Bozok J Eng Arch, 2026; 5(1) 90-99
DOI: 10.70700/bjea.1907686

Bozok
Journal of Engineering and Architecture

e-ISSN: 3023-4298
Research Article/ Araştırma Makalesi

YOLOv8n-Based football player detection and spatial density analysis using broadcast footage

Abdulsalam BOUAISHA[1], Nihan KAZAK ÇERÇEVİK[2]*

[1] Bilecik Şeyh Edebali University, Institute of Graduate Education, Department of Electronic and Computer Engineering, Bilecik, Turkey
[2] Bilecik Şeyh Edebali University, Faculty of Engineering, Department of Computer Engineering, Bilecik, Turkey

ARTICLE INFO

Article history:

Received
11.03.2026
Accepted
08.06.2026
Published
29.06.2026

Keywords:

Football analytics
YOLOv8
Player detection
Spatial distribution

ABSTRACT

This study presents a computer vision–based framework for football player detection and spatial density analysis using standard broadcast match footage. The proposed pipeline integrates the YOLOv8n object detection model with frame-to-frame player association and a normalized 32×32 grid-based spatial representation to analyze positional distributions under dynamic match conditions. Experimental evaluation was conducted using manually annotated broadcast-video sequences obtained from publicly available professional football recordings on YouTube. Player detections were associated across consecutive frames using a Hungarian algorithm–based matching strategy to preserve temporal consistency while maintaining computational efficiency. Spatial occupancy patterns were modeled through normalized density maps generated from player centroid projections onto a fixed grid structure. The evaluated framework achieved an average detection accuracy of 92.3%, while overall F1 scores above 0.98 were observed under the evaluated experimental conditions. The results further indicated relatively consistent frame-level detection behavior. The generated spatial density maps showed higher player occupancy in central midfield regions within the analyzed sequences, demonstrating the ability of the framework to produce interpretable positional summaries directly from broadcast footage. Although the experiments were conducted on a limited number of broadcast-video segments, the findings suggest that lightweight deep learning–based models can support computationally feasible football analytics without requiring specialized sensor systems or multi-camera infrastructure. The modular structure of the framework also provides a basis for future extensions, including ball tracking and event-level tactical analysis.

YOLOv8n ile maç yayını görüntülerinden futbolcu tespiti ve uzamsal yoğunluk analizi

MAKALE BİLGİSİ

Makale Tarihleri:

Geliş tarihi
11.03.2026
Kabul tarihi
08.06.2026
Yayım tarihi
29.06.2026

Anahtar Kelimeler:

Futbol analizi
YOLOv8
Oyuncu tespiti
Uzamsal dağılım

ÖZET

Bu çalışma, standart maç yayını görüntülerinden yararlanarak futbolcu tespiti ve uzamsal yoğunluk analizi için bilgisayarlı görüye dayalı bir çerçeve sunmaktadır. Önerilen yöntem, dinamik maç koşulları altında pozisyonel dağılımları analiz etmek için YOLOv8n nesne tespiti modelini kareler arası oyuncu ilişkilendirme ve normalize edilmiş 32×32 ızgara tabanlı uzamsal gösterimle entegre etmektedir. Deneysel değerlendirme, YouTube'da kamuya açık profesyonel futbol kayıtlarından elde edilen, elle etiketlenmiş yayın video sekansları kullanılarak gerçekleştirilmiştir. Oyuncu tespitleri, hesaplama verimliliğini korurken zamansal tutarlılığı sağlamak için Hungarian algoritmasına dayalı bir eşleştirme stratejisi kullanılarak ardışık kareler arasında ilişkilendirilmiştir. Uzamsal doluluk örüntüleri, oyuncu merkez noktalarının sabit bir ızgara yapısına yansıtılmasından oluşturulan normalize edilmiş yoğunluk haritaları aracılığıyla modellenmiştir. Değerlendirilen çerçeve, ortalama %92,3'lük bir tespit doğruluğuna ulaşırken, değerlendirilen deneysel koşullar altında 0,98'in üzerinde genel F1 puanları gözlemlenmiştir. Sonuçlar ayrıca, nispeten tutarlı kare düzeyinde tespit davranışını göstermiştir. Oluşturulan uzamsal yoğunluk haritaları, analiz edilen sekanslarda orta saha bölgelerinde daha yüksek oyuncu yoğunluğunu göstererek, çerçevenin yayın görüntülerinden doğrudan yorumlanabilir pozisyon özetleri üretme yeteneğini ortaya koymuştur. Deneyler sınırlı sayıda yayın videosu segmenti üzerinde gerçekleştirilmiş olsa da, bulgular, hafif derin öğrenme modellerinin, özel sensör sistemleri veya çok kameralı altyapı gerektirmeden hesaplama açısından uygulanabilir futbol analizini destekleyebileceğini göstermektedir. Çerçevenin modüler yapısı ayrıca, top takibi ve olay düzeyinde taktiksel analiz de dahil olmak üzere gelecekteki çalışmalar için bir temel sağlamaktadır.

