2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0 | 979-8-3503-7378-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/OTCON60325.2024.10688070

Advancing Football Game Analysis: Integrating Computer Vision, Deep Learning, and Hybrid Techniques for Enhanced Video Analytics

1st A H Benakesh
PG Student
Department of Computer Applications
Hindustan Institute of Technology and Science Chennai-India
benakeshjan23@gmail.com

2nd Dr. R . Rajeev
Associate Professor
Department of Computer Applications
Hindustan Institute of Technology and Science Chennai India
rajeevr@hindustanuniv.ac.in

Abstract
This research is about football analytics with the help of deep learning and computer vision, through which we can track the movement of the ball and all ball possessions a particular team has. This will assist teams in properly analyzing their game and identifying areas where they may be lacking. The fusion of YOLO which is one of the most popular model architectures and object detection algorithm and deep learning techniques like 3DCNN, CNN, Faster R-CNN, LSTM, BLSTM and computer vision has revolutionized the landscape of sports analytics, particularly in the domain of football (soccer). This abstract provides a glimpse into the exciting field of football analytics through the lens of advanced technologies. In recent years, the application of deep learning models to football data has enabled the extraction of invaluable insight from the game.

This abstract serves as a gateway to the exciting world of football analytics, where deep learning and computer vision come together to revolutionize the way we perceive and engage with the beautiful game. The possibilities are limitless, as ongoing research continues to refine models and algorithms, promising even deeper insights and a richer experience for football enthusiasts worldwide.

KEYWORD: Deep Learning, Computer Vision,3DCNN, CNN, RNN, YOLO.

I. INTRODUCTION

As nowadays in all sports data analysis and data science are widely used, with the help of which the particular team are able to understand what are they lacking in and what they should work on so that they can prepare accordingly and try to make minimal amount of errors in their upcoming matches. This research is about the football analytic with the help of **deep learning** and **computer vision.** With the help of which we can track the ball movement and all the ball possession the particular team is having which will help the teams to analyse their game properly and can find what are they lacking in. The fusion of deep learning techniques and computer vision has revolutionized the landscape of sports analytics, particularly in the domain of football (soccer). This abstract provides a glimpse into the exciting field of football analytics through the lens of advanced technologies.In recent years, the application of deep learning models to football data has enabled the extraction of invaluable insights from the game. Deep neural networks have been trained on vast datasets of football matches, allowing for the automatic recognition of player movements, ball trajectories, and critical events such as goals, assists, and fouls. instrumental in providing coaches, analysts, and fans with a comprehensive understanding of the game. Computer vision plays a pivotal role in this synergy, as it enables the automatic tracking of players and the ball in real-time using video feeds from multiple camera angles. Cutting-edge tracking algorithms, often combined with object detection and pose estimation, provide high-resolution data on player positions and actions. This data is further enriched with contextual information such as the pitch dimensions, weather conditions, and team formations. The benefits of football analytics with deep learning and computer vision are multifaceted. Coaches can make data-driven decisions about player positioning, strategies, and substitutions, leading to improved team performance.

Analysts can create visually engaging dashboards and reports that highlight key moments and trends within a match, enhancing the fan experience. Moreover, stakeholders in the football industry can monetize this technology through innovative applications like virtual reality simulations, interactive broadcasts, and augmented reality fan experiences.

This project serves as a gateway to the exciting world of football analytics, where deep learning and computer vision come together to revolutionize the way we perceive and engage with the beautiful game. The possibilities are limitless, as ongoing research continues to refine models and algorithms, promising even deeper insights and a richer experience for football enthusiasts worldwide.

The project's scope specifies that the recordings should originate solely from a singular camera and a specific angle. Nevertheless, in the interest of time efficiency, additional replays and alternative angles have been retained within the input recordings. Attempts have been made to filter out these frames during program execution, such as disregarding frames lacking sufficient green in the image center (indicative of no grass). However, this filtering process is not entirely foolproof. It's worth noting that the proportion of frames originating from replays and alternate angles is minimal compared to those from the designated "correct" angle. Consequently, the impact on results is expected to be negligible.

