CounterEval: A Deep Learning Framework for Performance Evaluation in Soccer Counterattacks Using Tracking Data

Yixiao Chen
Department of Statistics and Data Science
Yale University
New Haven, CT, USA 06511
yixiao.chen@yale.edu

Yuanhao You
Department of Statistics and Data Science
Yale University
New Haven, CT, USA 06511
yuanhao.you@yale.edu

Qifan Zhang
Department of Statistics and Data Science
Yale University
New Haven, CT, USA 06511
qifan.zhang@yale.edu

December 16, 2024

Abstract

This study introduces *CounterEval*, a deep learning framework for evaluating player performance in counterattacks using tracking data from 632 games across MLS (2022), NWSL (2022), and international women's soccer (2020 to 2022). The dataset includes detailed information on player and ball locations, velocity, and acceleration, and was refined through our quality improvements, such as correcting mislabeled counterattacks and filling missing player IDs. The core of the framework comprises two neural network models: a movement model, a variational autoencoder model fitting the distribution of a player's next location given the spatial positions of all players, and a graph-based counterattacking success model, which estimates the probability of a successful counterattack given the current game context. By combining the two models, we develop a real-time evaluation metric that quantifies player performance during counterattacks. The framework is robust and scalable, with potential for further enhancement through higher-quality data and model fine-tuning.

1 Executive Summary

In modern soccer, counterattacking strategies are pivotal to scoring opportunities, yet quantifying individual player contributions within these fast-paced scenarios remains challenging. To address this, we introduce *CounterEval*, a deep learning framework that evaluates player performance during counterattacks by leveraging tracking data from 632 professional games, including MLS (2022), NWSL (2022), and international women's soccer (2020–2022).

The CounterEval framework integrates two core models:

- A **movement prediction model**, implemented as a Transformer-enhanced Variational Autoencoder (VAE), predicts players' next plausible locations under similar circumstances.
- A **tactical success model**, built using Graph Attention Networks (GATs), estimates the probability of counter-attack success based on the current game context.

By comparing a player's actual movement to the predicted average movements of other players in similar scenarios, CounterEval generates **performance metrics** that quantify each player's influence on the success of the counterattack.

Key Results and Insights

1. Movement Prediction: The movement prediction model successfully captures realistic player movements, as shown in Figure 10. By predicting alternative positions for players, this model enables counterfactual comparisons to evaluate decision-making quality.

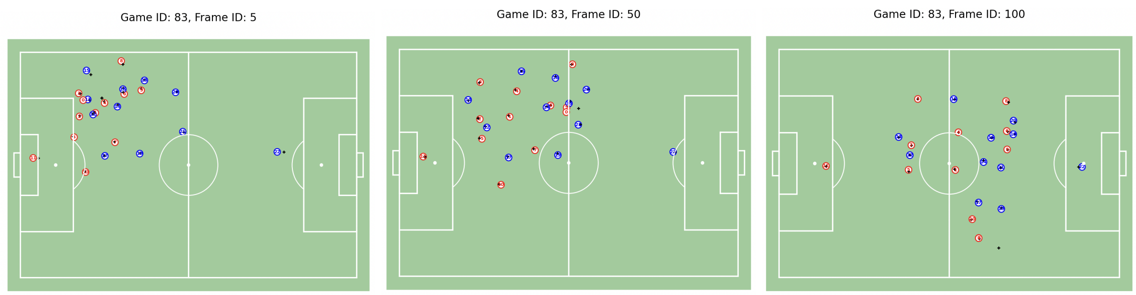

Figure 1: Movement prediction samples: Predicted player locations (black) closely align with actual observed movements.

2. Tactical Analysis: The GAT-based tactical success model achieved a testing AUC of 0.62. While there is room for improvement, it provides meaningful tactical insights. For instance, Figure 12 shows how counterattack success probability evolves as the play progresses:

- Initially, the offense team struggles with no numerical advantage, leading to a low success probability.
- As the ball advances into the attacking third, the success probability rises, reflecting better positioning.
- Near the penalty area, the probability fluctuates, driven by critical decision points such as dribbling or passing attempts.

