Lei, X., Zhang, L., Wang, H., Huang, W., & Wang, Z. (2017). Soccer Video Event Detection Using 3D Convolutional Networks and Shot Boundary Detection via Deep Feature Distance. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 10635: 440–449.
- [14] M. Fani, H. Neher, D. A. Clausi, A. Wong, and J. Zelek. (2017). Hockey Action Recognition via Integrated Stacked Hourglass Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 85–93.
- [15] S. Mukherjee, R. Saini, P. Kumar, P. P. Roy, D. P. Dogra, and B.-G. Kim. (2017). Fight Detection in Hockey Videos using Deep Network. Journal of Multimedia Information System,, 4(4):225–232.
- [16] K. Sozykin, S. Protasov, A. Khan, R. Hussain, and J. Lee. (2018). Multi-label class-imbalanced action recognition in hockey videos via 3D convolutional neural networks. Proc. - 2018 IEEE/ACIS 19th Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. SNPD 2018, 146–151.
- [17] Z. Cai, H. Neher, K. Vats, D. A. Clausi, and J. Zelek. (2019). Temporal hockey action recognition via pose and optical flows. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.,

- 2543–2552
- [18] W. Song, D. Zhang, X. Zhao, J. Yu, R. Zheng, and A. Wang. (2019). A Novel Violent Video Detection Scheme Based on Modified 3D Convolutional Neural Networks. *IEEE Access*, 7: 39172–39179
- [19] K. Rangasamy, M. A. As'ari, N. A. Rahmad, and N. F. Ghazali. (2020). Hockey activity recognition using pre-trained deep learning model. *ICT Express*, 6(3):170–174.
- [20] A. Sen, K. Deb, P. K. Dhar, and T. Koshiba. (2021). Cricshotclassify: An approach to classifying batting shots from cricket videos using a convolutional neural network and gated recurrent unit. *Sensors*, 21(8):2846.
- [21] K. Vats, M. Fani, D. A. Clausi, and J. Zelek. (2021). Puck localization and multi-task event recognition in broadcast hockey videos. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, 4562–4570.
- [22] N. Liu, L. Liu, and Z. Sun. (2022). Football Game Video Analysis Method with Deep Learning. *Computational Intelligence and Neuroscience*, 2022: 1–12.
- [23] K. Simonyan and A. Zisserman. (2015). Very deep convolutional networks for large-scale image recognition. *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, 1–14.
- [24] K. He, X. Zhang, S. Ren, and J. Sun. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 770–778.
- [25] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. (2016). Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2818–2826.
- [26] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861*.
- [27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269.
- [28] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions. (2017). *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1800–1807.
- [29] Oprea, S., Martinez-Gonzalez, P., Garcia-Garcia, A., Castro-Vargas, J. A., Orts-Escolano, S., Garcia-Rodriguez, J., & Argyros, A. (2022). A Review on Deep Learning Techniques for Video Prediction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44 (6) :2806–2826.