II. RELATED WORKS

Within this segment, the project's relevant literature has been outlined.

Authors (L. Huang, Y. Huang, W. Ouyang, and L. Wang) [1] utilized Convolutional Neural Networks (CNNs) to extract features from video frames, merging them to construct video descriptors, improving temporal action localization accuracy in weakly supervised scenarios.

Authors (D. Zhang, L. He, Z. Tu, S. Zhang, F. Han, and B. Yang) [2], investigated methods for capturing motion representations crucial for real-time spatio-temporal action localization tasks. They proposed novel techniques using deep learning to effectively extract and encode motion information, enhancing the accuracy and efficiency of action localization in dynamic environments.

Authors (Z. Sun, Q. Ke, H. Rahmani, and G. Wang) [4] developed a network structure capable of spatial and temporal feature extraction in videos. In contrast to prior 3DCNNs, this method separated spatial and temporal convolutions, processing spatial data first followed by temporal convolution.

Authors(G. Yao, T. Lei, and J. Zhong) [4] provided a comprehensive analysis of the state-of-the-art in action recognition using Convolutional Neural Networks (CNNs). They discussed various CNN-based methods and their applications in pattern recognition.

Authors (Y. M. Sridevi and M. Kharde) in their research titled [5] delved into video summarization techniques employing highlight detection alongside a pairwise deep ranking model. Their approach aimed to distill key moments from videos efficiently, contributing to streamlined content analysis and retrieval.

Authors (Manzano, Calliess, de la Peña, and Limon) [6,7] explored enhancing Model Predictive Control through an exploration-exploitation approach for adaptability.

Authors(Z. Zhang, Z. Song, J. Lei, H. Lei, and Y. Peng) [8] introduced a Twin Network tracking Algorithm for Online Object Classification and Adaptive Template Update in their publication.

Authors (Sun R., Fang L., Liang Q., and Zhang X.) [9] developed an aerial photography target tracking algorithm integrating saliency and interference online learning under the twin network framework .

Authors (Zhang R., Song J., and Li S.) [10] presented a target tracking algorithm integrating an anchor-free mechanism and online update in their research paper, published in Computer Engineering and Applications, contributing to advancements in tracking.

III. THEORY

In this chapter, a concise overview will be provided on the various scientific techniques and methodologies employed in this thesis. This discussion aims to enhance the reader's comprehension of both the implementation process and the resultant outcomes. The chapter will delve into fundamental theories in computer vision, machine learning, and neural networks, elucidating the specific computer vision techniques and methodologies applied in the execution of this project.

3.1 Computer Vision: The process begins with image acquisition, where cameras or other sensors capture visual data. Computer vision algorithms then process this data to perform various tasks, such as image recognition, object detection, and scene understanding. Image recognition involves identifying and classifying objects or patterns within an image, while object detection goes a step further by locating and delineating multiple objects in a scene. As computer vision continues to evolve, its integration into various domains promises transformative advancements in technology and society.

For instance, introduced features such as CSPDarknet53 as the backbone, PANet, and Mish activation function.

3.2 YOLO V8: YOLO, or You Only Look Once, is an object detection algorithm known for its real-time processing capabilities. Each new version of YOLO usually introduces improvements in terms of accuracy, speed, and robustness. The advancements may include architectural enhancements, training techniques, or optimization strategies. Previous versions, like YOLOv4 and YOLOv5, brought improvements in accuracy,

