

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology



أكاديمية كاوست
KAUST ACADEMY

Beyond Spotting: Seq2Seq Event Forecasting of Soccer Ball Actions

A Benchmark of Seq2Seq Models (Random, LSTM, Transformer) for Soccer Ball Action Forecasting

---

Banan Almdni
balmdn200@gmail.com

Shouq Alsulami
shoog.raja@gmail.com

Talah Almuhawwis
talahalmuhawwis@gmial.com

August 27, 2025

Abstract

The prediction of future events in football represents a complex challenge due to the dynamic and stochastic nature of the game. This study investigates the application of sequence-to-sequence (seq2seq) modeling, originally developed for Natural Language Processing (NLP), to forecast in-game football actions. Using automatically generated event annotations from the SoccerNet Ball Action Spotting Challenge, we model football actions (e.g., passes, shots, fouls) as sequential tokens and formulate the task as predicting the immediate subsequent action given a historical context of five consecutive events. [3]

We implement and evaluate a Transformer-based seq2seq model, benchmarking its performance against Transformer and Long Short-Term Memory (LSTM) networks. Experimental results demonstrate that sequence models substantially outperform random and frequency-based baselines. Among the tested approaches, the Transformer architecture achieved the highest performance, with a Top-1 accuracy of 67% and Top-3 accuracy of 90%, surpassing the LSTM baseline (60% and 84%, respectively). These findings highlight the effectiveness of attention mechanisms in capturing complex temporal dependencies within football event data.

Furthermore, evaluation through Top-3 accuracy and Macro-F1 score addressed the challenges of class imbalance and the multi-outcome nature of football, providing a robust assessment framework. This research contributes to the growing field of sports analytics by demonstrating the feasibility and advantages of adapting language-modeling techniques for football event prediction. The proposed framework offers a foundation for advancing predictive modeling in sports, with potential applications in tactical analysis, match preparation, and real-time decision support systems.

Acknowledgments

The author acknowledges the use of artificial intelligence (AI) tools in the preparation of this report. These tools were employed to assist in refining the language, enhancing clarity, and improving the overall readability of the text. All interpretations, analyses, and conclusions presented remain the sole responsibility of the author.

Contents

Abstract **1**

1 Introduction **3**
1.1 Project Objectives and Goals 3
1.2 Problem Statement 4

2 System Overview **4**
2.1 Broadcast Video Input 4
2.2 Ball Action Spotting and Event Extraction 5
2.3 Event Tokenization 5
2.4 Sequence-to-Sequence Modeling 5

3 Methodology **6**
3.1 Tools and Frameworks 6
3.2 Dataset and Preprocessing 6
3.3 Model Architectures 6
3.4 Training Procedure 6
3.5 Evaluation Metrics 6
3.6 Modeling Rationale 7
3.7 Resources and Tools 7

4 Results Discussion **7**
4.1 Quantitative Results 8
4.2 Comparison with Expectations 8
4.3 Bottlenecks and Suggestions for Improvement 8
4.4 Overall System Evaluation 8

5 Future work **8**

6 Conclusion **9**

1 Introduction

Football (soccer) is a dynamic and complex sport where every event—whether a pass, foul, or shot—can significantly alter the course of the match. Being able to forecast future events in a match has growing significance in multiple real-world domains:

1. **Sports analytics:** Event forecasting provides tactical insights by modeling the likelihood of upcoming actions (e.g., whether a passing sequence is likely to culminate in a shot or be interrupted by a foul). Such predictions can support coaches and analysts in match preparation and in-game decision-making.

2. **Broadcasting and entertainment:** Predictive systems can be integrated into live broadcasts to provide audiences with richer narratives, such as probabilities of certain events occurring in the next phase of play.

3. **Player evaluation and scouting:** By forecasting likely outcomes of player actions, clubs can assess individual contributions in contexts beyond goals and assists, offering more nuanced performance analysis.

4. **Academic research:** Forecasting tasks in football are relatively unexplored compared to trajectory prediction or player tracking. By treating actions as sequences, this project contributes to a novel intersection of sports analytics and natural language processing (NLP) methodologies.

Despite its potential, football event forecasting has been limited by the lack of large, well-annotated datasets. Manual event annotation is time-consuming and costly. Recent advances in SoccerNet's Ball Action Spotting Challenge provide automatically generated action annotations, which, while not perfectly accurate, offer scalable pseudo ground truth that makes predictive modeling feasible on a larger scale.

