Bein Sport
(Football Possession Analysis)

By: Ahmed Magdy , Bassant Salem , Bola Milad , Tasneem Bahaa , Jonathan Monir , Shaimaa Hesham
Supervised by: Dr.Manal Tantawi, TA.Aya Nasser
Faculty of Computer and Information Sciences - Ain Shams University

Introduction

Traditional football possession metrics rely on manual timing or pass counts, which are subjective, inconsistent and ill-suited to the game's speed and complexity. Observers using stopwatches must react instantaneously to loose ball situations and rapid possession changes, often resulting in biased or incomplete data. Similarly, tallying passes overlooks critical context failed passes, defensive pressure or strategic play in dangerous areas leading to an oversimplified view of ball control.

This project aims to eliminate human error by developing an automated, video-based system that detects and tracks the ball in real time and assigns possession to each team. Using annotated match footage, we will train a deep-learning model to recognize the ball under varying lighting and player formations. The system will then generate objective possession statistics and intuitive visualizations, enabling coaches and analysts to make data-driven tactical decisions.

Methods

Our pipeline begins with converting raw match video into a sequence of standardized frames. These preprocessed frames feed into custom-trained YOLOv8 to detect ball and player bounding boxes with high accuracy. Detected objects are linked across time by two Norfair Kalman-filter trackers one optimized for the players' slower, predictable movement and another specialized for the ball's rapid, erratic trajectories ensuring robust persistence even through heavy occlusions. In parallel, player ROI color histograms are clustered via MiniBatchKMeans to automatically assign each track to a team, providing stable labels despite lighting changes or pose variations. The bottom center point of each bounding box is then projected into real-world pitch coordinates via an inverse homography matrix derived from annotated field landmarks. Finally, frame-by-frame Euclidean distances between the ball and each player determine possession each frame is credited to the team whose nearest player is closest to the ball and the overall possession percentage is calculated as the ratio of frames held by each team to the total number of valid frames.

