Journal of Information Analytics, 1(2), 14-20

Journal of Information Analytics

Submit	25.12.2025
Revision	09.01.2026
Accept	09.01.2026
Publish	10.01.2026

Open Access

Research Article

Data-Driven Soccer Ball Trajectory Prediction Using Recurrent Neural Networks

Hanan Abu Kwaider, Faruk Serin*

Department of Computer Engineering, Mersin University, Mersin, Türkiye

Abstract

Accurate prediction of soccer ball trajectories enables advanced tactical analysis and real-time sports insights. This study proposes a deep learning approach for forecasting future ball positions from real match footage. Using the SAM2 segmentation model, frame-by-frame ball coordinates were extracted and refined through frame-reduction methods. Bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Unit (GRU) were trained with ball velocity as an additional feature. GRU achieved the lowest one-step-ahead Euclidean error while maintaining a consistent MAPE of 0.03 across both methods. Keyframe extraction yielded greater stability over longer horizons (1–10 steps). The approach demonstrates superior accuracy while operating directly on real-world footage, underscoring its potential for real-time soccer analytics and AI-assisted coaching systems.

Keywords: Soccer ball trajectory prediction, BiLSTM, GRU, Motion-based frame reduction, Keyframe extraction, Sports analytics

Citation: Abu Kwaider, H. and Serin, F., (2025). Data-Driven Soccer Ball Trajectory Prediction Using Recurrent Neural Networks. *Journal of Information Analytics*, 1(2), 14-20.

*Corresponding Author: Hanan Abu Kwaider hanan.abuquader@gmail.com

A version of this work was presented in abstract form at *The Fibonacci International Conference on Engineering, Technology, and Mathematics*.

1. Introduction

The analysis of ball movements in sports has become an important research area, driven by advances in computer vision and deep learning. In soccer, predicting future ball positions provides valuable insights for tactical analysis, strategic planning, and the development of AI-assisted coaching (Naik et al., 2022). Recent progress in object detection, segmentation, and tracking has enabled the transformation of raw match footage into structured data, allowing for more accurate modeling of ball trajectories and game dynamics (Juhai, 2024).

Despite these advances, ball trajectory prediction remains challenging due to the high variability of motion—such as rapid speed changes, sudden direction shifts, and frequent occlusions. These dynamics require models capable of capturing complex temporal dependencies across sequential frames, especially under real-world match conditions.

This study introduces a deep learning-based pipeline for forecasting soccer ball trajectories from real match footage. The proposed pipeline integrates segmentation-based tracking using SAM2 (Ravi et al., 2024) with recurrent neural networks—Bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Unit (GRU) architectures (Chung et al., 2014), (Schuster & Paliwal, 1997)—to learn temporal motion patterns from sequential ball positions. To improve data quality and reduce redundancy, two frame-reduction methods are explored: keyframe extraction using PySceneDetect AdaptiveDetector (Detection Algorithms — PySceneDetect 0.6.6 Documentation, n.d.), and motion-based thresholding that filters frames with minimal displacement. Additionally, ball velocity is calculated and incorporated as an input feature to enhance motion representation.

This approach contributes to advancing sports analytics by offering a reliable and scalable method for short-term trajectory forecasting, with potential applications in tactical evaluation and AI-driven coaching systems.

2. Literature Review

With the growing adoption of computer vision and machine learning in sports analytics, numerous studies have addressed ball tracking and trajectory prediction across different sports. These works have established solid foundations for analyzing match footage and forecasting in-game events, but often face limitations such as reduced accuracy over long distances, challenges with real-world data, or a narrow focus on specific aspects of play.

Yutaro et al (2022) developed a model for predicting soccer passes by combining a 3D CNN, LSTM, and Transformer encoder to learn spatial relations between players and the ball. While effective for short-distance passes, it struggled with longer-range predictions.

Anar & Hande (2022) used optical tracking data and a neural network to estimate the ball's 2D position based on player behavior, achieving R^2 scores of 0.79 (x-axis) and 0.92 (y-axis), though performance declined when possession changed over large distances.

