# Pass2Formation: Multi-Receiver, Full-Team Next-Location Forecasting from Short Event Sequences in Soccer Games

Paper Track: Soccer
Paper ID: 163

## 1. Introduction

Predicting the complete next state of play from the current state in soccer is essential not only for tactical and counterfactual analysis, forecasting play outcomes, and optimizing decision-making, but also for powering the core mechanics of popular soccer video games such as "*FIFA*" and "*Score! Hero*". Accurately anticipating the next state enables these systems to generate smooth, realistic, and context-aware in-game actions, resulting in more natural, continuous, and intelligent machine behavior throughout gameplay. However, achieving such prediction is inherently challenging, as it requires evaluating possession and passing decisions while simultaneously performing detailed spatio-temporal multi-agent analysis of highly dynamic and rapidly evolving game states.

The foundational frameworks like Expected Possession Value (EPV), which represents the likelihood of a team scoring or conceding the next goal at any time instance [1], or Valuing Actions by Estimating Probabilities (VAEP), which assigns values to individual player actions by estimating their impact on short-term scoring and conceding probabilities [2], have established the importance of quantifying on-ball actions. The Potential Pitch Control Field (PPCF) similarly complements these models by estimating the probability that a team can control the ball at any location on the pitch, incorporating player positions, velocities, and ball trajectory to capture spatial dominance [3]. However, these approaches remain limited to assigning scalar values or probability maps to the current possession, team dominance or passing options, without providing any explicit insight into the configuration of the next game state.

Several other models focus on predicting specific outcomes from the current game state. StatsBomb's Expected Pass (xPass) models estimate pass-completion probability based on factors such as starting location, distance, angle, pressure, and body part used [4]. Seq2Event, a Transformer/RNN-based approach, predicts six attributes of the next event—including action type and location—using prior events and contextual information [5]. Honda et al. further explored pass receiver prediction by integrating video features with player trajectories via 3D CNNs and LSTMs [6]. More recently, Wang et al. introduced TacticAI in collaboration with Liverpool FC and Google DeepMind, using geometric deep learning to analyze corner kicks and achieve top-3 receiver prediction by efficiently encoding pitch symmetries using geometric deep learning [7]. Despite these advances, such models remain limited to forecasting isolated components of the next state—such as completion probability, action type, receiver choice, or set-piece patterns—rather than addressing open-play passing scenarios.

The Baller2vec family has recently addressed next-state prediction in team sports. The original Baller2vec introduced a multi-entity Transformer for modeling multi-agent spatiotemporal dynamics in basketball [8], and then Baller2vec++ extended it with a look-ahead mechanism using specialized attention masks to capture dependencies among agent trajectories [9]. However, both models represent trajectories using discretized spatial grids rather than continuous coordinates, limiting the spatial precision needed for tactical analysis. Moreover, they also lack support for scenario expansion, where identical contexts can be evaluated under different receiver choices.

FootBots, an encoder–decoder Transformer leveraging equivariance, recently addressed motion prediction in soccer [10]. It models temporal and social dynamics through set-attention blocks and performs well on conditioned prediction tasks—such as forecasting player movement from ball position or predicting offensive motion from defensive structure. The same group later introduced TranSPORTmer, a unified Transformer framework for trajectory forecasting, imputation, ball inference, and state classification, also built on Set Attention Blocks to capture temporal and multi-agent interactions [11]. While achieving state-of-the-art results on multiple trajectory understanding tasks for soccer and basketball, both models focus on trajectory imputation and forecasting as separate objectives rather than providing deterministic next-state predictions across multiple receiver candidates.

Recently, SportsNGEN, a Transformer-decoder-based sports simulation engine trained on player and ball tracking data, was introduced to generate extended gameplay and evaluate counterfactual scenarios for coaching decisions. While initially applied to tennis, qualitative results suggest potential applicability to football [12]. However, SportsNGEN is tailored for long-horizon simulation rather than single-step next-state prediction, and its probabilistic sampling approach contrasts with the deterministic regression outputs required for direct scenario comparison.

In this work, we introduce a model that predicts the post-pass locations of all 22 players using only five prior events. By incorporating the positions of all players and the ball over these events, together with the details of the intended pass (passer, receiver, ball location, and pass features), we hypothesize that the immediate formations of both teams following the pass can be accurately estimated. This is particularly valuable for evaluating both real and alternative receiver scenarios, enabling richer tactical interpretation, counterfactual analysis, and seamless integration into game simulation systems.

# 2. Methods

## 2.1. Dataset

We trained, validated, and tested our proposed models on the PFF FC's 2022 World Cup publicly available dataset. To enable efficient retrieval of spatiotemporal and contextual information across all 64 matches, the JSON event files included in the dataset were organized into a relational database of multiple tables to structure the data: possession event features, player locations, ball locations, and player metadata.

The semi-finalists—France, Argentina, Morocco, and Croatia—along with England were selected to represent a diverse set of playing styles and tactical approaches, while also ensuring enough matches for each team. Accordingly, the 64 tournament matches were divided into two groups: Group A, containing all matches involving these reserved five teams (37 matches), and Group B, comprising the remaining matches (27 matches). Group B was used to train a generic baseline model applicable to any team, whereas Group A was used to fine-tune this baseline model, so it better reflects the tactics and playing style of each of the five selected teams.

Fig 1 summarizes the key steps of the data processing pipeline.