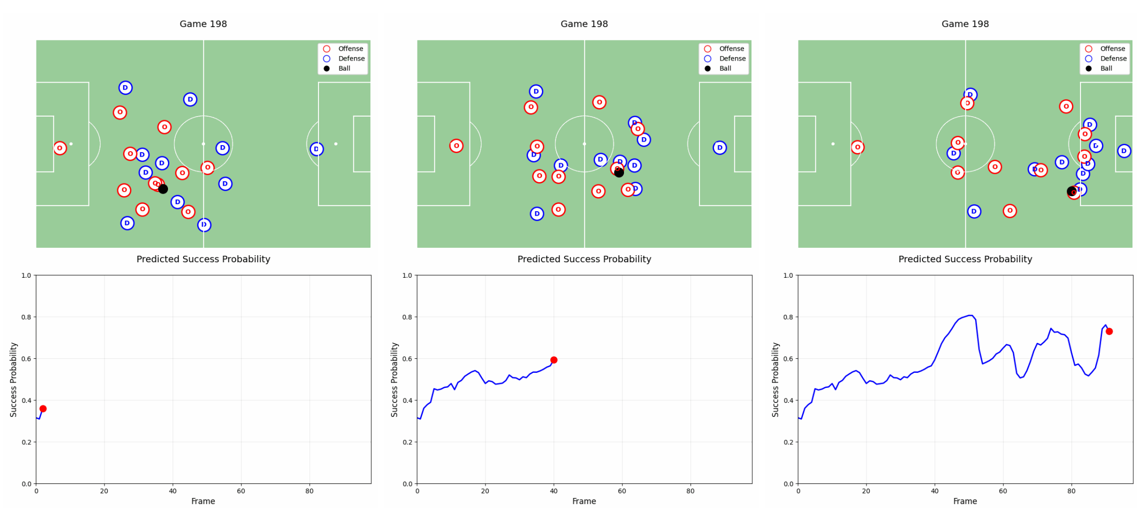

Figure 2: Tactical analysis of counterattack success probabilities over time.

3. Player Performance Metrics: Using the movement prediction and tactical success models, we compute player-specific performance scores that evaluate their contributions to counterattack success. Figure 14 illustrates the performance metrics for players in a specific game:

- Offensive players 1, 8, and 9 exhibit high scores due to effective off-ball movements and dribbling actions.
- Defensive players 17 and 20 receive positive scores for effectively marking key offensive players.
- Player 14, on the other hand, underperforms due to challenges in man-marking.

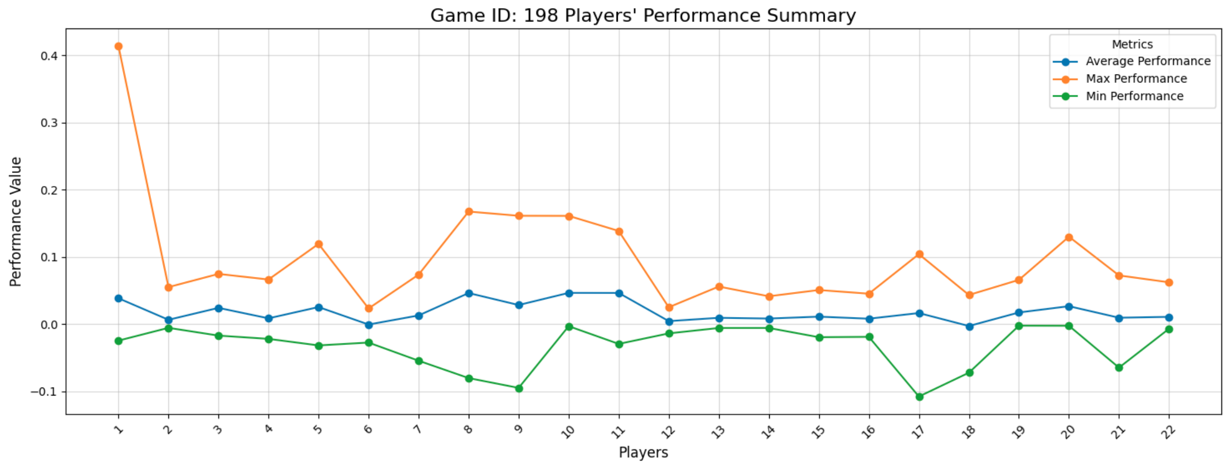

Figure 3: Player performance scores during a counterattack event.

Future Directions

The CounterEval framework provides a robust foundation for analyzing player performance in dynamic counterattack scenarios. However, several avenues for improvement and further research remain:

- **Counterfactual Paths:** Instead of evaluating single-step movements, incorporating entire movement paths could provide deeper insights into players' decision-making processes.
- **Time-Resolved Data:** Incorporating fixed time intervals would improve the accuracy of velocity and acceleration calculations.
- **Refined Event Definition:** A more rigorous and data-driven definition of a counterattack event would improve model consistency and accuracy.
- **Alternative Baselines:** Expanding beyond "average players" to include expert-defined baselines could enhance the evaluation framework.

Conclusion

CounterEval offers a scalable and interpretable framework for evaluating player performance during counterattacks. By combining movement prediction with tactical success analysis, it identifies key players and actions that influence success. While data quality and volume present challenges, the framework demonstrates promising results and offers actionable insights for coaches, analysts, and researchers in sports analytics. Future work will focus on refining the model and exploring its applications in other phases of play.