Samriddha (2021) proposed a dual-model framework—the Opponent Proximity and Pass Region models—to predict pass recipients from abstract top-view visualizations, achieving strong performance on a custom dataset.

Yang et al (2023) designed a stereo vision-based system integrating an ANN for detection and 3D trajectory estimation, yielding mean errors of 29.6 cm (x), 7.2 cm (y), and 11.7 cm (z) in simulation.

Kim et al (2023) applied Set Transformers with BiLSTM networks to infer ball motion from player trajectories, obtaining less than 3.7 m mean position error and 64.7% player-ball possession accuracy.

Overall, prior approaches demonstrate valuable progress but typically rely on player tracking or simulated data. This study addresses these limitations by predicting ball trajectories directly from real match footage using segmentation-based tracking and recurrent neural networks.

3. Methodology

This study presents a deep learning-based approach for predicting future soccer ball positions from real match footage. The workflow includes data collection, segmentation, and tracking using the SAM2 model, data cleaning and feature engineering, and model training with BiLSTM and GRU architectures. Each model forecasts the ball's motion over short temporal windows and is evaluated across multiple prediction horizons using standard error metrics.

3.1. Data Collection

Match videos were obtained from publicly available YouTube sources. Separate datasets were prepared for training and testing to ensure model generalization. Each video was converted into individual frames and processed using the SAM2 segmentation and tracking model. Prior to running SAM2, players and the ball were manually labeled to improve segmentation accuracy. SAM2 then generated JSON outputs containing frame-level information such as object labels, bounding boxes, and centroid coordinates. For this study, only ball-related data were extracted for subsequent preprocessing and modeling.

3.2. Data Preparation and Preprocessing

While SAM2 effectively tracked the ball, occasional failures produced missing coordinates (x, y = 0, 0). These values were corrected using linear interpolation to preserve temporal continuity. Rare visual misidentifications, typically limited to two or three frames per video, had a negligible impact on data quality.

To minimize redundancy and enhance efficiency, two frame-reduction strategies were applied. The first used PySceneDetect's AdaptiveDetector to identify scene changes between consecutive frames. A threshold of 2.0 was empirically selected to eliminate redundant frames while preserving motion continuity. The second strategy employed motion-based thresholding, discarding frames with ball displacement below 8 pixels to remove near-static intervals. Threshold values were determined experimentally to balance information retention and compactness.

Following frame reduction, positional data were Min-Max normalized across training and testing sets. A velocity feature, derived from frame-to-frame displacement, was added to the normalized x and y values, forming a three-dimensional feature vector per time step. Finally, data were organized into 10-frame input sequences to predict ball positions for n future steps (1–10). Predictions were expressed in frames rather than seconds due to variability in ball speed.

3.3. Modeling

Recurrent neural networks (RNNs) are well-suited for sequential data modeling, making them appropriate for forecasting soccer ball trajectories (Wang et al., 2017). In this study, two architectures were implemented: Bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Unit (GRU).

BiLSTM captures temporal dependencies from both past and future contexts (Xu et al., 2024), while GRU offers a simpler, more computationally efficient alternative capable of modeling nonlinear motion patterns. Comparing both allows examination of the trade-off between accuracy and model complexity. Both models followed the same structure:

- Input layer
- Recurrent layer with 128 units (bidirectional for BiLSTM, unidirectional for GRU),
- Dropout layer (rate = 0.15),
- Second recurrent layer with 64 units,
- Another dropout,
- Dense layer with 32 ReLU-activated neurons,
- Final dropout layer.
- The output layer produced 2 × n values representing the x and y coordinates of the ball across n future frames.

Training used the Adam optimizer (learning rate = 0.0001) with the Huber loss function for 150 epochs, a batch size of 16, and a 10% validation split.

4. Results and Discussions

The accuracy of the models was evaluated using 1-step-ahead trajectory predictions under both preprocessing strategies—keyframe extraction and motion-based thresholding, and metrics including Mean Absolute Percentage Error (MAPE), R^2 for x and y, and the average Euclidean distance error.