ORCID ID: Abdulsalam Bouaisha: 0009-0003-7545-3901; Nihan Kazak Çerçevik: 0000-0001-5339-0055

***Corresponding author/Sorumlu yazar:** Department of Computer Engineering, Bilecik Şeyh Edebali University, Bilecik/Turkey
Tel: +90 228 214 2131
E-mail:nihan.kazak@bilecik.edu.tr

To cite this article/Bu makaleye atıfta blunmak için: Bouaisha A., Kazak Çerçevik N. "YOLOv8n-Based football player detection and spatial density analysis using broadcast footage", Bozok Journal of Engineering and Architecture, vol. 5, no. 1, pp. 90-99, 2026.

1. INTRODUCTION

In recent years, advances in deep learning–based object detection have enabled direct analysis of football players from broadcast footage, eliminating the need for sensor systems or specialized tracking infrastructure [1, 2]. These developments have expanded the role of video-based analysis in football and enabled more accessible approaches for evaluating player positioning, team organization, and tactical behavior [3-5]. Despite these advances, broadcast–based football analytics still faces several challenges. Rapid player interactions, dense formations, frequent occlusion, and sudden camera movements often reduce detection stability and temporal consistency [6-8]. Players may overlap visually or disappear from view, making frame-to-frame association difficult [7, 9]. Moreover, while recent studies have achieved high detection accuracy, many approaches emphasize raw positional outputs and provide limited interpretation of collective spatial behavior or positional organization [10, 11].

Prior work has focused on player detection, tracking, tactical analysis, and action recognition using computer vision techniques [2, 12, 13]. YOLO-based architectures have become particularly popular because they balance computational efficiency and detection accuracy in sports video analysis [3-5]. For example, Wang [3] combined YOLOv5 with DeepSORT for player tracking, while Sharma et al. [4] applied YOLO-based detection for real-time match analysis. Liu et al. [5] proposed a YOLOv8-based framework for possession analysis and team classification. These studies demonstrate the potential of deep learning systems to achieve reliable player localization under broadcast conditions. However, many advanced systems rely on computationally intensive tracking algorithms, multi-camera setups, or additional sensor technologies, which limit their practical applicability in broadcast video environments.

Besides player detection, temporal consistency and spatial interpretation have also become important research topics in football analytics. Tracking approaches based on DeepSORT, association algorithms, and appearance-based models have been used to improve frame-to-frame continuity under dynamic gameplay conditions [7, 9]. Similarly, spatial analysis methods such as heat maps, occupancy analysis, and positional clustering have been employed to evaluate collective team behavior and tactical organization [10-12]. However, many advanced football analytics systems still rely on computationally intensive tracking architectures, multi-camera setups, or additional sensor technologies that may reduce their practical applicability in real-time broadcast environments.

Although previous studies have contributed significantly to football video analysis, comparatively fewer studies have focused on lightweight and computationally efficient frameworks capable of simultaneously performing player detection, temporal association, and interpretable spatial density analysis directly from standard broadcast footage. In addition, many existing systems emphasize raw positional outputs rather than structured spatial representations that can support practical interpretation of player distribution and collective positioning.

The contribution of the present study is not the development of a new detection architecture but the integration of existing methods into a practical broadcast-video pipeline. Specifically, this study combines YOLOv8n-based player detection, Hungarian algorithm–based frame association, and normalized grid-based spatial density analysis within a unified framework. A 32×32 spatial grid structure is employed to generate interpretable density-based maps of player occupancy while maintaining computational efficiency.

The objective of this study is to evaluate whether a YOLOv8n-based broadcast-video pipeline can provide stable player detection, maintain temporal consistency, and generate interpretable spatial density outputs under dynamic match conditions. The framework is evaluated using manually annotated professional match segments to examine detection behavior and spatial analysis capability.

2. LITERATURE REVIEW

Player detection is one of the most fundamental components of computer vision–based football analytics. Early approaches relied on traditional image processing techniques such as motion detection, background subtraction, and color segmentation. However, these methods often produced unstable results under broadcast match conditions because of camera movement, illumination changes, and player overlap [1].

With the development of deep learning, object detection performance in football videos has improved significantly. Convolutional neural network–based models such as YOLO have been widely adopted because they balance computational efficiency with detection accuracy [2-5]. Lightweight YOLO variants are particularly attractive for computationally efficient football analytics. For instance, Wang [3] combined YOLOv5 with DeepSORT to achieve stable tracking, Sharma et al. [4] proposed a YOLO-based framework for football analytics, and Liu et al. [5] integrated YOLOv8 with clustering for team classification. These studies confirm the feasibility of deep learning–based detection under broadcast conditions, yet they often remain limited in integrating interpretable spatial analysis. Nevertheless, football videos still present important challenges for object detection systems. Rapid player movement, dense formations, partial occlusion, and abrupt camera transitions may reduce localization stability, especially during crowded midfield situations [6, 8]. Therefore, maintaining stable and interpretable player detection outputs under real broadcast conditions remains an active research problem.