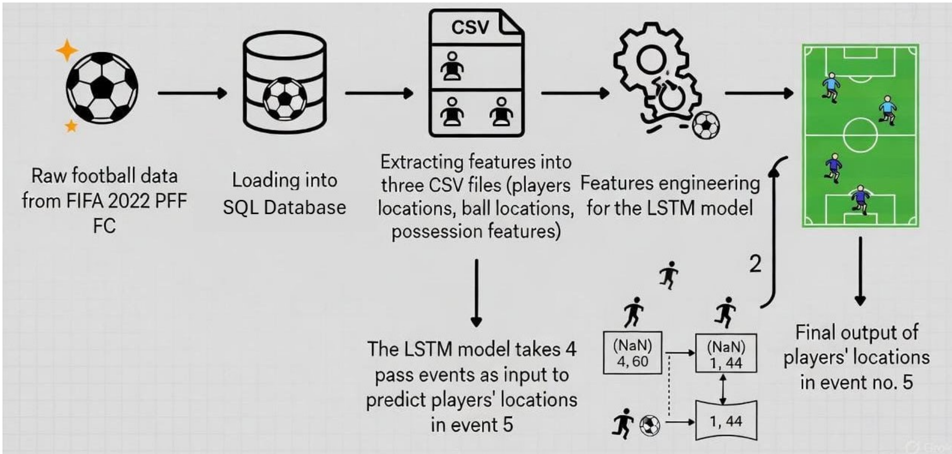


Figure 1. An overview of the data processing workflow.

### 2.1.1. Sequence Construction

All 64 matches were curated into different sequences, each sequence consists of N consecutive pass events, where N ranges between 3 and 8. Each pass event has 60 features: x and y coordinates of the 22 players' locations, ball coordinates (x,y,z), 6 pass contextual variables (passer id, receiver id, receiver location, pass type, pass outcome, pressure type), and 7 next-timestep context (next ball coordinates, next passer/receiver coordinates). Any sequence with missing or null values in any of these fields is excluded from downstream analysis to ensure data completeness. All positions were normalized to the pitch dimensions and a unified offense direction.

## 2.2. Model Design and Training

Fig. 2 presents a schematic of our proposed model that was developed as a sequence-to-next-state two-layers LSTM prediction model (128, 64 units with ReLU activation), batch normalization, dropout (30%), and a linear output layer. For an input sequence of N consecutive pass events, the model input is N×60 feature tensor, and the output is a vector of the next-event (post-pass) x and y positions for all 22 players, not including ball coordinates. The model was trained with Adam optimizer (lr=0.001), Mean Squared Error (MSE) loss masked on alternatives, batch size 64, early stopping, for up to 100 epochs. We split stratified by sequence into 80% train, 10% validation, 10% test sets, and employed the Mean Absolute Error (MAE) as our Evaluation Metric.

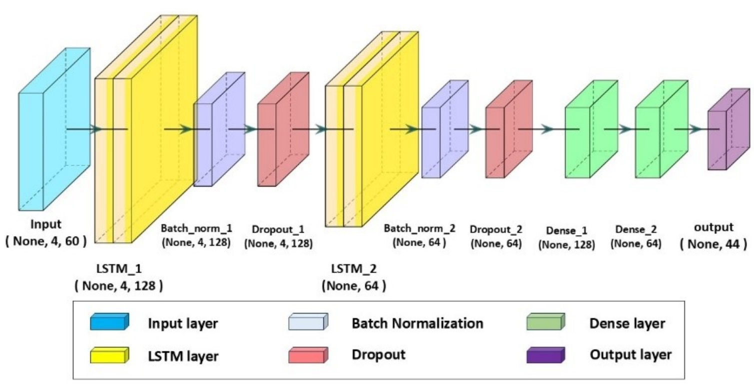


Figure 2. Schematic representation of the proposed model architecture.

## 2.3. Baseline Model

To pick the best number of consecutive pass events per sequence (N) to use, we ran a pilot study where we trained, validated and evaluated our baseline model on the matches of Group B explained earlier (i.e. all matches excluding the 5 reserved teams) for different N ranging between 3 and 8. This resulted in the best model performance/output at N=5, where the model is fed with consecutive pass events and outputs the location of the 22 players at the fifth event. Upon that, we fixed the number of consecutive pass events per sequence to be 5 across the remaining experiments of the study.

We then evaluated the 4-pass input baseline model systematically across the five reserved teams, i.e. France, England, Argentina, Croatia, and Morocco datasets. This cross-team evaluation protocol assesses the model's generalization capability to diverse tactical patterns, passing styles, and formation dynamics specific to each national team.

## 2.4. Fine-Tuned Models

Following cross-team evaluation of the baseline model, five team-specific fine-tuning experiments are conducted to adapt the pre-trained baseline model to the tactical and spatial characteristics of each reserved team. The fine-tuning process involves initializing each team-specific model with baseline model weights and then training exclusively on that team's data using a reduced learning rate strategy to preserve generalizable features learned during baseline training while adapting to team-specific patterns.

Five independent fine-tuned models are created, one for each reserved team: France, England, Argentina, Croatia, and Morocco. Each fine-tuning experiment follows an identical procedural pipeline with team-specific data substitution. However, unlike baseline training which uses a fixed learning rate, fine-tuning employs adaptive learning rate reduction to allow the model to make larger early adaptations when validation loss is high, then make finer adjustments as the model approaches convergence. Once the baseline model is fine-tuned for each reserved team, we apply a matrix-style cross-team evaluation to assess generalization. Each fine-tuned model is tested on its own data (within-team) as well as on all other teams' datasets (cross-team). This dual evaluation setup measures both the strength of team-specific adaptations and the model's ability to generalize these adaptations across different teams.

## 2.5. Detailed Prediction Analysis

Following model training, fine-tuning, and cross-team evaluation, detailed prediction analysis is conducted to decompose aggregate performance metrics into constituent components of different categories. During that, we evaluated the model predictions in terms of 1) Coefficient of Determination ($R^2$) which represents how well the model prediction match the observed data, and 2) Mean Average Error (MAE) which represents the magnitude of prediction error compared to the true state. Results are tracked across baseline to all fine-tuned models, and also within-team to cross-team evaluations.