1.1 Project Objectives and Goals

The overarching aim of this project is to explore whether sequence modeling techniques can accurately forecast future football events using automatically detected annotations. The work pursues several concrete objectives:

1. Seq2Seq Transformer Development: To build a Transformer-based sequence-to-sequence model that treats football events as tokens, drawing inspiration from language modeling in NLP.

2. Benchmarking against baselines: To evaluate the Transformer against multiple baselines, including random predictors, frequency-based (weighted random) predictors, Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks.

3. Evaluation using robust metrics: To address class imbalance and the multi-possibility nature of football, we evaluate the models using Macro-F1 score, Top-1 accuracy, and Top-3 accuracy. The Top-3 metric is particularly important, as it reflects the realistic uncertainty in football where multiple outcomes are plausible.

4. Investigation of class imbalance effects: To analyze how the imbalance across the 12 unique actions influences model learning and performance.

Ultimately, the goal is not only to build a working predictive model, but also to assess the strengths and weaknesses of different sequence modeling paradigms in this novel domain.

1.2 Problem Statement

This research investigates the challenge of next-event prediction in football, where the task is to forecast the subsequent ball action based on a sequence of five preceding actions obtained from the SoccerNet Ball Action Spotting model. The problem lies in determining which sequence modeling paradigm—LSTM or Transformer—can most effectively capture temporal dependencies and uncertainties in football dynamics to deliver accurate and reliable predictions.

2 System Overview

The proposed system is designed to forecast upcoming football events by leveraging annotated match data from the SoccerNet Ball Action Spotting Challenge and processing them through sequence-to-sequence (seq2seq) neural models. The architecture of the system is illustrated in Figure 1, which outlines the key stages: data acquisition from broadcast videos, automatic event spotting, event tokenization, sequence modeling, and final action prediction.

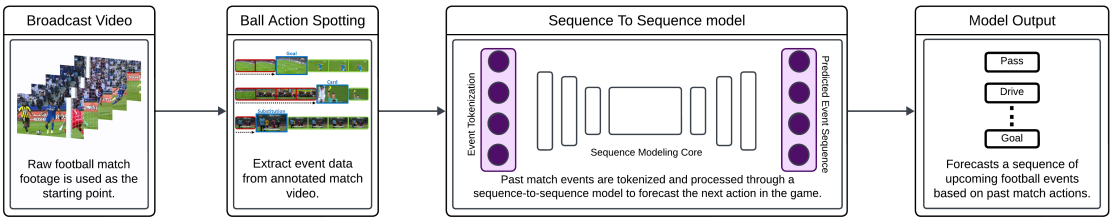


Figure 1: Proposed system architecture for football event

2.1 Broadcast Video Input

The system begins with raw football match footage as its primary input. Broadcast video provides a continuous and naturalistic record of all in-game actions but is not immediately suitable for predictive modeling. Directly processing video requires large-scale spatio-temporal learning, which is computationally expensive and noisy. To overcome this challenge, the system leverages pre-annotated event data from SoccerNet's Ball Action Spotting model, which extracts semantically meaningful football actions (e.g., pass, foul, substitution, shot).

2.2 Ball Action Spotting and Event Extraction

In the second stage, events are automatically detected from the broadcast videos using the SoccerNet Ball Action Spotting Challenge 2025 model. This step converts raw, unstructured video into structured, categorical information by assigning timestamps and labels to key actions. [2]

- Each event belongs to one of 12 unique classes (e.g., pass, foul, shot, goal).
- The total dataset includes 12,428 events, which serve as the foundation for our sequence modeling task.

These annotations serve as a pseudo ground truth: while they may contain minor inaccuracies due to model limitations, they provide a scalable and standardized event stream that makes downstream predictive modeling possible.

2.3 Event Tokenization

Once events are extracted, they are treated analogously to tokens in a natural language processing (NLP) task. This transformation is inspired by the analogy between football action sequences and sentences:

- A sequence of five consecutive events is considered the context (similar to a phrase or sentence fragment).
- The next event in the match is the prediction target (similar to predicting the next word in a sentence).

Tokenization allows sequence models (e.g., LSTMs and Transformers) to process football events in the same way language models process text.