In addition to frame-level player detection, maintaining temporal consistency across consecutive frames is critical for reliable football analytics. Tracking methods aim to preserve player continuity over time and reduce instability caused by temporary detection loss or visual ambiguity. Recent tracking systems commonly integrate deep learning–based object detectors with algorithms such as SORT, DeepSORT, and appearance-based association methods [3, 7, 9]. These approaches improve tracking continuity by combining motion estimation and visual feature matching. For example, Zhang et al. [7] demonstrated that combining YOLO detection with DeepSORT tracking improves temporal stability in dynamic video sequences. Similarly, Ranasinghe and Thayasivam [9] proposed an appearance-based football tracking framework designed to improve occlusion recovery in crowded match environments. The Hungarian algorithm is another widely used method for frame-to-frame object association because of its computational efficiency and deterministic structure [14]. However, centroid-based association methods may still experience identity-switching limitations during prolonged occlusion or dense player interactions. By minimizing spatial matching cost between detections in consecutive frames, the algorithm enables stable player association without requiring highly complex appearance modeling. Although lightweight association approaches may experience difficulties during prolonged occlusion events, they remain suitable for real-time football analytics where computational simplicity is important.

Despite recent improvements, maintaining stable player tracking in football footage continues to be challenging because of rapid gameplay transitions, camera motion, and dense player interactions [8, 9]. Therefore, balancing temporal robustness and computational efficiency remains an important issue in broadcast-video–based football analytics.

Besides player detection and tracking, spatial analysis has become an important component of football analytics because positional information can provide insights into collective team behavior and tactical organization [10]. Spatial analysis methods are commonly used to evaluate player compactness, positional occupation, movement tendencies, and team structure during different phases of the game. Several studies have investigated spatial occupancy analysis, heat-map generation, and positional clustering using tracking data extracted from football matches [10-12]. Grid-based spatial representations are particularly useful because they summarize positional behavior over time instead of focusing only on isolated player coordinates. Memmert et al. [10] emphasized the importance of positional data analysis for tactical evaluation in football, while Gu et al. [11] proposed a machine learning framework for analyzing space-control efficiency using positional information extracted during gameplay. Similarly, Wang et al. [12] demonstrated how artificial intelligence systems can support tactical interpretation and decision-making in football environments.

Although recent studies have introduced more advanced tactical analysis methods, many systems still depend on multi-camera infrastructures, wearable sensors, or computationally expensive tracking frameworks [10, 11]. In addition, spatial analysis is often treated as a secondary component rather than being directly integrated into lightweight broadcast-video analysis framework. Consequently, relatively fewer studies have focused on computationally efficient systems capable of simultaneously performing player detection, temporal association, and interpretable spatial density analysis directly from standard broadcast footage.

The reviewed studies show that substantial progress has been achieved in football player detection, tracking, and spatial analysis using computer vision techniques. Existing approaches demonstrate that deep learning–based models can provide reliable player localization, while tracking and positional analysis methods can support tactical interpretation [3, 7, 10]. However, several limitations remain evident in the current literature. First, many studies mainly prioritize detection accuracy without integrating interpretable spatial analysis into the same framework. Second, advanced tracking systems frequently rely on computationally intensive appearance-based models or specialized infrastructures that may reduce practical applicability in real-time broadcast environments [9, 10]. Third, relatively limited attention has been given to lightweight broadcast footage analysis framework capable of simultaneously maintaining temporal consistency and generating structured spatial density representations directly from standard broadcast footage.

The novelty of the present study lies in its applied integration: YOLOv8n-based detection, Hungarian algorithm–based frame association, and normalized 32×32 grid-based spatial density analysis are combined into a unified broadcast-video pipeline. Rather than proposing a new detection architecture, the study focuses on developing a computationally efficient and interpretable framework capable of generating meaningful positional summaries under dynamic match conditions.

3. PROPOSED FRAMEWORK

The framework was designed to detect football players, associate detections across frames, and analyze spatial density patterns from broadcast footage. Unlike sensor-based or multi-camera tracking systems, the framework operates directly on publicly available single-camera broadcast footage, thereby improving practical applicability and reducing infrastructure requirements. The overall pipeline follows a modular structure in which detection, frame association, and spatial analysis components operate sequentially while remaining computationally efficient under dynamic match conditions. This modular design also allows future integration of additional analytical modules without requiring structural modification of the pipeline. The main components of the present approach and their corresponding functions within the analysis pipeline are summarized in Table 1.

Table 1. Overview of algorithms and their functions

Algorithm / Module	Type	Function
YOLOv8 (Ultralytics)	Deep Learning (Object Detection)	Detects players and outputs bounding boxes with confidence scores
Player Detector	Wrapper Module	Loads and manages YOLOv8 model execution
linear_sum_assignment	Optimization (Hungarian Algorithm)	Associates detections across consecutive frames
Heat Map Generator	Spatial Analysis	Projects player positions onto fixed grid for density mapping

Experimental evaluation was conducted using publicly available professional football broadcast videos obtained from YouTube. The analyzed footage included high-definition broadcast recordings containing rapid player interactions, partial occlusion, dense midfield formations, and camera motion commonly observed during football matches. Video processing was performed using OpenCV-based frame acquisition, and supported formats included MP4, AVI, and MOV with resolutions ranging from 720p to 1080p. The primary evaluation sequence consisted of a 70-second segment corresponding to 2,100 frames at 30 fps. Additional longer sequences were also evaluated to examine performance stability under extended gameplay conditions.