3.3 Football Event Detection:

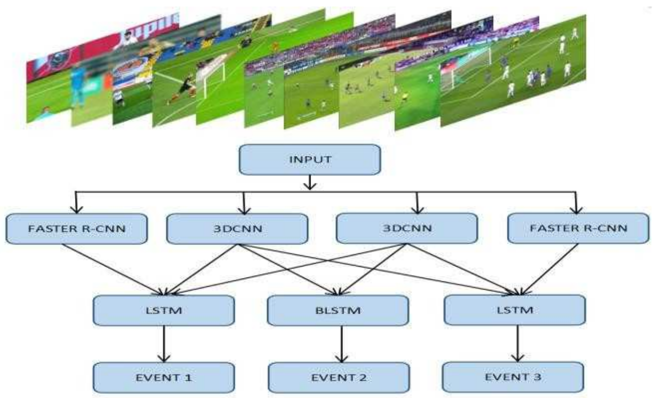

Figure 1: Football Event Detection

Event detection involves identifying the temporal boundaries of occurrences within a football video and subsequently categorizing them. The event detection framework presented in this section extends upon the classification model by incorporating a time series feature integration module. Within this model, the football video undergoes segmentation into frame sequences of predetermined lengths. Subsequently, the entire video undergoes scanning through a sliding window mechanism, enabling the prediction of the event's commencement by amalgamating and synthesizing features extracted from multiple frame sequences.

IV. METHODOLOGY

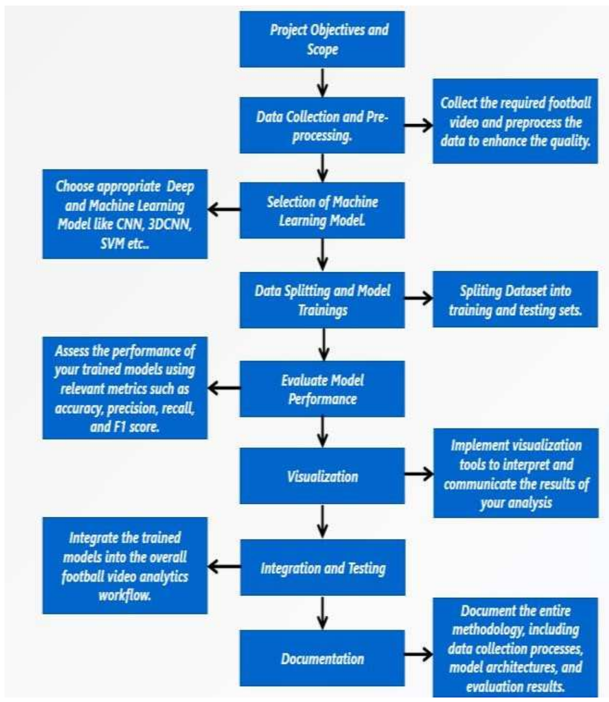

Figure 2 : Methodology of football game video analytics

4.1 Project Objective: The main objective of the project is to develop a system for player detection and ball possession analysis in football videos. This involves accurately identifyingv and tracking players throughout the video frames, as well as determining which team player currently possesses the ball. The system aims to provide valuable insights into player movements, team strategies, and ball distribution during football matches.

4.2 Data Collection: High-quality videos of football matches are collected from various sources, ensuring a diverse range of scenarios for robust analysis. The dateset used in this design is a self-made Football match video dateset, with a total of 200 soccer match videos, including FIFA World Cup 2022, Bundesliga 2024, and UEFA champions League 2024. Each game video spans approximately 20 minutes in duration, with a frame rate of 25 frames per second. The dataset encompasses shots and various events from a specific football match.