### 2.5.1. Teams and Game Dynamics

In this section, we evaluate our models with respect to several key factors that influence both team behavior and overall game dynamics. Specifically, we examine whether prediction accuracy differs across player position groups (e.g., defenders, midfielders, forwards) within each team, between the possession team (actively controlling the ball) and the opponent team (defending), and across the various phases of the match—Period 1 (first half), Period 2 (second half), as well as Period 3 (first extra time), and Period 4 (second extra time) when extra time is played at the knockout matches.

### 2.5.2. Pass Features

In this section, we evaluate our models against pass type, pass outcome, and pressure applied—represent distinct tactical scenarios with potentially different formation complexity. Pass type (e.g., cutback, through ball, standard pass) reflects the intended ball-progression method. Pass outcome (e.g., complete, blocked, out of play) reflects the success or failure of the pass execution. Pressure type (e.g., player pressured, passing lane pressure, attempted pressure, no pressure) reflects defensive intensity. Pass-context-based error analysis examines whether prediction accuracy varies systematically across these contextual dimensions, revealing tactical scenarios where formation prediction becomes more or less challenging.

### 2.5.3. Player Movements and Locations

In this section, we evaluate our models with respect to players' average movement, team spatial dispersion—defined as the mean Euclidean distance of all team members from their spatial centroid—and players' on-pitch locations. These three factors capture key aspects of player mobility, team compactness, and positional context, all of which are critical to formation prediction difficulty and may systematically affect prediction accuracy.

# 3. Results

Fig. 2 shows the mean average error (MAE) of all players when predicting their positions using the baseline model, and the models fine-tuned on specific teams. While the baseline model has an almost uniform low performance across the different teams, each of the fine-tuned models exhibits better MAE for the team it was fine-tuned on.

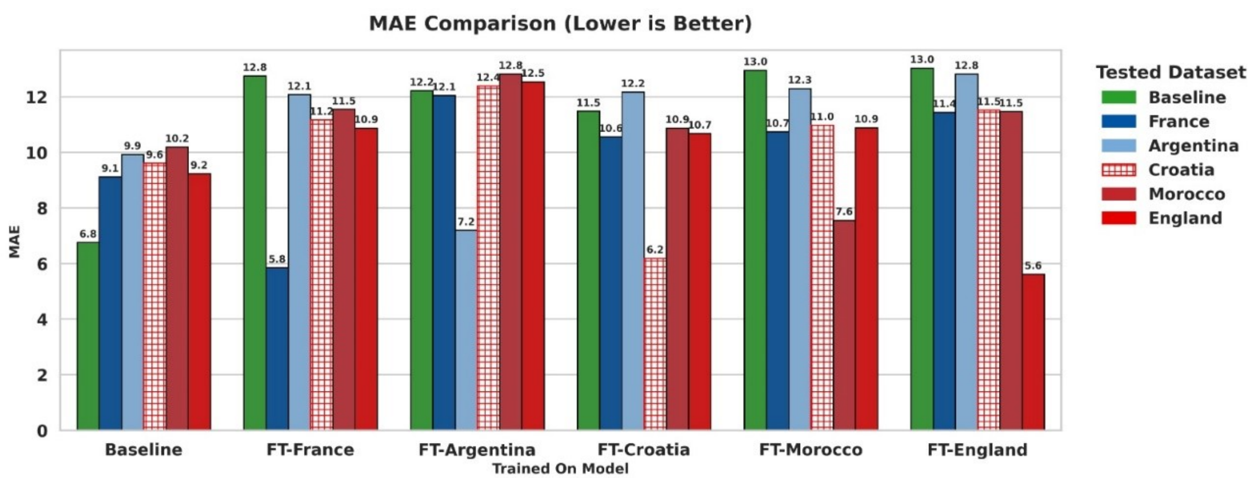


Figure 2. MAE results for the baseline and fine-tuned models.

The baseline model achieves consistent generalization across all five reserve teams. Cross-team MAE values range from 9.1m to 10.2m, representing a nearly 34% to 50% increase relative to baseline test performance (6.8m). This variance reflects differences in tactical diversity, passing patterns, and formation dynamics specific to each national team's play style. All $R^2$ values remain positive (0.69–0.71), indicating that the model captures meaningful variance in player positioning across teams despite being trained exclusively on non-reserved teams. On the other hand, all five team-specific models demonstrate substantial improvement over their baseline model equivalents, with MAE reductions ranging nearly from 26% to 39%. England achieves the largest improvement (39%, from 9.23m to 5.62m MAE), suggesting that its distinct tactical patterns—including compact defensive shape, wider progression and wing overloads —are well-captured through fine-tuning. Morocco shows the smallest improvement percentage (26%) but maintains $R^2$ = 0.778, indicating consistent prediction quality. All $R^2$ values exceed 0.77, demonstrating that fine-tuned models capture meaningful team-specific variance in player positioning.

Fig. 3 illustrates an example for the model's ability to predict complete future scenarios for different possible pass options. In Fig. 3a, the player positions at timestep 4 (before the pass) are shown, with colored arrows representing four hypothetical pass directions, including the actual pass (black arrow). Figure 3b then shows the resulting player positions after the actual pass is executed. Finally, Figs. 3c–3e present the predicted player positions for each of the hypothetical passes, demonstrating how the model anticipates player movement under alternative scenarios.