Player detection was performed using the YOLOv8n object detection architecture implemented through the Ultralytics framework. YOLOv8n was selected because it provides a favorable balance between computational cost and detection accuracy compared with larger YOLOv8 variants. Each frame was processed at its native resolution, and detected players were represented using bounding boxes accompanied by confidence scores ranging between 0 and 1. A minimum confidence threshold of 0.3 was applied to suppress low-probability predictions and reduce false-positive detections in visually crowded scenes. Detected player positions were represented by the centroids of the predicted bounding boxes.

To preserve temporal continuity across consecutive frames, player detections were associated using the Hungarian assignment algorithm implemented via the linear_sum_assignment function. Matching was performed according to centroid-based spatial distance between detections in adjacent frames (See Eq. 1).

$$d_{ij} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - j_j\right)^2} \tag{1}$$

where d_{ij} represents the Euclidean distance between the centroids of player(i) and player(j) in consecutive frames. The Hungarian optimization procedure was then applied to minimize the total spatial matching cost and preserve short-term positional continuity throughout the analyzed video sequences.

Compared with more computationally intensive frame-to-frame association approaches such as DeepSORT or appearance-based re-identification methods, the adopted association strategy prioritizes computational simplicity while maintaining stable frame-to-frame correspondence under moderate occlusion and camera motion conditions. Although prolonged occlusion events may still introduce temporary association instability, the approach was considered suitable for lightweight broadcast-video analysis.

Temporal consistency was quantified using the inverse normalized variation of frame-level F1 scores across consecutive video segments (See Eq. 2).

$$C = 1 - \sigma_{F1} \tag{2}$$

where C denotes the consistency score and σ_{F1} represents the standard deviation of frame-level F1 scores. Higher consistency values indicate more stable detection behavior across consecutive gameplay frames.

Spatial analysis was performed using a normalized 32×32 grid representation superimposed onto the football field. Player centroid coordinates extracted from each frame were projected onto the corresponding grid cells, where occupancy frequencies were accumulated over time to generate density-based spatial representations. The 32×32 grid structure was selected to balance spatial granularity and computational overhead.

To improve comparability across different gameplay intervals, occupancy values were normalized with respect to both the number of processed frames and the number of detected players. This normalization procedure ensured that the resulting density maps reflected relative spatial occupation intensity rather than absolute positional counts. In addition to density visualization, the playing field was divided into three coarse tactical regions representing defensive, midfield, and attacking zones. Regional occupancy percentages were then calculated based on the accumulated centroid distributions.

All positional and statistical outputs were organized using the Pandas library to facilitate structured analysis of frame-level and sequence-level data. The complete pipeline operated sequentially on the incoming video stream while preserving frame-level temporal information throughout the analysis process. Final outputs consisted of normalized spatial density maps, temporal detection statistics, confidence distributions, and zone-based occupancy summaries designed to provide interpretable representations of collective player positioning from broadcast footage. The experimental evaluation was conducted using the Google Colab environment with GPU acceleration enabled. Reported processing-speed measurements therefore reflect the evaluated cloud-based execution environment and should not be interpreted as optimized deployment performance.

4. RESULTS AND DISCUSSION

The proposed framework was evaluated using manually annotated broadcast-video sequences obtained from publicly available professional football recordings. Ground-truth annotations were manually generated by the authors through frame-level player bounding-box labeling. Detection performance was evaluated using the Intersection over Union (IoU) criterion with a threshold of 0.5 for true-positive assignment.

The temporal distribution of detected players throughout the primary 70-second evaluation sequence is presented in Figure 1. The number of detections per frame generally ranged from 15 to 20. Temporary decreases in player detections were observed during short intervals involving dense player overlap, rapid gameplay transitions, and abrupt camera motion. However, these fluctuations remained localized and did not persist across extended frame intervals, indicating stable recovery behavior following short-term visual disturbances.