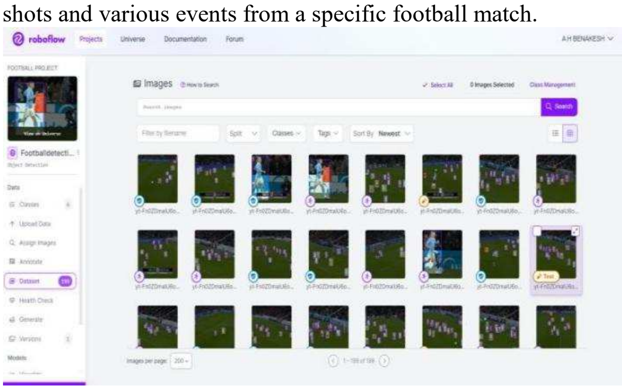

Figure 3: Image of Sample datasets collected

4.3 Preprocessing: The videos undergo preprocessing steps including frame extraction, resolution standardization, and noise reduction to enhance the quality of the footage and facilitate accurate analysis. This can be done by using some online tools like roboflow. The Annotation of the video is done frame by frame by using online tool i.e, roboflow because of which we can label the particular object in the video and can create our own datasets. In the Figure[1] the annotation is done manually and Football and Players are labeled frame by frame. And the frames are preprocessed and datasets are divided into 70% Training, 20% Validation and 10% Testing Set.

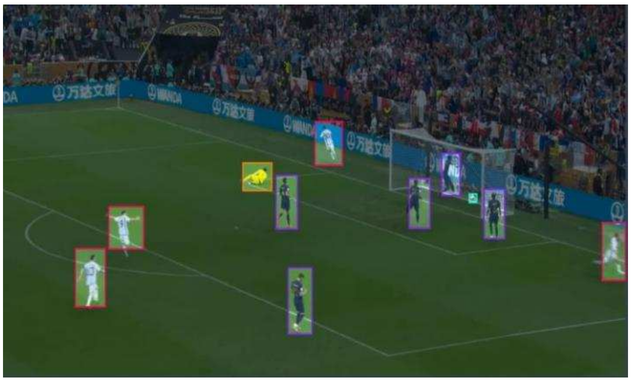

Figure 4: Annotation of Player and Football

TABLE I: Distribution of various events

Event	Training set	Test set
Shot	1365	620
Corner kick	686	245
Free kick	674	294
Yellow card	258	126
Foul	1683	687

4.4 Model Testing:

1. Confusion matrix
2. Accuracy
3. Precision
4. Recall
5. F1 score

- ***Confusion matrix:*** The confusion matrix is a fundamental tool used in the evaluation of machine learning models, including those in the context of sign language recognition ex-plored in the research paper. Here in Fig for example 83% of time player was detected correctly.

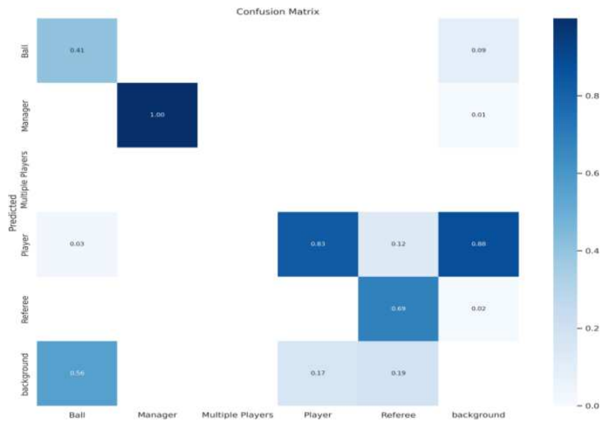

Figure 5 : Confusion matrix of predicted objects of the video.

- ***Accuracy:*** Accuracy in model training gauges the proportion of correctly predicted instances, serving as a fundamental metric for assessing the model's overall correctness.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

TABLE II : Accuracy Results

Algorithms	Accuracy
SVM	95.60 %
Random Forest	96.03%

- ***Precision:*** Precision quantifies the percentage of true positive instances relative to all instances predicted as positive by the model. It provides insight into the model's accuracy specifically when it claims positive identifications, highlighting its precision within the entire datasets.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- ***Recall***: Recall gauges the model's ability to accurately detect true positive instances among all actual positive instances in the dateset. This metric illuminates the model's effectiveness in capturing all pertinent positives, revealing any instances of true positives that the model may miss while correctly identifying positive cases.