The BiLSTM and GRU models were evaluated across prediction horizons ranging from 1 to 10 frames using two preprocessing strategies: keyframe extraction and motion-based thresholding. Performance was assessed through average Euclidean distance error, R^2 scores for the x and y coordinates, combined R^2, and MAPE. Tables 1 and 2 present representative results for 1-, 5-, and 10-step-ahead predictions, illustrating short-, mid-, and long-term forecasting performance for threshold-based and keyframe-based, respectively.

Table 1. Prediction Performance for Trajectories Obtained Using Threshold-Based Frame Reduction

n_future	Model	Avg Euclidean Error (px)	R^2 X	R^2 Y	Combined R^2	MAPE
1	BiLSTM	20.95	0.9829	0.9841	0.9975	0.03
1	GRU	26.18	0.9775	0.9571	0.9958	0.03
5	BiLSTM	65.35	0.7724	0.9041	0.9721	0.05
5	GRU	67.22	0.8071	0.8709	0.9742	0.06
10	BiLSTM	90.27	0.5451	0.8331	0.9459	0.07
10	GRU	111.86	0.3357	0.7944	0.9227	0.07

Table 2. Prediction Performance for Keyframe-Extracted Trajectories

n_future	*Model*	*Avg Euclidean Error (px)*	*R^2 X*	*R^2 Y*	*Combined R^2*	MAPE
1	BiLSTM	47.04	0.8342	0.9862	0.9893	0.02
1	GRU	37.46	0.884	0.9726	0.9917	0.03
5	BiLSTM	50.15	0.7723	0.9498	0.984	0.03
5	GRU	58.59	0.7076	0.9404	0.9798	0.04
10	BiLSTM	86.57	0.259	0.9159	0.952	0.05
10	GRU	78.58	0.4138	0.9061	0.9607	0.04

Both models achieved high short-term accuracy, with error gradually increasing as prediction horizons extended. Under the keyframe extraction approach,

- GRU exhibited superior short-term performance (steps 1–4),
- while BiLSTM maintained higher stability across medium-range horizons (steps 5–9).
- At the longest horizon (n = 10), GRU slightly regained advantage, suggesting better long-term generalization.

Using threshold-based reduction, GRU consistently achieved lower Euclidean errors and higher R^2 scores than BiLSTM across most horizons, particularly for short-term forecasts. The performance gap narrowed mid-range but widened again at extended horizons, indicating that GRU benefits more from compact, motion-focused inputs. These findings emphasize that frame-reduction methods influence not only model accuracy but also their comparative strengths—GRU excelling in short-range prediction and BiLSTM showing better temporal consistency.

Figures 1–2 visualize 1-step-ahead predictions for both models under keyframe and threshold-based preprocessing, confirming GRU's advantage in capturing immediate motion trends.