2 Introduction

Counterattacking is an effective strategy for scoring in modern football, championed by renowned coaches like José Mourinho, Diego Simeone, and Antonio conte. A successful counterattack is often executed by key players, making it valuable to develop metrics that evaluate individual contributions to the play. The analysis of soccer player movements and team tactics requires sophisticated computational approaches that can capture both individual dynamics and collective behavior. Our work addresses this challenge through a two-stage framework: probabilistic movement modeling through generative deep learning and tactical analysis through graph-based approaches. The integration of these components enables the construction of performance metrics scores that reflect each player's impact on counterattack success.

The remainder of this paper follows a structured analysis of our framework and its applications. Section 2 contains data exploration and visualization, including a description of the source of dataset and data cleaning processes.In Section 3, two predictive models, a transformer-enhanced conditional VAE for forecasting player movements, and Graph Attention Networks architectures for tactical analysis, and the methodology of computing performance metrics are presented. Section 4 then visualizes and interprets the resulting model outputs, including an in-depth examination of the CounterEval Contribution Scores to identify key player actions influencing counterattack success. Finally, Section 5 presents conclusions and recommendations for practical applications in player evaluation and tactical analysis, along with promising directions for future research in dynamic game situation analysis.

3 Data exploration and visualization

In this project, we focus on using MLS(2022) dataset from [3], that contains tracking data from 632 games across MLS (2022), NWSL (2022), and international women's soccer (2020 to 2022). Each frame of tracking data in the dataset is represented as a heterogeneous graph. Nodes in the graph correspond to individual players (offensive and defensive), while edges capture spatial and temporal relationships between players. A detailed summary of the node and edge features is presented in Table 1.

Feature Type	**Description**
	Node Features
x Coordinate	x-coordinate on the 2D pitch for the player/ball.
y Coordinate	y-coordinate on the 2D pitch for the player/ball.
vx	x-component of the velocity vector.
vy	y-component of the velocity vector.
Velocity	Magnitude of the velocity vector.
Velocity Angle	Angle of the velocity vector.
Distance to Goal	Distance of the player from the goal post.
Angle with Goal	Angle made by the player with the goal.
Distance to Ball	Distance of the player from the ball (always 0 for the ball).
Angle with Ball	Angle made by the player with the ball (always 0 for the ball).
Attacking Team Flag	1 if the team is attacking; 0 otherwise (1 for the ball as well).
Potential Receiver	1 if the player is a potential receiver; 0 otherwise.
	Edge Features
Player Distance	Distance between two connected players.
Speed Difference	Speed difference between two connected players.
Positional Sine Angle	Sine of the angle between the positional vectors of two connected players.
Positional Cosine Angle	Cosine of the angle between the positional vectors of two connected players.
Velocity Sine Angle	Sine of the angle between the velocity vectors of two connected players.
Velocity Cosine Angle	Cosine of the angle between the velocity vectors of two connected players.

Table 1: Summary of Node and Edge Features in the Dataset

To identify and label counterattacks in the dataset, [3] apply a systematic labeling algorithm based on on-ball event sequences. The process involves two stages: identifying counterattacks and determining whether they are successful.

A sequence of events is classified as a counterattack if it satisfies the following criteria:

- The starting point of the sequence occurs in the **defensive half** of the pitch.
- The sequence of events does not include a set piece (e.g., freekick, corner, throw-in).
- The ball moves forward by at least **10 meters**.
- The ball travels with a forward velocity of at least **4 meters per second (m/s)**.

A counterattack is labeled as **successful** if the following conditions are met:

- The ball and the player enters the **18-yard penalty box**.

- The last event in the sequence occurs within the attack-third.

However, the data quality of this dataset was poor, with over 50% of outcomes mislabeled, unrealistic movements of players and the ball, and incorrect attacking team flags. To address these issues, we manually checked and corrected the errors using a custom visualization tool built with Streamlit. This process involved relabeling the outcomes and attacking flags to ensure data accuracy. For those samples containing unrealistic movements of players and the ball, we directly delete the samples. Figure 4 shows a sample of good quality. Figure 5 shows a sample when the ball is stopping in the same location for too long (in the middle of the player 10 and the player 16). Figure 6 shows an example when the label is incorrect since the player 10 has dribbled the ball into the penalty area.