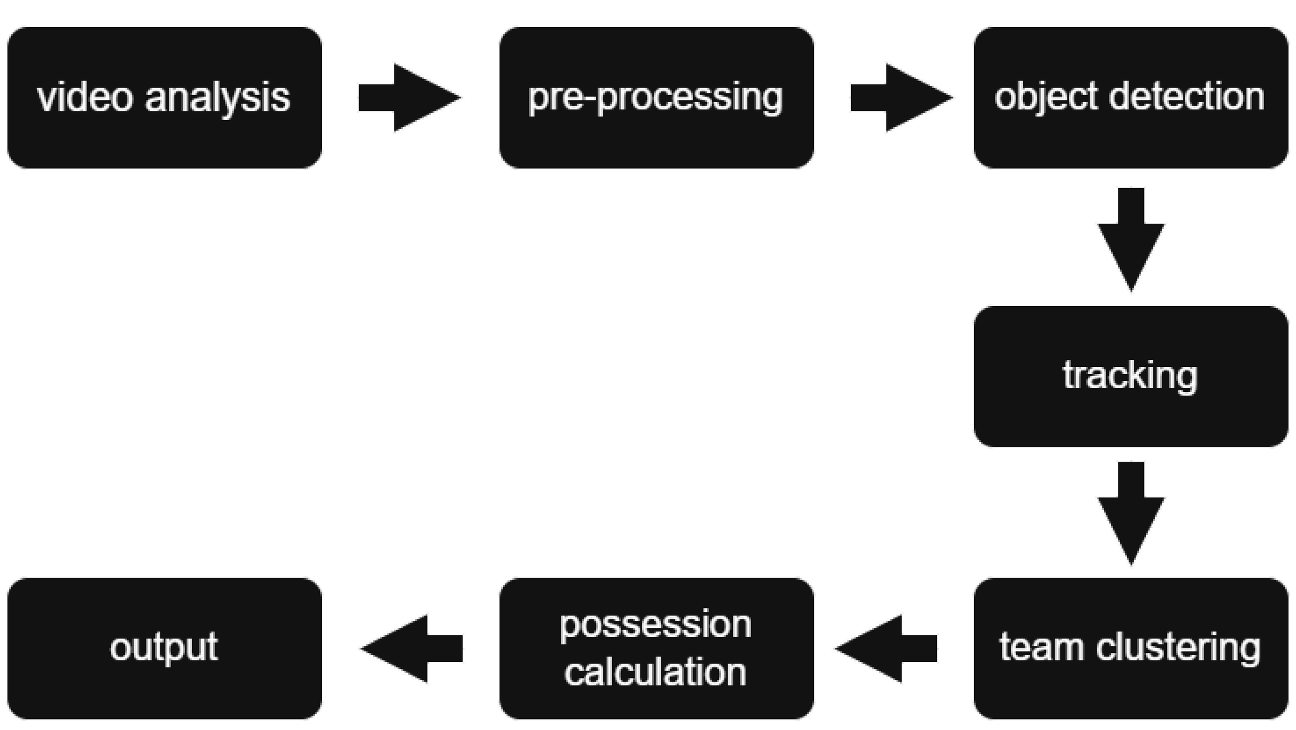


Figure 1: System Architecture

Results

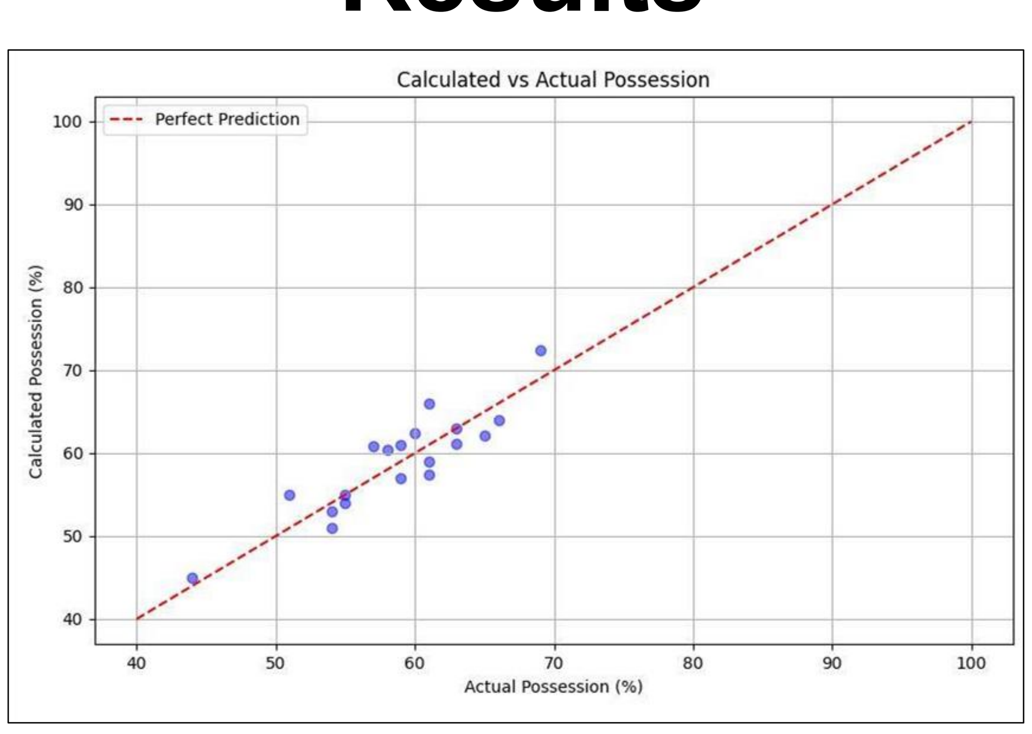


Figure 2 : Actual possession vs predicted possession

Our possession calculation was validated on 20+ full matches, producing a Mean Absolute Error (MAE) of 2.28%, a Mean Squared Error (MSE) of 6.92, and a maximum single-match deviation of 5.0%. Importantly, over 80% of our possession estimates fall within ±3% of the ground-truth values, demonstrating high reliability even during rapid game phases. For object detection, we evaluated our YOLO model on 3,298 images containing 35,684 labeled instances. It achieved a Precision of 0.91 and a Recall of 0.851, with a mean Average Precision (mAP) of 0.898 at IoU = 0.50 (and 0.632 when averaged across IoU thresholds from 0.50 to 0.95), confirming robust performance across both ball and player classes.