$$\text{Recall} = \frac{TP}{TP + FN}$$

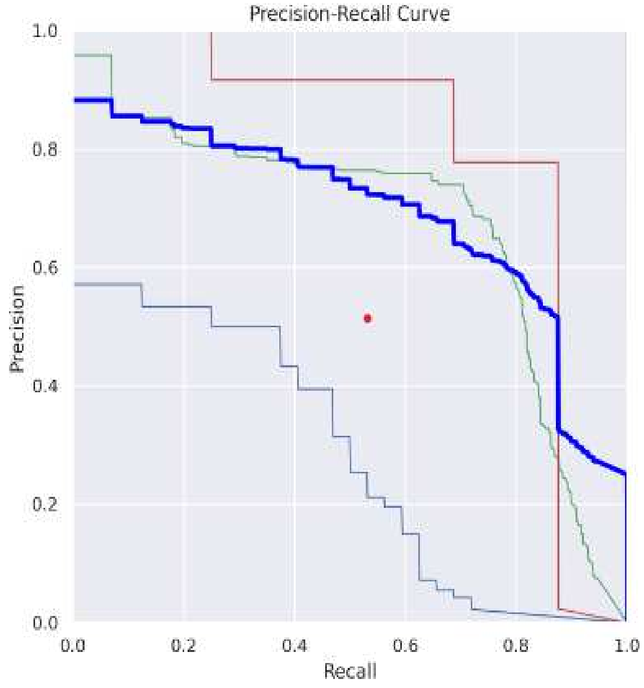

Figure 6: Precision and Recall Curve

- ***F1 Score:*** The F1 score is a harmonic mean of precision and recall, considers the balance be-tween the two metrics, emphasizing the need for both to be high for a better overall performance. The F1 score is particularly sensitive to decreases in either precision or recall, making it a comprehensive metric.

$$\text{F1 Score} = \frac{2}{\frac{1}{precision}+\frac{1}{recall}} = \frac{2*precision*recall}{precision+recall}$$

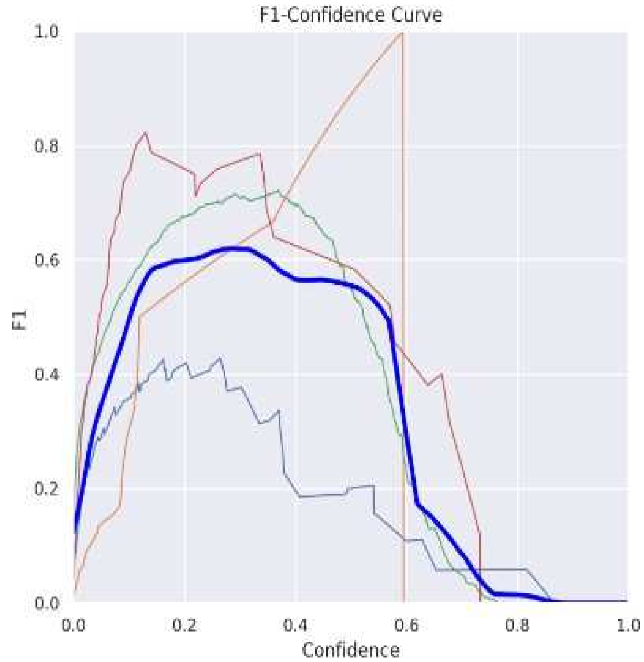

Figure 7: F1 Score Confidence Curve

4.5 Object Detection: Object detection algorithms such as YOLO (You Only Look Once) or Faster R-CNN are employed to identify the ball in each frame of the match footage. These algorithms provide precise localization of the ball, which is crucial for accurate tracking.

Figure 8: Sample Object Detection of Player and Ball With Confidence value

4.6 Ball Tracking: Once the ball is detected in each frame, tracking algorithms are utilized to follow its movement overtime. Techniques such as centroid tracking or Kalman filters estimate the ball's position in successive frames, allowing for continuous monitoring of its trajectory.

4.7 Possession Determination: Possession rules are defined based on criteria such as proximity to players and ball control. By analyzing the movement of the ball relative to players from both teams, possession is determined at any given time during the match.