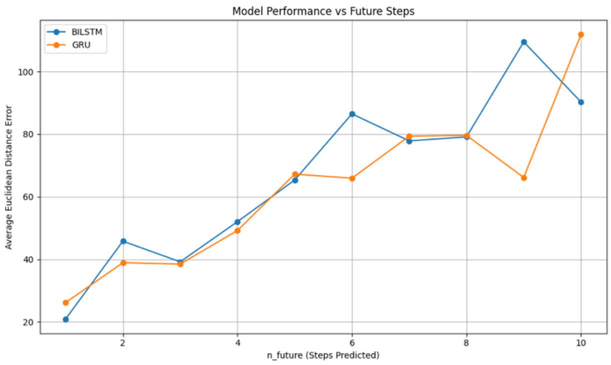

Figure 1 - 1-step-ahead prediction with threshold preprocessing

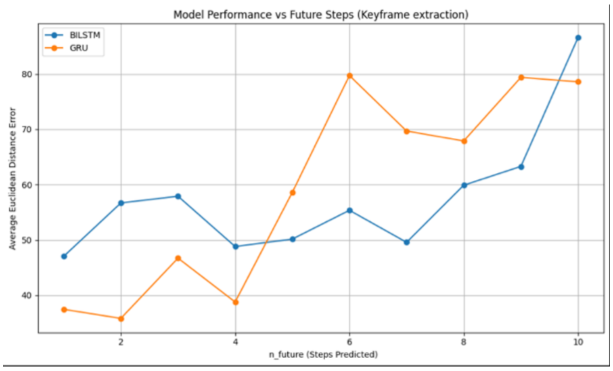

Figure 2- 1-step-ahead prediction with keyframe preprocessing

In practical terms, achieving sub-30 px average errors for immediate steps and below 90 px for 10-step forecasts demonstrates strong predictive capability for broadcast-resolution data. This precision supports reliable trajectory overlays and tactical visualization in live analytics environments.

5. Conclusion

This study introduced a deep learning approach for forecasting soccer ball trajectories directly from real match footage. The proposed pipeline integrates the SAM2 segmentation model with BiLSTM and GRU architectures to learn motion dynamics from ball trajectories obtained from videos with two frame-reduction strategies. The approach is lightweight, independent of player tracking, and suitable for real-world sports analytics applications.

Experimental results demonstrate that the proposed approach achieves strong predictive accuracy while operating directly on unstructured video data, highlighting its practical potential for real-time analytics and coaching applications.

Potential extensions of this work involve improving segmentation robustness under occlusions and camera motion, expanding the dataset, and incorporating player trajectory prediction for more comprehensive tactical modeling.

References

Anar, A., & Hande, A. (2022). Prediction of the ball location on the 2D plane in football using optical tracking data. Academic Platform Journal of Engineering and Smart Systems. https://doi.org/10.21541/apjess.1060725

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling (No. arXiv:1412.3555). arXiv. https://doi.org/10.48550/arXiv.1412.3555

Detection Algorithms—PySceneDetect 0.6.6 documentation. (n.d.). Retrieved May 5, 2025, from https://www.scenedetect.com/docs/latest/api/detectors.html#scenedetect.detectors.adaptive_detector.AdaptiveDetector

Juhai, W. (2024). Analyzing players' tactical positions and movement trajectories in soccer matches using machine vision algorithms. Journal of Electrical Systems. https://doi.org/10.52783/jes.3126

Kim, H., Choi, H.-J., Kim, C. J., Yoon, J., & Ko, S.-K. (2023). Ball trajectory inference from multi-agent sports contexts using set transformer and hierarchical Bi-LSTM. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 4296–4307. https://doi.org/10.1145/3580305.3599779

Naik, B. T., Hashmi, M. F., & Bokde, N. D. (2022). A comprehensive review of computer vision in sports: Open issues, future trends and research directions. Applied Sciences, 12(9), Article 9. https://doi.org/10.3390/app12094429

Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., & Feichtenhofer, C. (2024). SAM 2: segment anything in images and videos (No. arXiv:2408.00714). arXiv. https://doi.org/10.48550/arXiv.2408.00714

Samriddha, S. (2021). Who will receive the ball? Predicting pass recipient in soccer videos. Journal of Visual Communication and Image Representation. https://doi.org/10.1016/j.jvcir.2021.103190

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673–2681. https://doi.org/10.1109/78.650093

Wang, L., z, l, & Yi, Z. (2017). Trajectory predictor by using recurrent neural networks in visual tracking. IEEE Transactions on Cybernetics, PP, 1–12. https://doi.org/10.1109/TCYB.2017.2705345

Xu, Y., Pan, Q., Wang, Z., & Hu, B. (2024). A novel trajectory prediction method based on CNN, BiLSTM, and multi-head attention mechanism. Aerospace, 11(10), Article 10. https://doi.org/10.3390/aerospace11100822

Yang, Y., Kim, D., & Choi, D. (2023). Ball tracking and trajectory prediction system for tennis robots. Journal of Computational Design and Engineering, 10(3), 1176–1184. https://doi.org/10.1093/jcde/qwad054

Yutaro, H., Rei, K., Ryota, Y., Kenta, K., & T., N. (2022). Pass Receiver prediction in soccer using video and players' trajectories. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Wo Rkshops (CVPRW). https://doi.org/10.1109/CVPRW56347.2022.00394