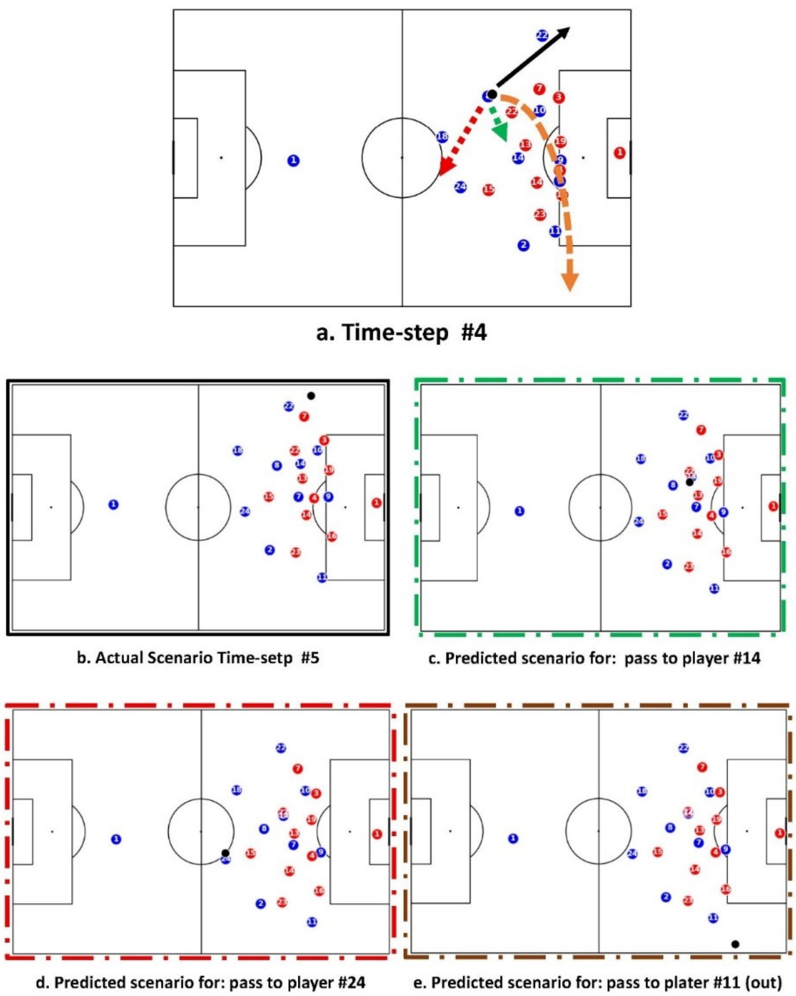


Figure 3. Predicted scenarios for different pass options. a) Player positions at timestep #4 immediately before the pass, b) Player positions following the execution of the actual pass (black arrow in panel a), c-e) Predicted player positions for hypothetical passes to players #14, #24, and #11, corresponding to the green, red, and orange arrows shown in panel (a).

Fig. 4 presents another example of the model's predictions for different pass scenarios. Panel (a) shows the player positions at timestep 4, along with colored arrows indicating the hypothetical pass options. In the black-arrow scenario, the model predicts that the pass would be intercepted, whereas in the gold-arrow scenario the pass is expected to be successfully completed. Finally, in the green-arrow scenario, the pass is predicted to result in an assist, leading to an expected shot.