4.8 Statistical Analysis: Possession metrics, including time of possession and percentage of possession for each team, are calculated using the determined possession details. These metrics provide quantitative insights into team dynamics and performance.

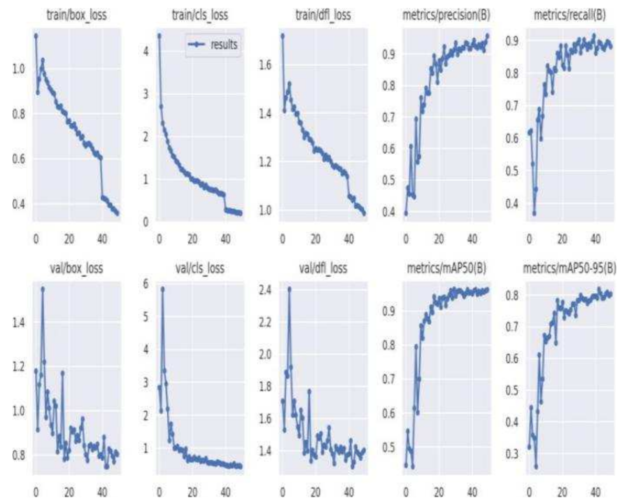

Figure 9: These are the training loss and validation loss on football events has been classified.

In Fig 9 the Training loss and validation loss of has been calculated. The behavior of the model is convincing the model is converging, Training more will give better results

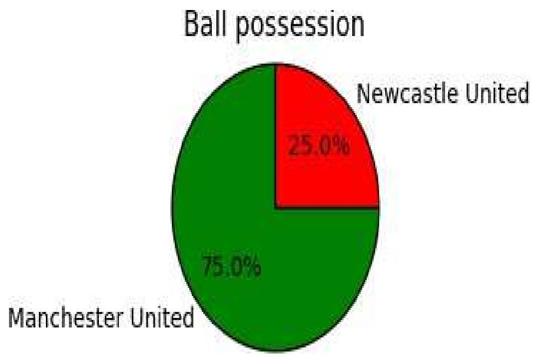

Figure 10: Comparison of Two Teams Ball Possession

In this figure 10 comparison of ball possession of two teams is show.

TABLE III: Football event classification results.

Event	Precision	Recall
Shot	79.8	86.9
Corner kick	92.1	98.6
Free kick	79.9	65.8
Yellow card	97.9	81.2
Foul	87.1	89.9
Goal	33.6	19.9
Average	79.2	75.7

In table III classification of events has taken place different events like shot, corner kick, free kick, yellow card, foul, goal average. And precision and recall of the particular events has been calculated.

4.9 Visualization: Possession patterns are visualized using various visualization tools such as bar graph, heat maps and spatial distributions etc. These visualizations help in understanding possession dynamics and identifying trends and patterns in team strategies.

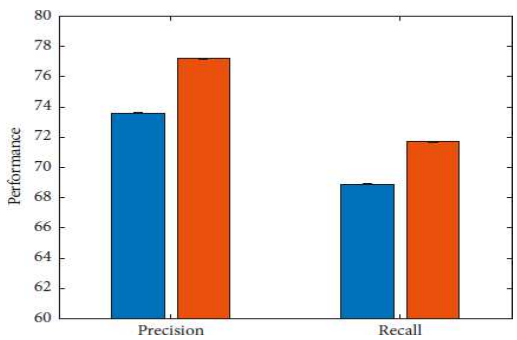

Figure 11: Comparison of Inception v2 and 3D CNN

In Figure 11 we have shown the comparison between inception v2 and 3D CNN.

4.10 Validation: The accuracy and reliability of the obtained possession details are validated through comparison with ground truth data and consistency checks. This ensures the credibility of the analysis results.

4.11 Integration: The ball possession analysis module is integrated into the broader football video analytics system, enabling seamless interaction with other analytics modules such as player tracking and event detection. This integration enhances the comprehensiveness of the analysis and provides a holistic view of match dynamics.