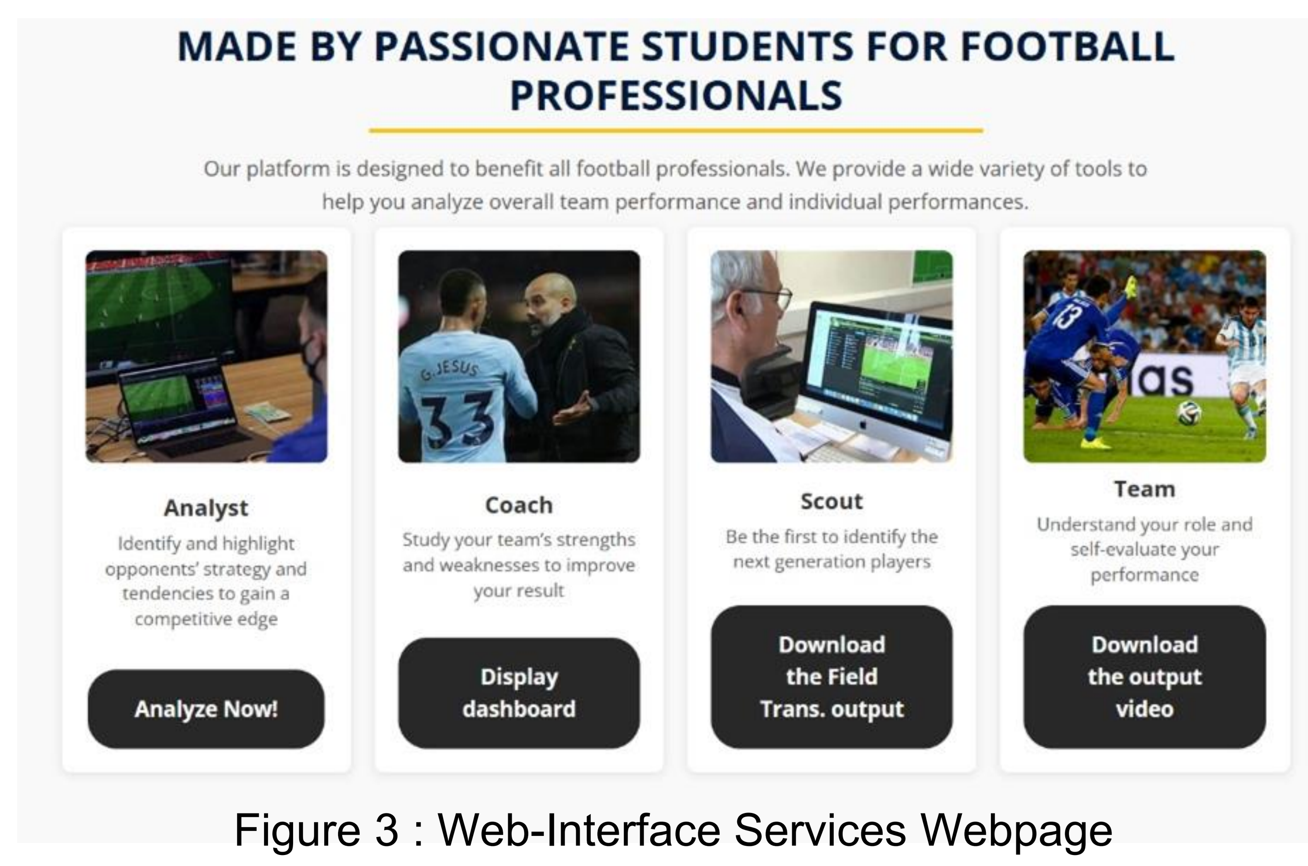
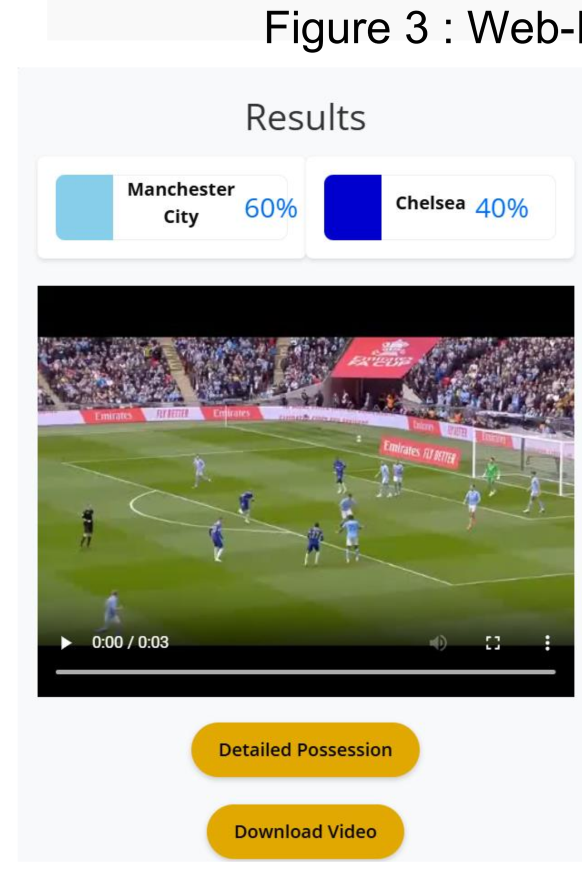