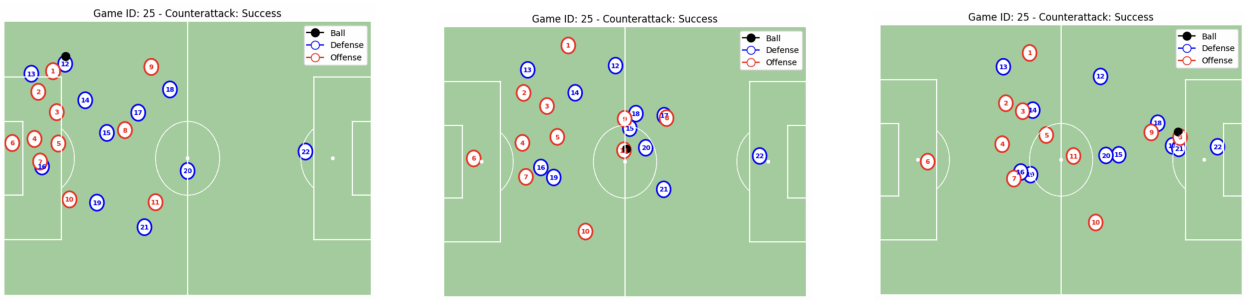

Figure 4: A sample of good quality

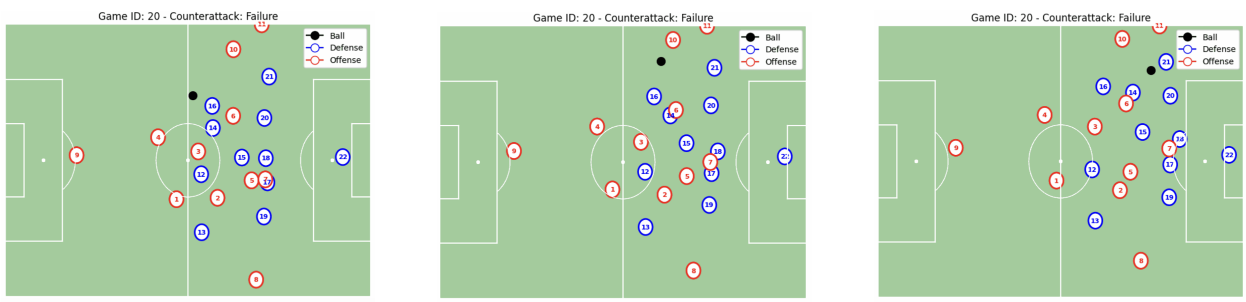

Figure 5: A sample when the ball is stopping in a location too long

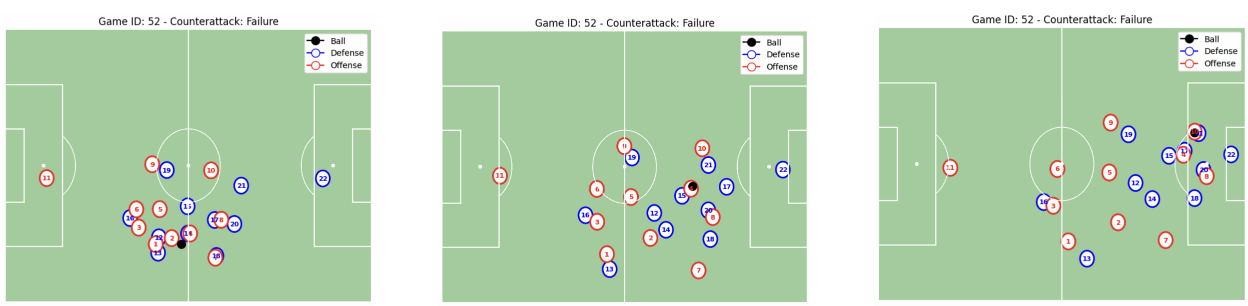

Figure 6: A sample when the label is incorrect

Additionally, instead of using the heterogeneous graph approach proposed in [3], which incorporates specific node and edge features to represent connections between players and the ball, we adopted a fully connected homogeneous graph structure. This design choice enhances the effectiveness of the attention mechanism in our graph models. Furthermore, we excluded edge features, as they are derived directly from node features and do not contribute additional unique information. For a detailed explanation of the graph construction process, refer to Section 4.2.

With these data cleaning and preparation steps complete, we proceed with the dataset, accompanied by summary statistics presented in Table 2 and illustrative graph samples 7.

Statistic	Value
Number of Graphs	78,583
Number of Nodes in Each Graph	23
Number of Edges in Each Graph	253
Number of Edge Features	19
Number of Node Labeled Classes	2