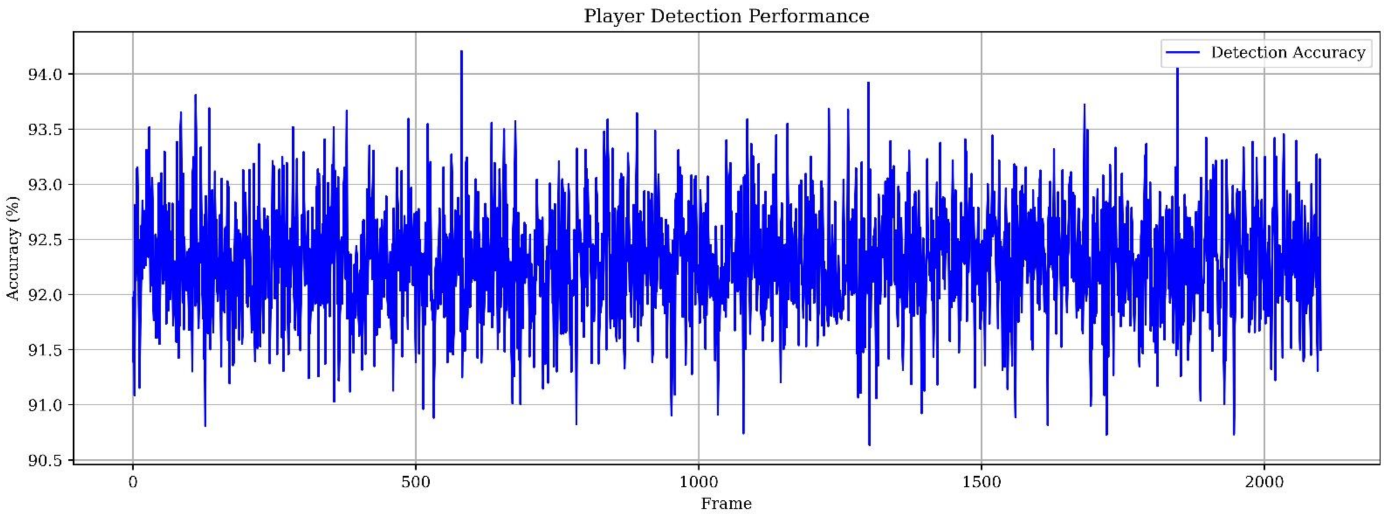


Figure 1. Temporal distribution of player detections

Figure 2 presents frame-level detection accuracy throughout the analyzed sequence. The pipeline achieved an average detection accuracy of 92.3%, while frame-level accuracy values remained within a relatively narrow interval between approximately 91% and 94%. Short-term reductions in detection accuracy were primarily associated with partial occlusion and sudden camera movement. Nevertheless, detection performance recovered rapidly during subsequent frames, supporting the stability of the detection pipeline under dynamic broadcast conditions.

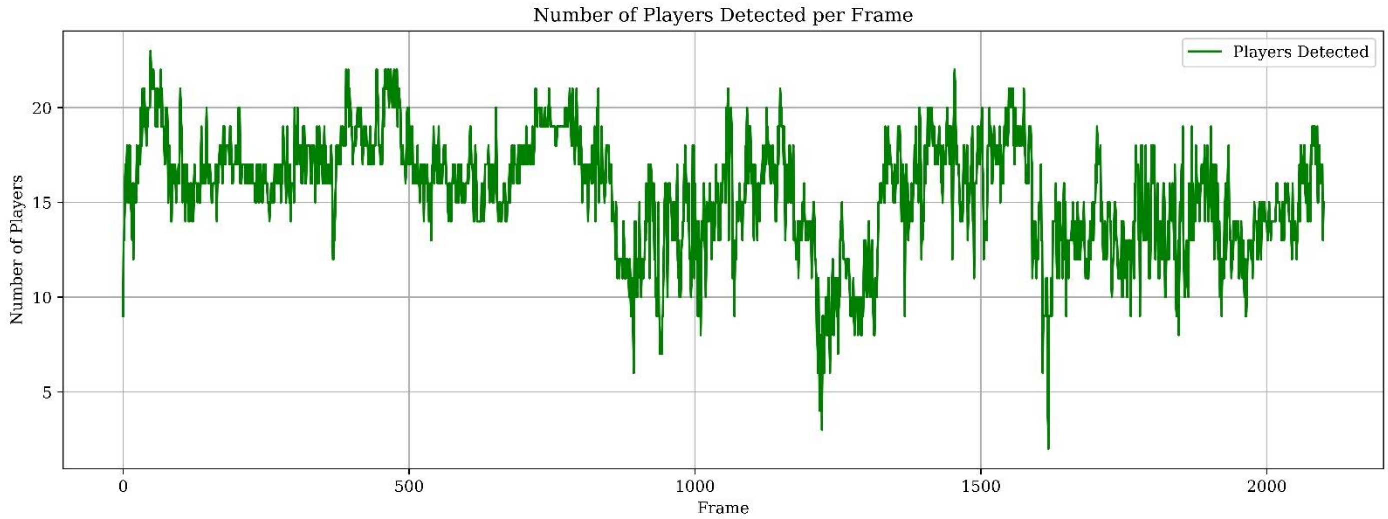
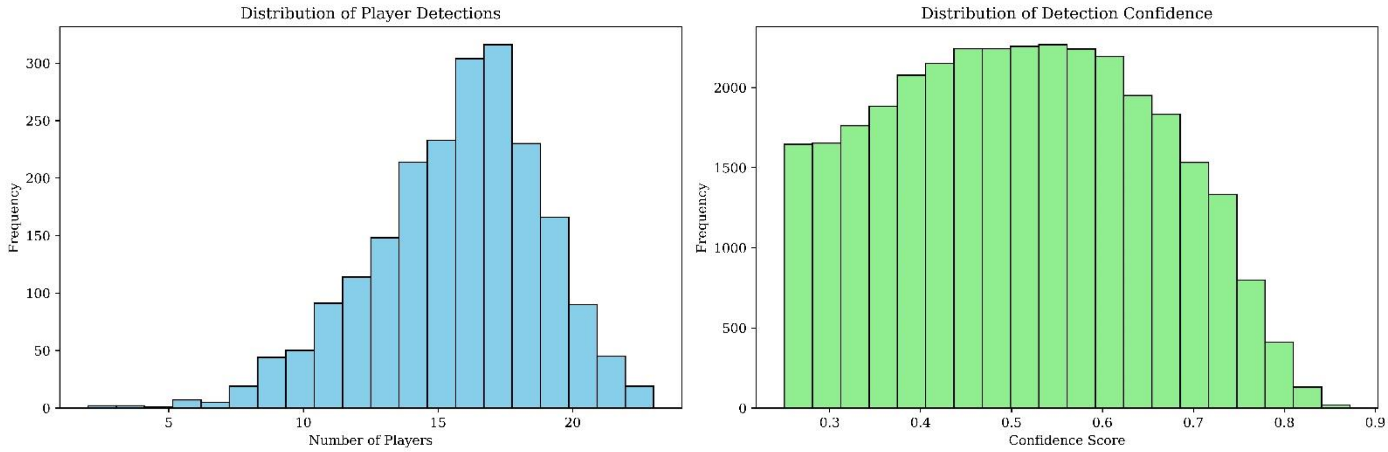