V. PSUEDO CODE

In this chapter we have explained how the outline of the project how the processing will take place in each frame for the football video analytics for player detection and ball possession.

This pseudo-code outlines the main steps involved in processing each frame of the football video for event detection. It starts by opening the video file, then enters a loop to read each frame sequentially. Within the loop, it detects players in the frame, processes the player detection, performs event detection based on player interactions, and displays the annotated frame.

The loop continues until the user presses the 'q' key to exit, at which point the video is released, and the program terminates.

1. Open the football video file

2. Loop:
   a. Read the next frame from the video
   b. Detect players in the frame
   c. Process player detection (e.g., filter out false positives)
   d. Perform event detection based on player positions and interactions
   e. Display the frame with player detection and event annotations
   f. Check if the user has pressed the 'q' key to exit
- If yes, release the video and exit the loop

3. End loop

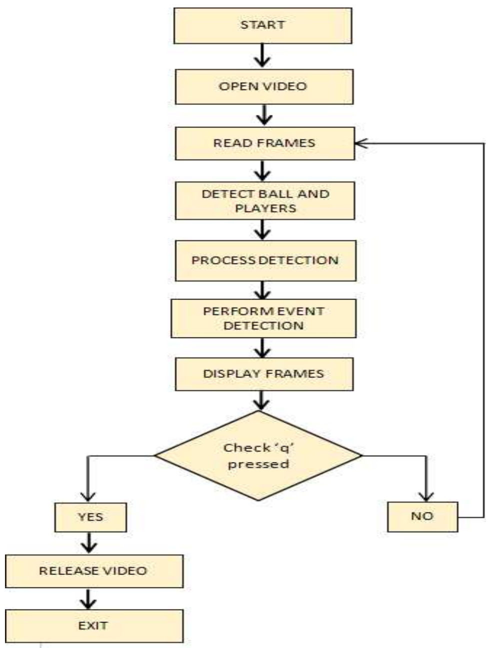

Figure 12: Flow Chart of the Pseudo Code

VI. FUTURE SCOPE

In this chapter, we outline potential directions for future research and development based on the findings and limitations of our current work. These avenues represent opportunities for extending the impact and applicability of our project in the field of football video analytics.

6.1 Integration of Multi-Modal Data: While our current project focuses primarily on analyzing football videos, future research could explore the integration of multi-modal data sources. Incorporating additional sensor data, such as player tracking data from wearable devices or match statistics from official records, could enrich the analysis and provide deeper insights into player performance, team strategies, and match dynamics.

6.2 Generalization to Other Sports: While our project focuses specifically on football video analytics, the techniques and methodologies developed could be generalized to other sports and domains. Future work could explore the adaptation and extension of our approach to analyze videos from sports such as basketball, soccer, hockey, and tennis, opening up new avenues for interdisciplinary collaboration and knowledge transfer.

6.3 Human-Centric Analysis and Interpretation: In addition to automated analysis, there is a growing interest in human-centric approaches to football video analytics. Future research could explore the development of interactive visualization tools, user-friendly interfaces, and explainable AI techniques to empower coaches, analysts, and enthusiasts to interpret and interact with the analysis results effectively.

VII. CONCLUSION

In conclusion, our project represents a significant step forward in the realm of football video analytics, leveraging cutting-edge technologies such as deep learning and computer vision to extract valuable insights from match footage. By focusing on player detection and ball possession analysis, we have provided coaches, analysts, and fans with a powerful tool for understanding the intricacies of the game and identifying areas for improvement.

Through the fusion of deep learning techniques and computer vision algorithms, we have demonstrated the ability to track player movements, determine possession dynamics, and detect key events in real-time. Our methodology, validated through rigorous testing and evaluation, has showcased promising results in terms of accuracy, reliability, and scalability.

In essence, our project serves as a testament to the transformative power of technology in sports, offering a glimpse into a future where data-driven insights drive performance, enhance strategy, and enhance the collective experience for both players and fans. As we continue to push the boundaries of innovation, the possibilities for football analytics are truly limitless.