Table 2: Summary Statistics of the Dataset and Graph Characteristics

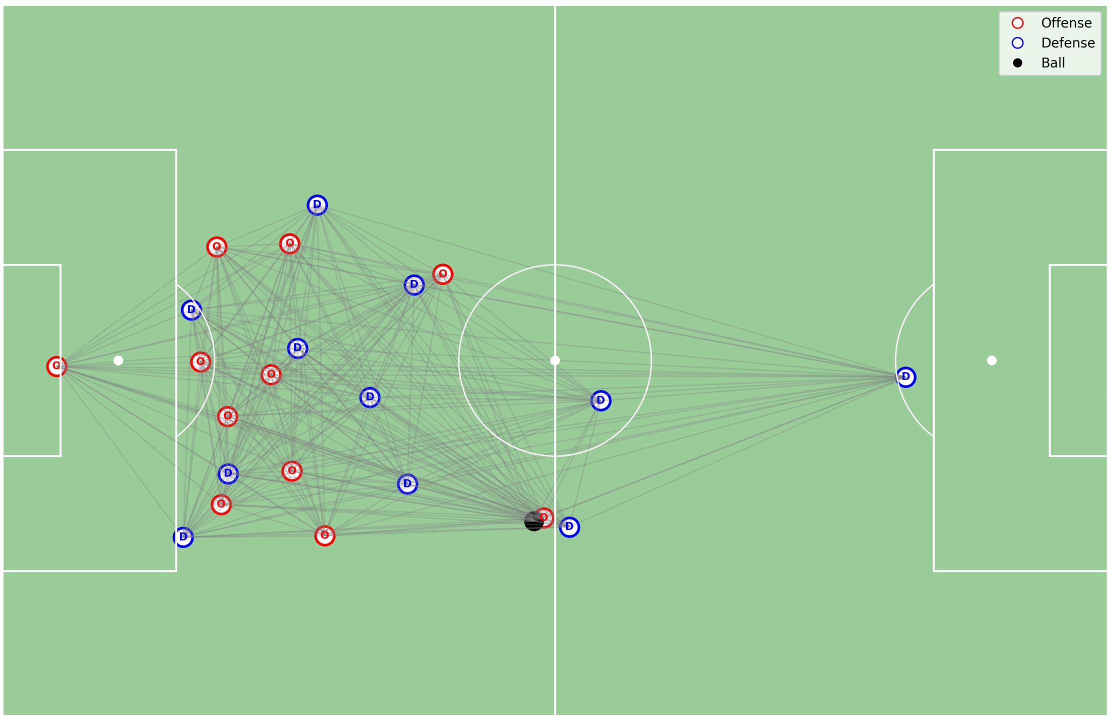

Figure 7: Sample Graph

4 Methodology

CounterEval is a framework for real-time evaluation of player performance during counterattacks. We begin by presenting a general method for calculating performance metrics applicable to various scenarios. We define a tick-level metric m that assesses how well a player's action a performs compared to the actions a' of other players, sampled from a distribution π, at a specific time t:

$$m(p,t) = f(a|\theta(t)) - \mathbb{E}_{a' \sim \pi}[f(a'|\theta(t))],$$

where $\theta(t)$ denotes the context at time t and f is a scoring function. We then identify the time interval $[t_1, t_2]$ for the counterattack and integrate the metrics m over this period to evaluate the player's performance M:

$$M(p) = \int_{t_1}^{t_2} m(p,t)\rho(t)dt,$$

with ρ as a time-based importance function.

To model M, we must consider a, f, π, θ, and ρ. In soccer, action a includes various elements such as player movement, passing types, and communication. For this study, we focus solely on players' movement direction, as it is more easily observable. Players' movements can be represented by a series of two-dimensional locations (x, y) included in our tracking datasets. Thus we model action a as the next location (x, y) of a player given the context $\theta(t)$.

We model score function f as the predicted probability of a successful counterattack given the game status through a deep learning model. Tracking data in sports is a complex form of spatiotemporal data. While physics-based models are interpretable and useful, they often require extensive parameter fitting and can lack accuracy [2, 4]. Since this study aims for both accuracy and generality, the data-driven, assumption-free deep learning models are more suitable in the study given their effectiveness shown in other complex scenarios. Similar to the score function f, we model the distribution of other players' action π using a generative deep learning model. Instead of using the uniform distribution for the next location as in [3], we propose learning the distribution of the next locations for average players. Comparing individual players to the average in similar scenarios is crucial, as most are well-trained and avoid random actions. However, highly skilled and experienced players can outperform others, raising the probability of scoring.

The context $\theta(t)$ is represented by features in the tracking datasets, including players' location, velocity, and acceleration. For simplicity, we represent ρ as a constant, although we believe that the beginning and end of a counterattack event are more significant than the middle.

Based on the discussion above, we now present the detailed implementation of the CounterEval framework. The score function is implemented by a graph-based neural network f_s and the distribution π, which we also name as movement prediction model, is implemented by a variational autoencoder model f_m. At time t, we observe the actual location of a player $l_t = (x_t, y_t)$. Next, we collect the historical context data $\{\theta(t-\delta)\}_{\delta=1}^{W}$, where W is the window size, and obtain the predicted movement made by other players, $f_m(\{\theta(t-\delta)\}_{\delta=1}^{W})$. We then sample from $f_m(\{\theta(t-\delta)\}_{\delta=1}^{W})$ to get a set of possible location $\{l'_{i,t}\}_{i=1}^{N}$ made by other players in the same scenario. The tick-level performance metric m is calculated as

$$m(p,t) = f_s(l_t|\theta(t)) - \frac{1}{N}\sum_{i=1}^{N} f_s(l'_{i,t}|\theta(t)).$$