Figure 2. Detection accuracy over time

The statistical distributions of player detections and confidence scores are illustrated in Figure 3. The number of detected players per frame followed an approximately normal pattern centered around 17–18 detections, which corresponds to typical midfield-oriented broadcast views. Confidence scores were predominantly concentrated between 0.55 and 0.60, with an average confidence value of 0.520. These moderate confidence values should be interpreted cautiously because broadcast football footage contains frequent player overlap, partial occlusion, and motion blur. The adopted confidence filtering strategy intentionally suppressed ambiguous low-confidence detections in order to reduce false positives and preserve temporal consistency.



(a) (b)

Figure 3. The distribution analysis of (a) player detections and, (b) detection confidence

Spatial occupancy patterns derived from the normalized 32×32 grid representation are shown in Figure 4. The resulting density map indicates a higher concentration of player activity within central midfield regions during the analyzed sequence. Rather than representing isolated positional events, the generated spatial density map map reflects aggregated occupancy behavior accumulated throughout the evaluation interval. Because the density values were normalized according to both frame count and player count, the spatial representation reflects relative occupation intensity rather than raw positional frequency.

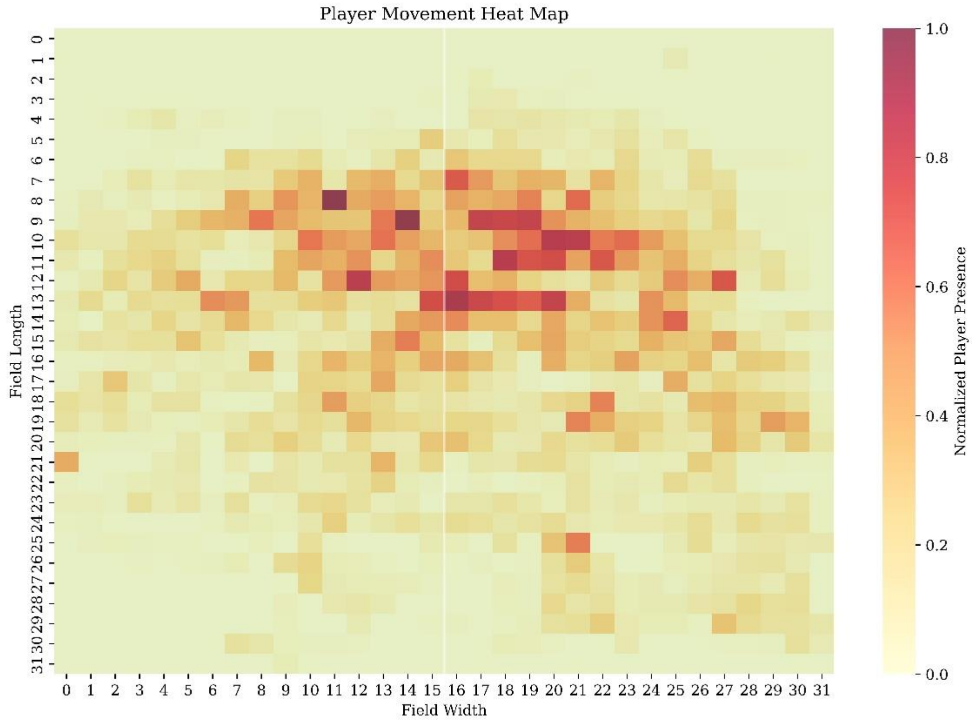


Figure 4. Normalized player spatial density map based on a 32x32 grid

Examples of individual player detections extracted from broadcast frames are presented in Figure 5. The PlayerImageExtractor module isolated player regions using bounding-box coordinates together with minimum confidence and size constraints. The examples demonstrate that this framework can preserve player localization under visually crowded match conditions, including partial overlap situations and moderate camera motion.