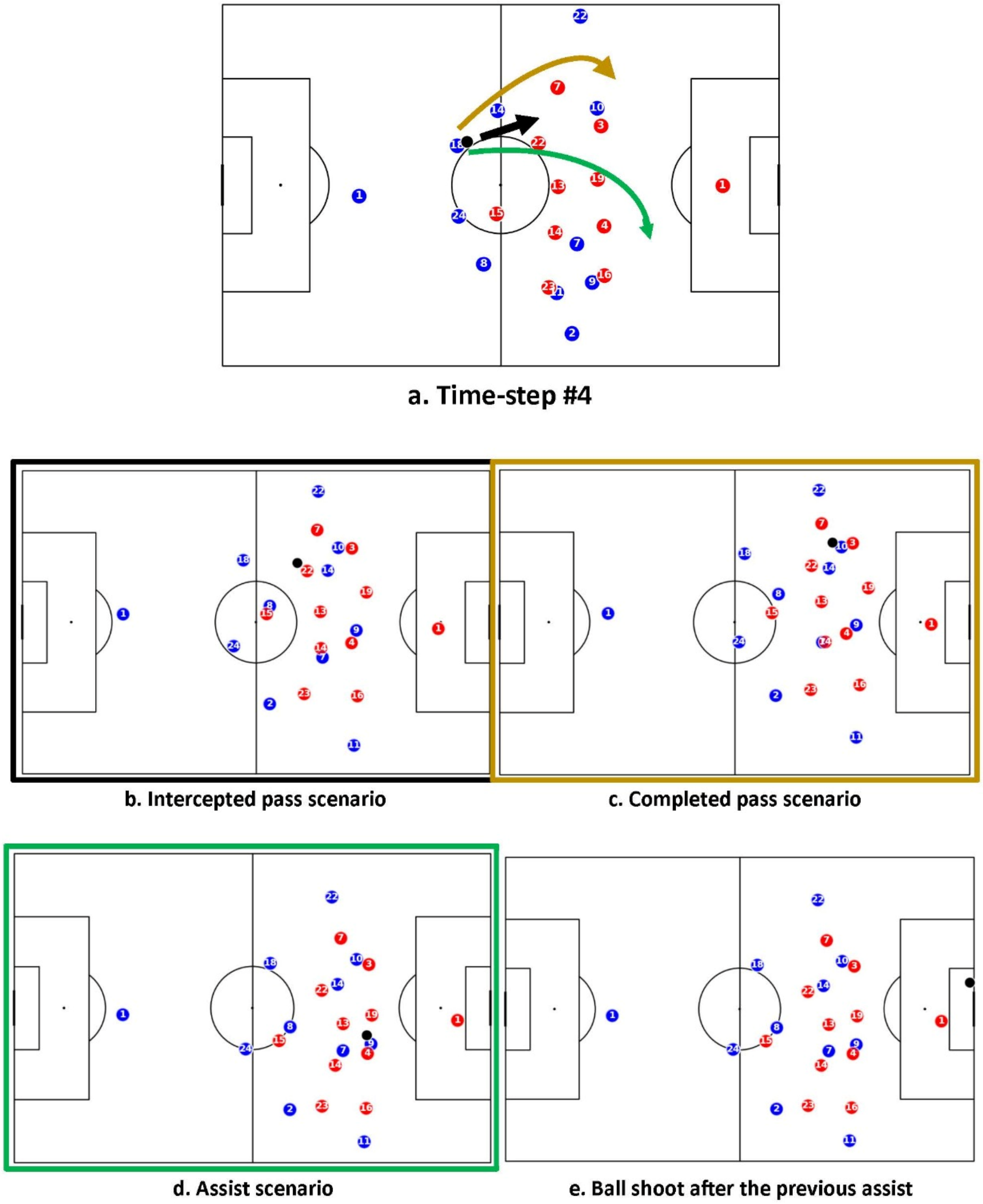


Figure 4. Predicted outcomes for different pass options. (a) Player positions at timestep 4, immediately before the pass. (b) Predicted player formation for the black-arrow pass option shown in panel (a), where the ball is expected to be intercepted. (c) Predicted formation for the gold-arrow pass option, where the pass is expected to be successfully completed. (d) Predicted formation for the green-arrow pass option, where the pass is expected to result in an assist followed by a direct shot on goal in (e).

Figure 5 presents the prediction error across different positional groups. Players whose roles involve relatively limited movements—such as goalkeepers and center-backs—show lower prediction error, whereas positions characterized by higher mobility and greater positional variability, such as wingers and midfielders, exhibit higher error levels.

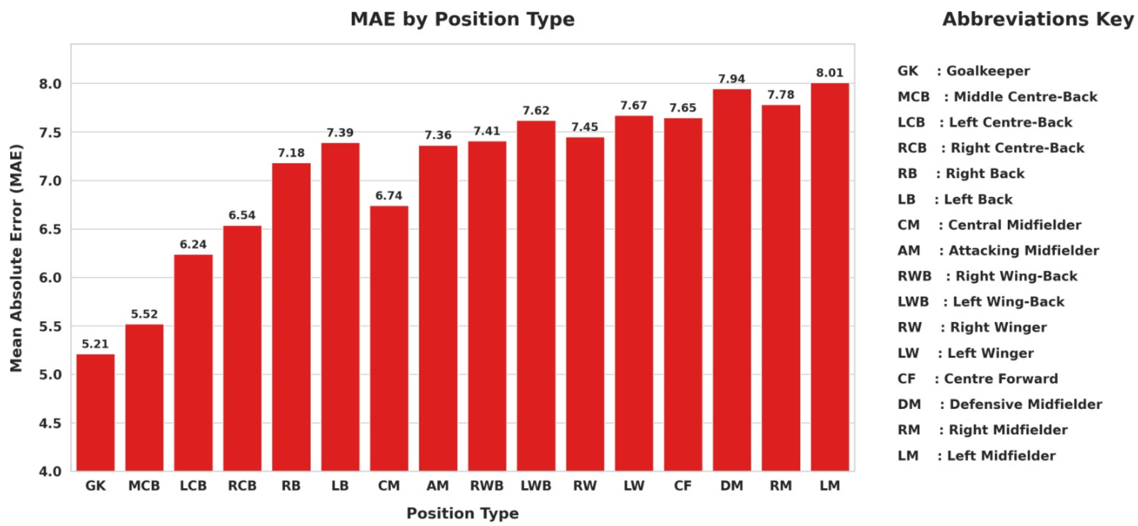


Figure 5. MAE for different players positional groups.

Fig. 6 reports the average prediction error for the in-possession and out-of-possession teams, both for the baseline model and for each of the five reserved teams. Overall, the results indicate consistently lower accuracy when predicting the positions of the possession team compared with the opposing team. The only exceptions are Argentina and Morocco, each reflecting well-known tactical particularities during the tournament. Morocco's compact, defensively oriented structure produced tightly clustered player positions with minimal variability, whereas Argentina's opponents showed unusually high positional variability due to their persistent efforts to contain Messi, who was the focal point of most defensive schemes.

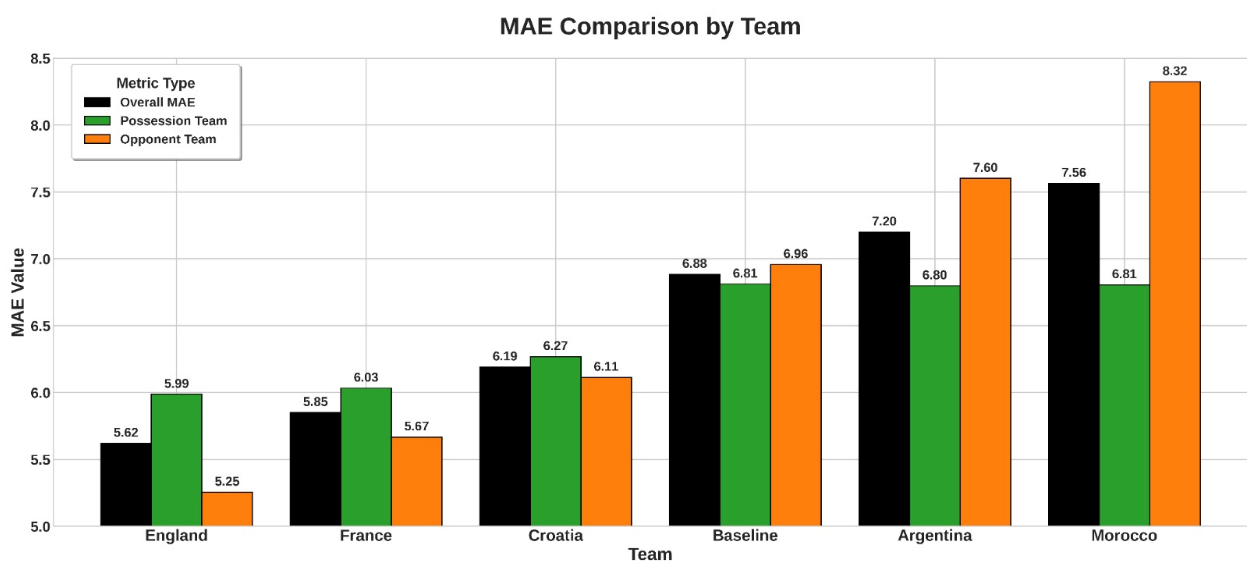


Figure 6. MAE for in-possession team vs. the opponent team.