The player's performance in the event is then summarized as

$$M(p) = \frac{1}{T}\sum_{i=1}^{T} m(p, t_i).$$

4.1 Movement Prediction Model

The movement prediction model models the distribution π of other players' action as mentioned above. In this study, we use a Variational Autoencoder (VAE) model to fit the distribution. We set the window size $W = 1$ for simplicity. The inputs of the encoder are the players' actual location and the context $\theta(t-1)$ and the outputs are two-dimensional latent variable μ and $\log \sigma$. Then we draw a sample z from $\mathcal{N}(\mu, \sigma^2)$, where we also apply the reparametrization trick. The input of the decoder are z and the context $\theta(t-1)$ and the output is the predicted location of the player at time t. Since the context variable θ also includes information of other players in the game, we use a Transformer model to generate the embedding of the context variables for the downstream VAE. Figure 8 shows the architecture of the movement prediction model. For other details of VAE and Transformer, please refer to [5, 1].

4.2 Tactical Analysis with Graph Neural Networks

This section outlines the tactical analysis component of the CounterEval framework, which employs Graph Attention Networks (GATs) to model f_s as the probability of successfully counterattack. The game context at any given time t is represented as a graph $G = (V, E)$ similar to Figure 7, where:

- **Nodes (V):** Each node $v \in V$ represents a player, with features X_v capturing their positional and kinematic information:

  $$X_v = [x, y, v_x, v_y, d_{\text{goal}}, \theta_{\text{goal}}, d_{\text{ball}}, \theta_{\text{ball}}, \text{team\_flag}, \text{potential\_receiver}]$$

  Here:
  - (x, y): Player's position on the field.
  - (v_x, v_y): Player's velocity components.
  - $d_{\text{goal}}, \theta_{\text{goal}}$: Distance and angle to the opponent's goal.
  - $d_{\text{ball}}, \theta_{\text{ball}}$: Distance and angle to the ball.

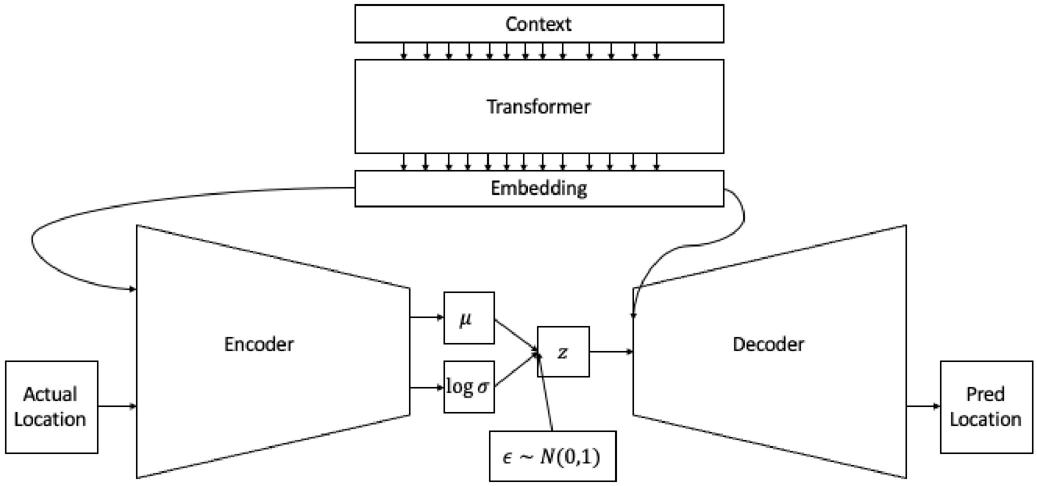

Figure 8: Architecture of the Movement Prediction Model

  - team_flag: Binary indicator of the player's team (attacking or defending).
  - potential_receiver: Binary indicator of pass viability.
- **Edges (E):** Connections are established between all player pairs and between each player and the ball, capturing spatial and tactical relationships.

Figure 9 shows the graph model architecture. GATs uses attention mechanisms to learn the importance of neighboring nodes to generate node embeddings. For more details of GAT, please refer to [6]. The output embeddings from GATs are passed to Multi-Layer Perceptrons (MLPs), which further refine the embeddings through deep learning. Finally, the model outputs a probability score, representing the likelihood of counterattack success.