Figure 5. Player detection examples

The zone-based occupancy analysis presented in Figure 6 further quantified spatial distribution patterns. The midfield zone accounted for 48.9% of cumulative player occupancy, followed by the attacking zone with 33.5% and the defensive zone with 17.6%. Although these findings correspond only to the analyzed broadcast segment and should not be interpreted as complete tactical conclusions, they demonstrate that the framework can generate interpretable positional summaries beyond raw coordinate extraction.

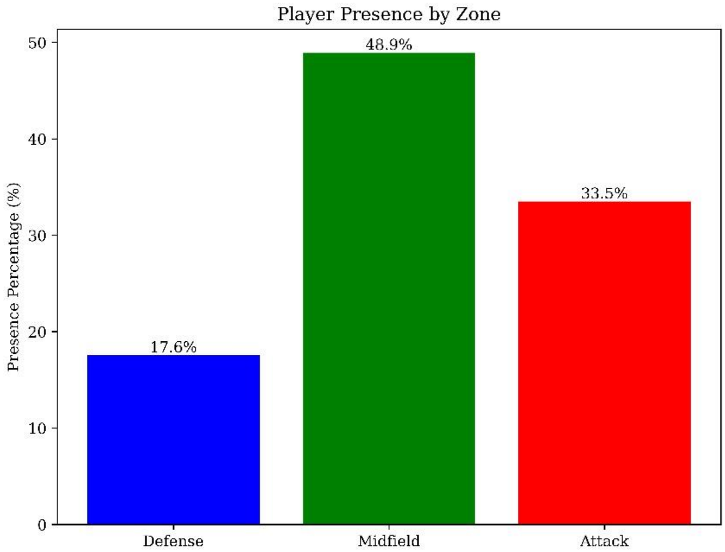


Figure 6. The zone-based analysis

To further examine performance stability, the present approach was evaluated using three temporal configurations consisting of 3105.8-second, 124.8-second, and 70-second sequences. Quantitative performance metrics for all evaluated configurations are summarized in Table 2. Across all temporal settings, the overall F1 score remained above 0.98, indicating strong balance between false-positive and false-negative detections. The highest temporal consistency score was obtained in the 70-second sequence, where the consistency value reached 0.9751 with a relatively low F1 standard deviation of 0.0246. Longer sequences exhibited slightly larger variability because of accumulated occlusion events and abrupt camera transitions; however, overall performance remained stable throughout the evaluated durations.

Table 2. Comparative performance metrics for different test configurations

Metric	3105.8 s	124.8 s	70 s
Overall F1 Score	0.9868	0.9864	0.9886
F1 Std. Deviation	0.0644	0.0512	0.0246
Consistency Score	0.9346	0.9481	0.9751
Confidence Interval (95%)	[0.984 - 0.985]	[0.985 - 0.989]	[0.987 - 0.989]
Overall Precision	0.9803	0.9794	0.9815
Overall Recall (Sensitivity)	0.9933	0.9935	0.9958
Total Frames Processed	93175	3744	2100
Total Detections	556635	24783	23482
Average Detections per Frame	5.97	6.62	11.18
Processing Speed (FPS)	1.7	1.5	1.8
Average Processing Time (ms/frame)	601.7	671.4	549.9
Detection Throughput (detection/s)	9.9	9.9	20.3

The pipeline achieved overall precision values between 0.9794 and 0.9815 and recall values between 0.9933 and 0.9958 across all evaluated sequences. These results demonstrate that the system maintained high detection coverage while limiting false-positive predictions under broadcast match conditions. Confidence interval analysis also demonstrated relatively stable frame-level detection behavior across different temporal scales.

Processing speed measurements indicated frame-processing performance between approximately 1.5 and 1.8 FPS under the evaluated hardware configuration. Although the current implementation does not achieve full real-time inference speed, the results demonstrate that lightweight object detection models can still support computationally feasible football analytics directly from broadcast footage without requiring specialized sensor systems or multi-camera infrastructure.

From a spatial-analysis perspective, the proposed grid-based density representation provides a structured mechanism for summarizing collective positioning behavior during dynamic gameplay. The observed concentration of activity in central midfield areas reflects positional compactness within the evaluated sequence rather than definitive tactical interpretation. Consequently, the generated spatial outputs should be interpreted as occupancy-based positional indicators rather than complete tactical conclusions.

Several limitations of the present study should be acknowledged. First, the experimental evaluation was conducted using a limited number of broadcast-video sequences and therefore does not fully represent the variability of different leagues, stadium environments, or broadcast styles. Second, the framework relies exclusively on single-camera broadcast footage, which may reduce detection continuity during prolonged occlusion or abrupt zoom transitions. Third, the proposed analysis focuses primarily on player localization and occupancy modeling without incorporating ball tracking, player re-identification, or event-level annotations. These components may improve tactical interpretation in future extensions of the framework.