Fig. 7 shows the average error across the different periods of the game demonstrating a slightly higher prediction error in Period 2, but a significantly higher error in the extra times, when extra time were played at the knockout matches, compared to the main two periods.

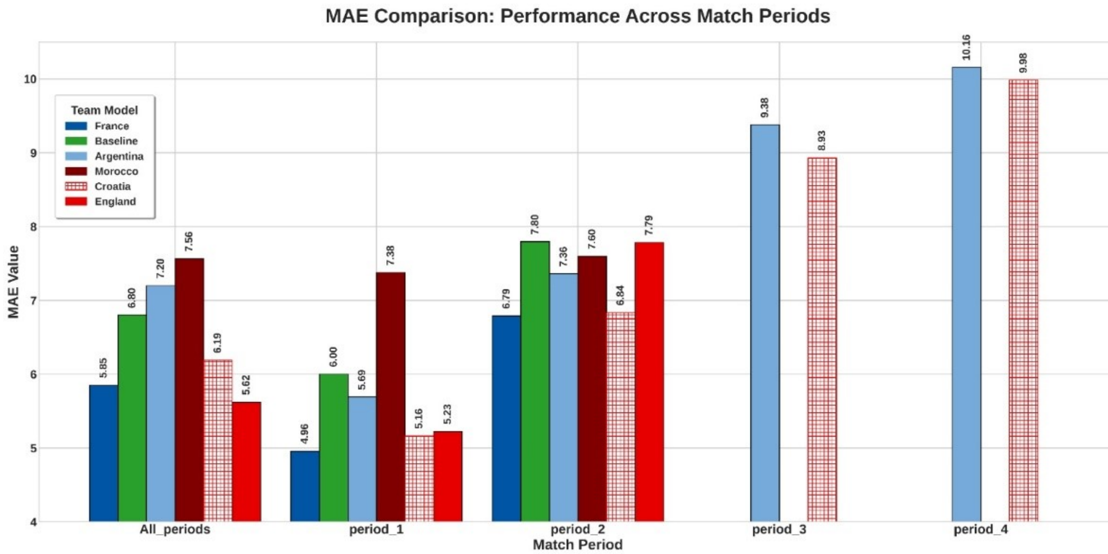


Figure 7. MAE for the four periods of the game

Fig. 8 illustrated the MAE for different pass features of the last pass before prediction as well as the pressure type of the team in-possession. While MAE ranges from 6.51m to 7.43m (~15% increase) for the different types of defensive pressure, the range increases dramatically for the different pass types (~87% increment) and different pass outcome (~43% increment, excluding the “out of play” outcome since it leads to sudden interruption of the formation).

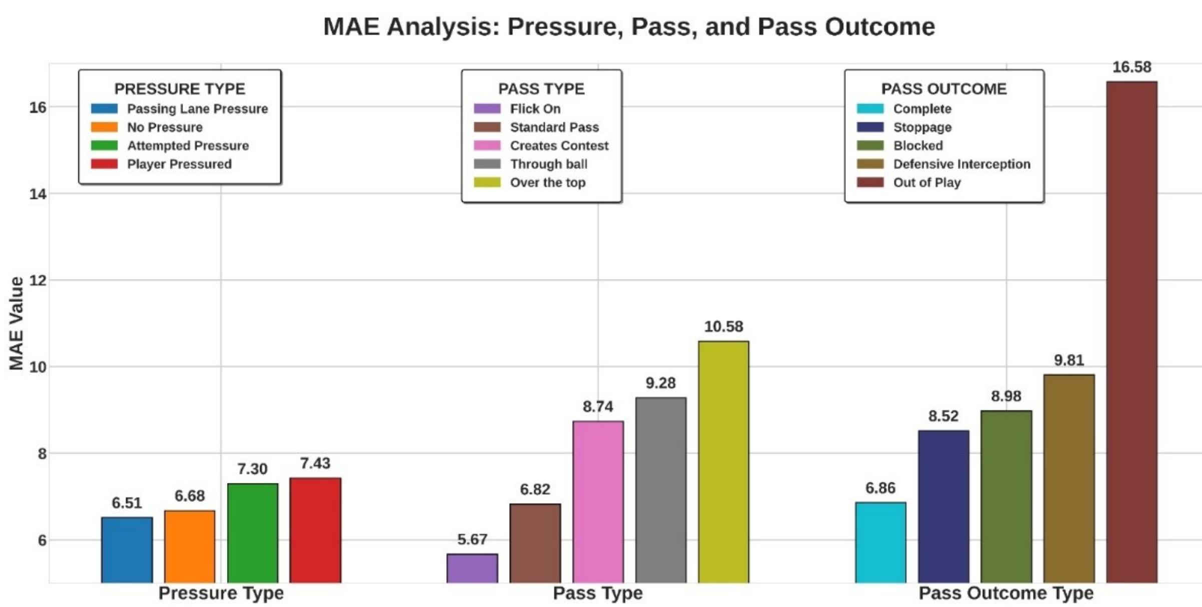


Figure 8. MAE for different pressure types, pass types, and pass outcomes.

Fig. 9 shows the MAE for different dynamics of the pass like the average distance moved by the players between events #4 and #5, and the players' dispersion across the field. The model shows better prediction with lower players' dispersion or movements directly before the pass.