VIII. REFERENCES

[1] L. Huang, Y. Huang, W. Ouyang, and L. Wang, “Modeling sub-actions for weakly supervised temporal action localization,” IEEE Transactions on Image Processing, vol. 30, pp. 5154–5167, 2021.

[2] D. Zhang, L. He, Z. Tu, S. Zhang, F. Han, and B. Yang, “Learning motion representation for real-time spatio-temporal action localization,” Pattern Recognition, vol. 103, Article ID 107312, 2020.

[3] Z. Sun, Q. Ke, H. Rahmani, and G. Wang, “Human action recognition from various data modalities: a review,” 2020, https://arxiv.org/abs/2012.11866.

[4] G. Yao, T. Lei, and J. Zhong, “A review of Convolutional- Neural-Network-based action recognition,” Pattern Recognition Letters, vol. 118, pp. 14–22, 2019.

[5] Y. M. Sridevi and M. Kharde, “Video summarization usinghighlight detection and pairwise deep ranking model,” Procedia Computer Science, vol. 167, pp. 1839–1848, 2020.

[6] Manzano J. M., Calliess J., de la Peña D. M., Limon D. Online learning robust MPC: an exploration-exploitation approach. IFAC-PapersOnLine . 2020;53(2):5292–5297. doi: 10.1016/j.ifacol.2020.12.1210,

[7] F. Xiao, Y. J. Lee, K. Grauman, and J. Malik, “Audiovisual slowfast networks for video recognition,” 2020, https://arxiv.org/abs/2001.08740.

[8] Z., Zhang Z., Song J., Lei H., Peng Y. Twin Network tracking Algorithm for Online Object Classification and Adaptive template Update. Journal of Communications . 2021;42(08):151–163.

[9] Sun R., Fang L., Liang Q., Zhang X. Aerial photography target tracking algorithm integrating saliency and interference online learning under the twin network framework. Journal of Electronics and Information . 2021;43(05):1414–1423.

[10] Zhang R., Song J., Li S. "Target tracking algorithm based on anchor-free mechanism and online update. Computer Engineering and Applications" . 2021;57(20):210–220.

[11] Chen C., Liu B., Wan S., et al. "An edge traffic flow detection scheme based on deep learning in an intelligent transportation system." IEEE Transactions on Intelligent Transportation Systems . 2020;22(3):1840–1852.

[12] Z. Wu, Y. G. Jiang, X. Wang, and H. Ye, “Fusing multi-stream deep networks for video classification,” 2015, https://arxiv.org/abs/1509.06086.

[13] H. H. Pham, L. Khoudour, A. Crouzil, P. Zegers, and S. A. Velastin, “Learning to recognise 3D human action from a new skeleton-based representation using deep convolutional neural networks,” IET Computer Vision, vol. 13, no. 3, pp. 319–328, 2019.

[14] Purnama F., Usagawa T. Using real-time online preprocessed mouse tracking for lower storage and transmission costs. Journal of Big Data . 2020;7(1):p. 27. doi: 10.1186/s40537-020-00304-x.

[15] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, and K. Schindler, “MOT16: A Benchmark for Multi-Object Tracking,” Comput. Vis. Pattern Recognit, Accessed: Jul. 01, 2020. [Online]. Available:http://www.motchallenge.net/.

[16] T. Akiba, S. Suzuki, and K. Fukuda, “Extremely Large Minibatch SGD: Training ResNet-50 on ImageNet in 15 Minutes,”, Accessed: Jun. 30, 2020. [Online]. Available: http://arxiv.org/abs/1711.04325.

[17] M. Sridevi and M. Kharde, “Video summarization using highlight detection and pairwise deep ranking model,” Procedia Computer Science, vol. 167, pp. 1839–1848, 2020.

[18] P. Bauer and G. Anzer, “Data-driven detection of counter-pressing in professional football,” Data Mining and Knowl-edge Discovery, vol. 35, no. 5, pp. 2009–2049, 2021.