2.4 Sequence-to-Sequence Modeling

The core of the system lies in the sequence-to-sequence (seq2seq) modeling component. Here, the tokenized events are passed through a neural architecture that learns temporal dependencies and patterns in football event sequences. Multiple architectures are tested and compared, including:

- Random and Weighted Random Predictors (baselines).
- Long Short-Term Memory (LSTM) network.
- Transformer Seq2Seq model (proposed system).

The Transformer model is the primary contribution, as it uses self-attention mechanisms to capture both short- and long-range dependencies within football sequences. This is particularly advantageous compared to RNNs and LSTMs, which may struggle with long-context modeling.

3 Methodology

The objective of this work is to predict the next ball-related action in a football match sequence using outputs from the SoccerNet Ball Action Spotting Challenge 2025. We formulate the problem as a sequence-to-sequence prediction task, where actions are treated as discrete tokens, similar to words in natural language processing. This approach enables neural models to capture temporal dependencies and contextual patterns across sequences of events.

3.1 Tools and Frameworks

The model's were implemented using TensorFlow/Keras for model development, with NumPy and scikit-learn for preprocessing, token encoding, and evaluation. All training was conducted in Google Colab with GPU acceleration to handle computational demands.

3.2 Dataset and Preprocessing

The dataset consists of automatically annotated ball actions provided by SoccerNet. Each action was encoded into integer tokens, and training pairs were constructed by segmenting sequences into windows of five actions, with the following action as the target. To address class imbalance between frequent events (e.g., passes) and rare but critical ones (e.g., goals), we applied weighted categorical cross-entropy during training. [1]

3.3 Model Architectures

Two sequence modeling approaches were benchmarked:
LSTM (Recurrent Encoder): A single LSTM layer with 128 hidden units followed by dropout (0.5) and a dense softmax output layer. This architecture is effective for modeling short- to medium-term temporal dependencies.[4]
Transformer (Encoder-Only): An encoder with 4 layers, each with 8 self-attention heads, model dimension of 128, feedforward size 512, and dropout 0.15. Positional embeddings preserved temporal ordering, and the next action was predicted from the last token representation. The Transformer is particularly suited for long-range dependencies and complex temporal dynamics.

3.4 Training Procedure

Both models were implemented in TensorFlow/Keras and trained on Google Colab with GPU acceleration. Optimization was performed with the Adam optimizer, using ReduceLROnPlateau for learning rate scheduling. Early stopping and checkpointing were based on Top-3 Accuracy to prevent overfitting.

3.5 Evaluation Metrics

Model performance was assessed using the following:

- **Top-1 Accuracy:** measures exact prediction correctness.
- **Top-3 Accuracy:** accounts for scenarios where multiple outcomes are plausible, reflecting the dynamics of football in the real world.
- **Weighted Macro-F1:**evaluates class-wise balance, ensuring that rare but crucial events are not overshadowed by frequent actions.

3.6 Modeling Rationale

By treating actions as tokens, we leverage sequence modeling techniques from NLP to capture temporal and contextual dependencies in football. Furthermore, reporting Top-k metrics reflects the real-world uncertainty of match events, where several outcomes may be plausible at any point.

This methodology allows a systematic comparison between recurrent and attention-based architectures under a unified experimental framework, while also accounting for dataset imbalance and real-world interpretability.

3.7 Resources and Tools

This project was developed using a combination of publicly available datasets, deep learning frameworks, and cloud-based resources:

- **Dataset**: The SoccerNet Ball Action Spotting Challenge 2025 dataset, consisting of 7 EPL matches, 12,428 annotated ball actions, and 12 unique action classes (e.g., pass, shot, foul, goal). These annotations, although pseudo ground truth, provide a scalable and standardized basis for predictive modeling.
- **Development Frameworks:**
  - **TensorFlow/Keras** – for implementing, training, and evaluating the sequence models (LSTM and Transformer).
  - **NumPy and Pandas** – for data preprocessing, tokenization, and handling numerical operations.
  - **scikit-learn** – for additional preprocessing utilities and evaluation metric calculations.
- **Training Environment**: All experiments were conducted on Google Colab with GPU acceleration, which enabled efficient model training and hyperparameter tuning without requiring dedicated local hardware.