5. CONCLUSIONS

This study presented a computer vision–based framework for football player detection, temporal association, and spatial density analysis using standard broadcast footage. The proposed pipeline integrated YOLOv8n-based player detection with Hungarian algorithm–based frame association and a normalized 32×32 grid representation in order to generate interpretable positional summaries under dynamic gameplay conditions. Unlike sensor-based or multi-camera systems, the framework operated directly on publicly available broadcast videos, thereby reducing infrastructure requirements while preserving computational simplicity.

Experimental evaluation demonstrated that the framework was capable of maintaining stable player detection performance across different temporal configurations. The evaluated sequences produced overall F1 scores above 0.98, while the generated spatial density maps successfully represented occupancy tendencies within the analyzed match segments. The findings further indicated that lightweight deep learning–based detection models can support computationally feasible football analytics directly from broadcast footage, despite challenges associated with occlusion, camera motion, and dense player interactions.

The present approach does not aim to provide complete tactical interpretation or full-scale tracking robustness under all broadcast conditions. Instead, the study demonstrates the feasibility of combining player detection, temporal consistency analysis, and grid-based spatial modeling within a unified and interpretable analysis pipeline. The generated positional summaries should therefore be interpreted as occupancy-based spatial indicators rather than definitive tactical conclusions.

Several limitations should also be acknowledged. The experiments were conducted using a limited number of broadcast-video sequences and a single-camera setup, which may affect generalizability across different leagues, stadium conditions, and broadcast styles. In addition, prolonged occlusion events and abrupt camera transitions may still reduce frame-to-frame association continuity under certain gameplay situations.

Future studies may extend the framework by incorporating ball tracking, player re-identification, multi-object temporal modeling, and event-level tactical analysis. Additional evaluation on larger and more diverse broadcast datasets may also improve the generalizability and practical applicability of the proposed approach.

AUTHOR CONTRIBUTIONS

Abdulsalam Bouaisha: Conceptualization, Data Curation, Software, Investigation, Formal Analysis, Validation. **Nihan Kazak Çerçevik:** Methodology, Supervision, Results Interpretation, Writing – Original Draft Preparation, Writing – Review & Editing. All authors contributed to the analysis and writing the manuscript.

CONFLICT OF INTEREST

There is no conflict of interest between the authors.

ETHICS

There is no ethical problem in the publication of this article.

REFERENCES

[1] B. T. Naik, M. F. Hashmi, and N. D. Bokde, "A comprehensive review of computer vision in sports: open issues, future trends and research directions," Applied Sciences, vol. 12, no. 9, pp. 4429, 2022.

[2] S. Akan and S. Varlı, "Use of deep learning in soccer videos analysis: survey," Multimedia Systems, vol. 29, no. 3, pp. 897–915, 2023.

[3] B. Wang, "Football sports video tracking and detection technology based on YOLOv5 and DeepSORT," Discover Applied Sciences, vol. 7, no. 6, pp. 563, 2025.

[4] A. Sharma et al., "Real-Time Football Match Analysis: Leveraging YOLO for Enhanced Object Detection and Possession Tracking," Procedia Computer Science, vol. 252, pp. 312–321, 2025.

[5] Y. Liu et al., “Automatic Estimation of Football Possession via Improved YOLOv8 Detection and DBSCAN-Based Team Classification,” Sensors, vol. 26, no. 4, pp. 1252, 2026.

[6] A. Cioppa et al., “Setting a Baseline for Long-Shot Real-Time Player and Ball Detection in Soccer Videos,” arXiv preprint arXiv:2311.06892, 2023.

[7] J. Zhang et al., “Video object tracking based on YOLOv7 and DeepSORT,” arXiv preprint arXiv:2207.12202, 2022.

[8] Z. L. Crang et al., “Concurrent validity of computer-vision artificial intelligence player tracking software using broadcast footage,” arXiv preprint arXiv:2508.19477, 2025.

[9] C. Ranasinghe and U. Thayasivam, “Team-aware football player tracking with SAM: an appearance-based approach to occlusion recovery,” arXiv preprint arXiv:2512.08467, 2025.

[10] D. Memmert, K. A. Lemmink, and J. Sampaio, “Current approaches to tactical performance analyses in soccer using position data,” Sports Medicine, vol. 47, no. 1, pp. 1–10, 2017.

[11] C. Gu, V. De Silva, and M. Caine, “A machine learning framework for quantifying in-game space-control efficiency in football,” Knowledge-Based Systems, vol. 283, pp. 111123, 2024.

[12] Z. Wang et al., “TacticAI: an AI assistant for football tactics,” Nature Communications, vol. 15, no. 1, 2024.

[13] F. Wu et al., “A survey on video action recognition in sports: datasets, methods and applications,” IEEE Transactions on Multimedia, vol. 25, pp. 7943–7966, 2023.

[14] H. W. Kuhn, “The Hungarian method for the assignment problem,” Naval Research Logistics Quarterly, vol. 2, no. 1–2, pp. 83–97, 1955.