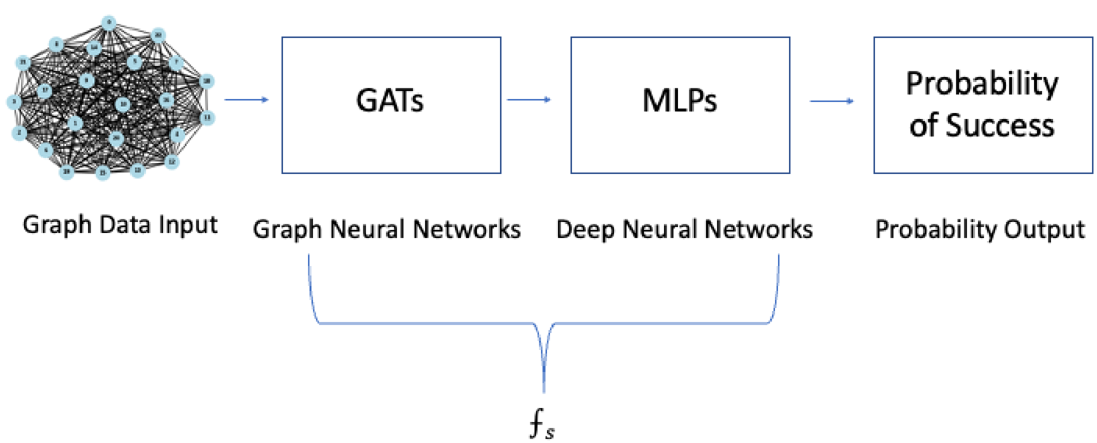

Figure 9: Architecture of the Graph Model

4.3 Performance Metrics Methodology

The central objective is to quantify how a single player's chosen movement at any given time influences the probability of a successful counterattack. To achieve this, we consider each player's actual movement decision—represented by their next observed on-field location—and compare it against a set of plausible alternative locations sampled from the learned movement model. By holding all other players' positions fixed at their actual observed locations, this procedure isolates the effect of the focal player's movement on the ensuing play.

Step-by-Step Procedure:

1. **Extract Actual Situation:** At a specific time tick t during a counterattack, let $\theta(t)$ denote the contextual game state, including the coordinates, velocities, and other relevant features of all players and the ball.
2. **Evaluate Actual Action:** Let $l_t = (x_t, y_t)$ be the observed next location of a particular player at time t. Using the tactical success model f_s, we compute the probability of a successful counterattack given this actual location:
$$P_{\text{actual}} = f_s(l_t \mid \theta(t)).$$
3. **Generate Counterfactual Locations:** To understand alternative decision outcomes, we rely on the trained movement model π. From π, we sample N plausible alternative next locations $\{l'_{i,t}\}_{i=1}^{N}$ for the player. Each sample $(x'_{i,t}, y'_{i,t})$ is a realistic counterfactual that respects the learned spatiotemporal distribution of player movements.
4. **Compute Counterfactual Success Probabilities:** For each counterfactual location $l'_{i,t}$, we hold all other players' locations fixed as observed. The only modification is the focal player's position. We then use the tactical success model to evaluate the success probability:
$$P_{\text{cf},i} = f_s(l'_{i,t} \mid \theta(t)), \quad i = 1, 2, \ldots, N.$$
5. **Aggregate Counterfactual Estimates:** We aggregate the counterfactual success probabilities to form an expected counterfactual outcome:
$$\overline{P_{\text{cf}}} = \frac{1}{N} \sum_{i=1}^{N} P_{\text{cf},i}.$$
6. **Compute Performance Score:** The instantaneous performance metric at time t is defined as:
$$m(p, t) = P_{\text{actual}} - \overline{P_{\text{cf}}}.$$
A positive $m(p, t)$ indicates that the player's actual decision improved the likelihood of a successful counterattack compared to typical plausible alternatives, while a negative $m(p, t)$ suggests the opposite.
7. **Event-Level Score:** Summarizing over the entire counterattack event $[t_1, t_2]$, we define:
$$M(p) = \frac{1}{T} \sum_{t=t_1}^{t_2} m(p, t),$$
where $T = t_2 - t_1$ is the number of time steps during the counterattack event. This provides a holistic measure of the player's overall contribution during the entire counterattack.

This methodology yields a player-specific metric that directly measures the marginal impact of the player's observed decision on the probability of a successful counterattack. By comparing the actual action to data-driven counterfactual scenarios while holding other players' positions fixed, this metric isolates and quantifies the player's influence in dynamic in-game situations.

5 Experimental Results & Analysis

5.1 Movement Prediction

We train the conditional VAE model on the MLS dataset, achieving a testing mean squared error of approximately 0.0017 (with standardized labels). Figure 10 shows a sample result from the movement prediction model, where the black dot represents a set of VAE samples reflecting the closest player's predicted location. We find that the predicted location is quite close to the players' actual location. This poses challenges for calculating performance metrics, as we anticipate notable differences in player movement during successful counterattacks compared to others. Consequently, we aim to analyze players' multi-step paths rather than focusing on single steps, as the differences in one step are minimal. This will guide our future exploration. Figure 11 illustrates a close-up of player 8's predicted location over three frames.