4 Results Discussion

In this study, we evaluated four models **Random baseline, Weighted Random, LSTM, and Transformer** for predicting the next ball action in football sequences. Actions were treated as discrete tokens, analogous to words in a sentence, allowing sequence models to capture temporal dependencies and patterns in the order of events.

4.1 Quantitative Results

The following table summarizes the Top-1 and Top-3 accuracy of each model.

Model	Top-1 Accuracy	Top-3 Accuracy	Weighted F1
Random	8%	26%	0.8
Weighted Random	40%	80%	0.39
LSTM	60%	84%	0.61
Transformer	67%	90%	0.63

Table 1: Quantitative results for next-action prediction across different models.

4.2 Comparison with Expectations

Overall, the models behaved largely as expected. Treating actions like tokens allowed sequence models to capture temporal dependencies effectively, resulting in high predictive performance. The Transformer exceeded expectations, achieving the highest Top-3 accuracy (90%), which confirms that multi-head self-attention can effectively model complex sequential patterns. The LSTM performed well on frequent actions, as anticipated, but underperformed on rare events, highlighting its limitations in capturing long-range dependencies. Random baselines provided reference points, confirming that modeling sequences is essential for accurate next-action prediction.

4.3 Bottlenecks and Suggestions for Improvement

- **Class Imbalance:** Rare actions (e.g., goals, free kicks) have low recall. Using oversampling, focal loss, or data augmentation could improve performance.
- **Limited Temporal Features:** Only sequences of actions were used. Adding time deltas, player positions, or contextual features may enhance predictions.
- **Data Requirements:** Transformers require large datasets; for smaller datasets, LSTM or hybrid models may generalize better.
- **Prediction Uncertainty:** Current predictions rely on Top-k accuracy. Probabilistic modeling or beam search could improve multi-action predictions.

4.4 Overall System Evaluation

Overall, the system effectively models football action sequences. The Transformer's high Top-3 accuracy (90%) demonstrates that treating actions as tokens is a valid and effective approach for next-action prediction. Main bottlenecks relate to rare event classes and limited contextual features, offering clear avenues for improvement.

5 Future work

The findings of this study lay the groundwork for a number of compelling research trajectories, each of which presents an opportunity to further advance the field. We could

explore longer sequences to improve long-term forecasting accuracy and incorporate multi-modal inputs by combining event data with visual and player tracking data to add spatial and contextual cues. A key next step is to predict not just the action type, but also the timing and location simultaneously (a "when, what, where" model) to create a more comprehensive predictive system. Finally, we can streamline the entire pipeline to enable real-time analytics for automatic in-match predictions and develop dedicated coaching tools that extend the system for advanced match analysis and decision support.

6 Conclusion

To conclude, this project successfully demonstrated the viability of using sequence-to-sequence (seq2seq) models to forecast future ball actions in soccer matches. By treating football events as a language, we were able to apply methodologies from Natural Language Processing (NLP) to a novel domain, with our results showing a clear superiority of the sequence models over simpler baselines. The Transformer model stood out as the most effective, achieving a Top-1 Accuracy of 67% and a Top-3 Accuracy of 90%, significantly outperforming the LSTM's 60% and 84% respectively, and far exceeding the Random (8%) and Weighted Random (40%) baselines. This performance validates the Transformer's ability to capture complex, long-range dependencies within the match data. The use of robust metrics like Top-3 Accuracy and Macro-F1 score provided a more realistic and nuanced evaluation, confirming that our model's predictions are not only precise but also account for the inherent unpredictability of a real-world game.

References

[1] Anthony Cioppa et al. *SoccerNet 2024 Challenges Results*. Sept. 2024. DOI: 10.48550/arXiv.2409.10587.

[2] Adrien Deliège et al. "SoccerNet-v2 : A Dataset and Benchmarks for Holistic Understanding of Broadcast Soccer Videos". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2021.

[3] Tiago Mendes-Neves, Luís Meireles, and João Mendes-Moreira. *Forecasting Events in Soccer Matches Through Language*. 2024. arXiv: 2402.06820 [cs.LG]. URL: https://arxiv.org/abs/2402.06820.

[4] Qiyun Zhang et al. "Sports match prediction model for training and exercise using attention-based LSTM network". In: *Digital Communications and Networks* 8 (Aug. 2021). DOI: 10.1016/j.dcan.2021.08.008.