Figure 3 : Web-Interface Services Webpage



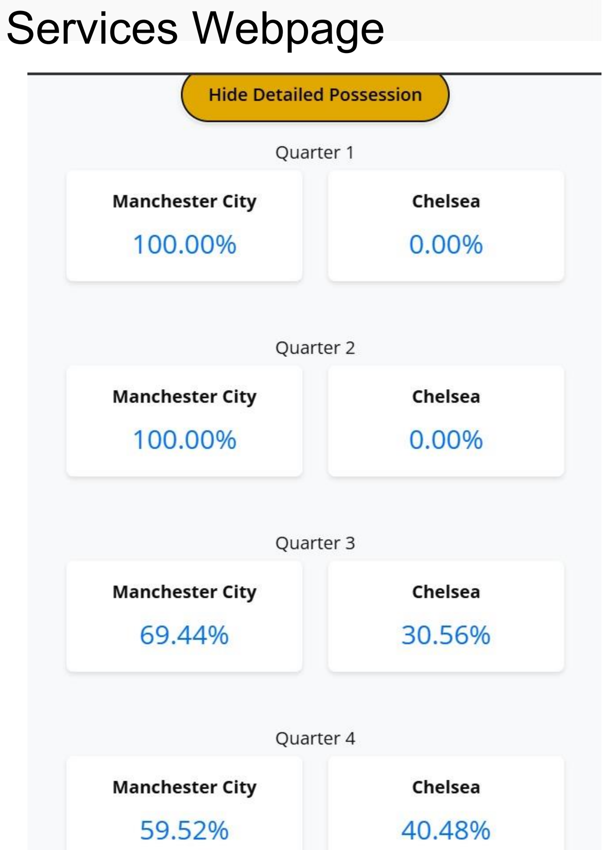


Figure 4 : Results Webpage
Figure 5: Detailed Possession Webpage

Conclusion

Football Possession Calculation project delivers a deep learning–powered system that transforms broadcast football footage into actionable possession insights: it accurately detects and tracks players and the ball, uses CNNs to assign team identities, and analyzes spatial and temporal patterns to determine clear or contested possession. Validated against manual annotations, the system achieved 91.2% precision and just a 4.1% mean absolute error. Encapsulated in an intuitive application designed for coaches and analysts, our solution streamlines performance evaluation and elevates the rigor and efficiency of match analysis.

Bibliography

[1] Chen, J., Little, J. J., Hsiao, E., & Zhu, J. (2021). SportsCap: Monocular 3D human motion capture and fine-grained understanding in challenging sports videos. International Journal of Computer Vision, 129(10), 2846-2864.

[2] Deliège, A., Cioppa, A., Giancola, S., Seikavandi, M. J., Dueholm, J. V., Nasrollahi, K., ... & Van Droogenbroeck, M. (2021). SoccerNet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp.4508-4519)

[3] Giancola, S., Amine, M., Dghaily, T., & Ghanem, B. (2018). SoccerNet: A scalable dataset for action spotting in soccer videos. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 1711-1721).

[4] Lu, J., Batra, D., Parikh, D., & Lee, S. (2019). ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. Advances in Neural Information Processing Systems, 32.