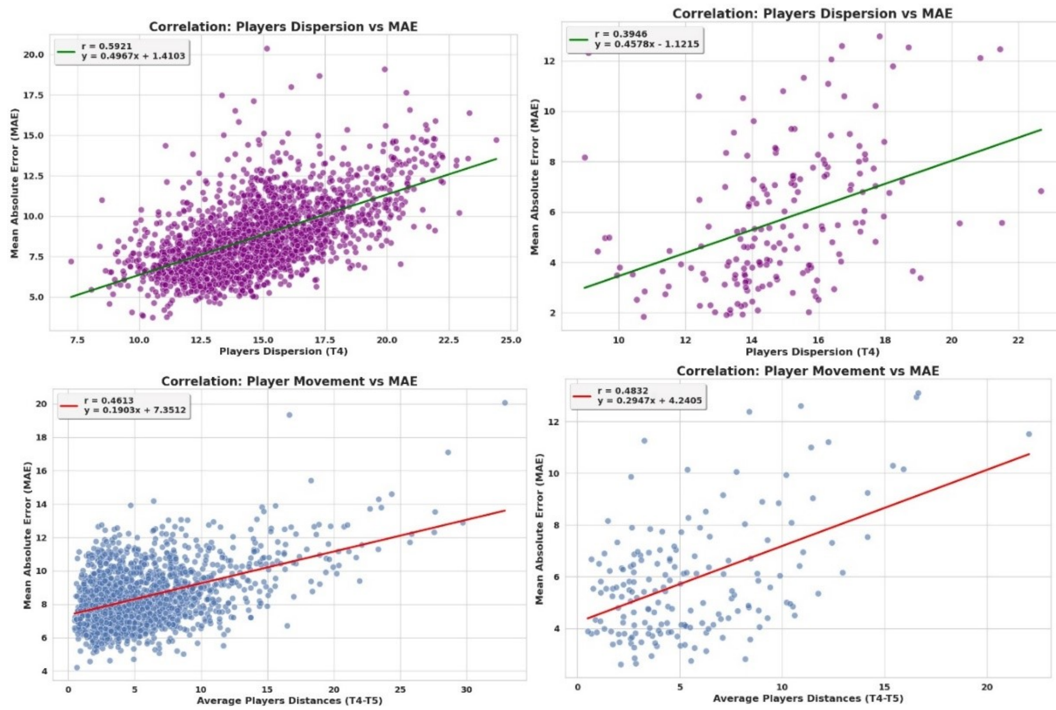


Figure 9. Correlation plots for MAE versus players dispersion (top-row) and average players movement (bottom-row) for team France. For each factor, results are obtained using the baseline models (left column) and France-finetuned models (right column). Note that different in points numbers due to the availability of all data point from France to be tested on the baseline but only 10% of them constitute the testing set for model finetuned on France data.

Figure 10 presents MAE heatmaps showing how prediction error varies across different areas of the pitch. Panel (a) displays the heatmap for the general baseline model, where predictions are more reliable in central areas and less accurate near the touchlines. Panels (b) and (c) show the corresponding maps for the fine-tuned models of France and Morocco, respectively. The Morocco heatmap exhibits a noticeable shift of the map regions toward the team goal line, reflecting the team's characteristically deep and compact defensive style.

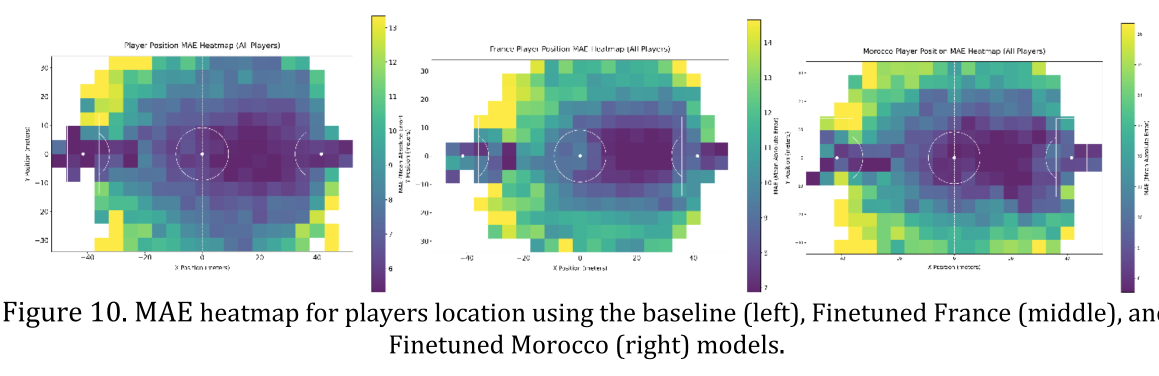


Figure 10. MAE heatmap for players location using the baseline (left), Finetuned France (middle), and Finetuned Morocco (right) models.

# 4. Discussion

In this work, we proposed a novel light-weighted soccer model that predicts complete post-pass formation, i.e. players' locations, of both teams given the players and ball locations in the previous five events to this pass. We first trained a baseline model on a diverse set of matches from the 2022 World Cup, then fine-tuned this model on selected reserved teams chosen to represent a broad range of tactical styles. We evaluated each model—baseline and fine-tuned—using a matrix-based setup that tests performance on its own test set (within-team) and on all other reserved teams plus the baseline data (cross-team). This dual evaluation quantifies both team-specific specialization and cross-team generalization, showing how team-specific tactical patterns help or hinder accuracy across different teams. Our models achieve strong accuracy (MAE = 6.8 m for the baseline and down to 5.6 m for the fine-tuned versions), comparable to performance reported in analogous tasks (6.8m in Graph Imputer [13], and 6.9m in Agent Imputer [14]).

Our results yield several notable insights. First, MAE values for in-possession and out-of-possession teams consistently diverge, likely reflecting the fundamentally different tactical behaviors of each role. Possession teams generally maintain compact, coordinated structures to facilitate ball progression, whereas opponent teams adopt more dispersed defensive shapes and apply pressing pressure—leading to systematically different prediction challenges for each side. Second, we observe an increase in prediction error across match periods, suggesting that the model may be sensitive to evolving match-state dynamics. Factors such as player fatigue, tactical adjustments, scoreline-driven formation changes, or increased late-game pressure likely contribute to this pattern. Although the models are trained using samples spanning all periods, the more condensed testing samples from the later stages of the match still exhibit higher errors compared to earlier phases.

Our results demonstrate that the proposed model's performance is strongly influenced by several key factors. First, *player movement distance*—the spatial displacement between consecutive timesteps—plays a central role in prediction difficulty. High-movement sequences often correspond to dynamic tactical moments such as counter-attacks, pressing actions, or transitions, whereas low-movement sequences typically reflect stable possession phases and are therefore easier to predict. Second, *team spatial dispersion*—the degree to which players are spread across the pitch—captures an essential tactical dimension. Low-dispersion states (compactness) commonly arise during defensive blocks or coordinated pressing, while high-dispersion states occur in expansive possession structures that exploit width or during defensive recovery. Prediction is less reliable when team spatial dispersion is high (expansive possession structures or defensive recovery), compared with the more stable and compact configurations observed in controlled possession phases. Finally, *prediction accuracy varies across pitch regions and player roles,* reflecting inherent differences in positional responsibilities and movement behaviors. Defensive players tend to operate within more constrained spatial zones, making their locations more predictable, while attacking players exhibit greater variability and freedom in movement. Similarly, predicting the goalkeeper or backline positions is structurally different from predicting midfield or forward locations due to their distinct tactical functions.

Our model complements the prior work *Seq2Event* [5] in a meaningful way. While *Seq2Event* predicts the optimal next action (e.g., a pass) based on the current formation, our model takes that predicted action together with the formations from the previous five events to forecast the resulting post-action formation. This naturally suggests integrating both models to address a very useful question: *"Given* *the players' current locations, what is the best action to take, and where will all players be positioned immediately afterward?".* Such an integrated setup would enable an automatic, step-by-step simulation of match progression by repeatedly answering this question. This capability is valuable not only for tactical analysis and counterfactual exploration but also for the gaming industry, where it can support dynamic in-game behavior that is not restricted to predefined scenarios.

One limitation of this work is the use of a single prediction model for both teams. Our results indicate that model performance consistently differs between the in-possession team and the opponent team across both the baseline and fine-tuned settings. This suggests that developing separate, role-specific models for each team may be more appropriate. We acknowledge this limitation but defer its investigation to future work.

# 5. Conclusion

Pass2Formation is the first practical, identity-anchored pipeline to forecast full next-state spatial configurations and systematically generate multi-receiver counterfactuals from open-play contexts. Its compact architecture and strong positional accuracy make it immediately deployable for team-shape analysis, tactical decision support, and what-if scenario ranking—bridging the gap between value-only models and heavyweight simulators and benefitting both the professional analytics community and game/simulators developers.

# References


[1] J. Fernández, L. Bornn, and D. Cervone, “A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions,” *Mach. Learn.*, vol. 110, no. 6, pp. 1389–1427, 2021.

[2] T. Decroos, L. Bransen, J. Van Haaren, and J. Davis, “Actions Speak Louder than Goals: Valuing Player Actions in Soccer,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1851–1861.

[3] W. Spearman, “Beyond expected goals,” in *Proceedings of the 12th MIT sloan sports analytics conference*, 2018, pp. 1–17.

[4] Hudl StatsBomb, “Hudl StatsBomb Release New Models: Expected Pass (xPass) nda Pass Clustering.” 2023.

[5] I. Simpson, R. J. Beal, D. Locke, and T. J. Norman, “Seq2Event: Learning the Language of Soccer Using Transformer-based Match Event Prediction,” in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3898–3908.

[6] Y. Honda, R. Kawakami, R. Yoshihashi, K. Kato, and T. Naemura, “Pass Receiver Prediction in Soccer using Video and Players’ Trajectories,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022, pp. 3502–3511.

[7] Z. Wang *et al.*, “TacticAI: an AI assistant for football tactics,” *Nat. Commun.*, vol. 15, no. 1, p. 1906, 2024.

[8] M. A. Alcorn and A. Nguyen, “baller2vec: A Multi-Entity Transformer For Multi-Agent Spatiotemporal Modeling,” *arXiv Prepr. arXiv2102.03291*, 2021.

[9] M. A. Alcorn and A. Nguyen, “baller2vec++: A Look-Ahead Multi-Entity Transformer For Modeling Coordinated Agents.” 2021.

[10] G. Capellera, L. Ferraz, A. Rubio, A. Agudo, and F. Moreno-Noguer, “FootBots: A Transformer-based Architecture for Motion Prediction in Soccer.” 2024.

[11] G. Capellera, L. Ferraz, A. Rubio, A. Agudo, and F. Moreno-Noguer, “TranSPORTmer: A Holistic Approach to Trajectory Understanding in Multi-Agent Sports.” 2024.

[12] L. Thorpe, L. Bawden, K. Vendal, J. Bronskill, and R. Turner, “SportsNGEN: Sustained Generation of Realistic Multi-Player Sports Gameplay,” in *Proceedings of the 12th International Conference on Sport Sciences Research and Technology Support*, 2024, pp. 119–130.

[13] S. Omidshafiei *et al.*, “Multiagent off-screen behavior prediction in football,” *Sci. Rep.*, vol. 12, no. 1, p. 8638, 2022.

[14] G. Everett, R. J. Beal, T. Matthews, J. Early, T. J. Norman, and S. D. Ramchurn, “Inferring Player Location in Sports Matches: Multi-Agent Spatial Imputation from Limited Observations.” 2023.