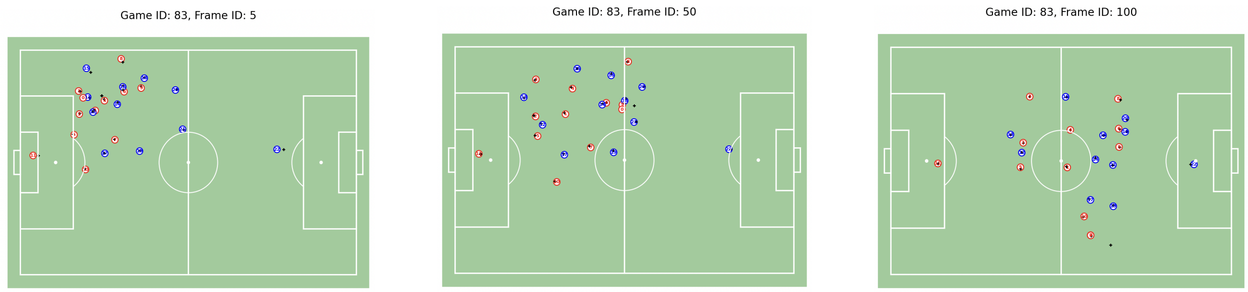

Figure 10: Movement prediction

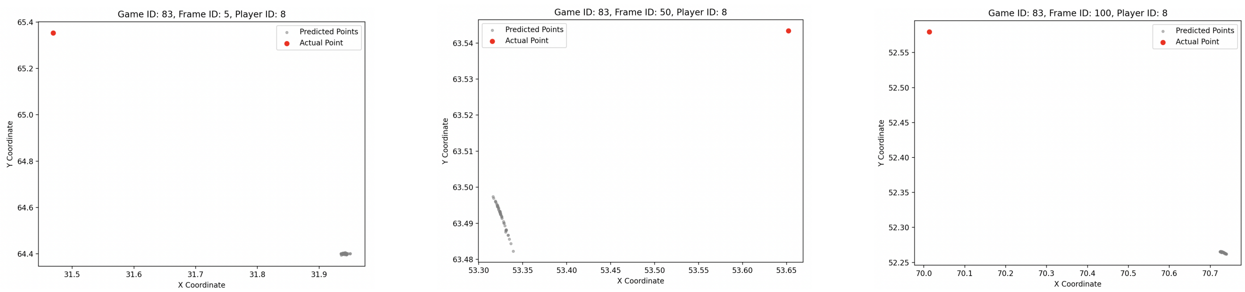

Figure 11: Movement prediction for the Player 8

5.2 Tactical Analysis Visualization

We train the GAT model on the MLS dataset and achieve a testing AUC of 0.62. While this presents a challenge in terms of predictive accuracy, the model still provides a solid foundation for conducting tactical analysis. We believe that improvements in data quality and an increase in the number of data points could further enhance the model's performance in future studies. Despite the current limitations of the model, these analyses demonstrate its ability to provide valuable insights into tactical scenarios, allowing for a detailed examination of player movements and decision-making during counterattacks. Figure 12 presents a sample tactical analysis. The top graph illustrates the current situation on the field, while the bottom plot depicts the corresponding success probability over time.

In the first frame, we observe that at the start of the counterattack, the offensive team holds the ball near midfield. They lack a numerical advantage in the attacking third, and two defensive players are closing in on the ball carrier. This scenario corresponds to a low counterattack success probability, as the offensive team faces significant defensive pressure.

In the second frame, the offensive team successfully advances the ball into the attacking third, though they still lack a numerical advantage in this area. Moreover, there are limited potential receivers available in advantageous positions. However, since the ball has moved closer to the goal, the success probability slightly increases, reflecting the team's improved positioning.

Finally, in the third frame, the offensive team pushes closer to the penalty area. Despite the proximity to the goal, the team faces significant challenges: there are limited options for passing or crossing, and the defensive players are well-positioned to neutralize the attack. At this point, the counterattack success probability fluctuates. It depends heavily on the actions of the ball carrier, such as the ability to execute a precise pass or successfully dribble through defenders. These decisions directly influence the likelihood of a successful counterattack, as reflected in the higher but volatile success probability.

5.3 Performance Metrics Visualization

The CounterEval performance evaluation systematically assesses players' actions during counterattacks by comparing them to average actions in analogous scenarios. This section provides a case study on game 198, as illustrated in three screenshots in Figure 13. Within the offensive team, players 1, 8, and 9 demonstrate exemplary off-ball movement, with player 1 advancing into the penalty area, while player 2 exhibits minimal activity. On the defensive side, player 14 encounters challenges with man-marking, whereas players 17 and 20 effectively contain players 1 and 8. These observations are corroborated by Figure 14, which indicates that players 1, 8, and 9 achieve high performance scores

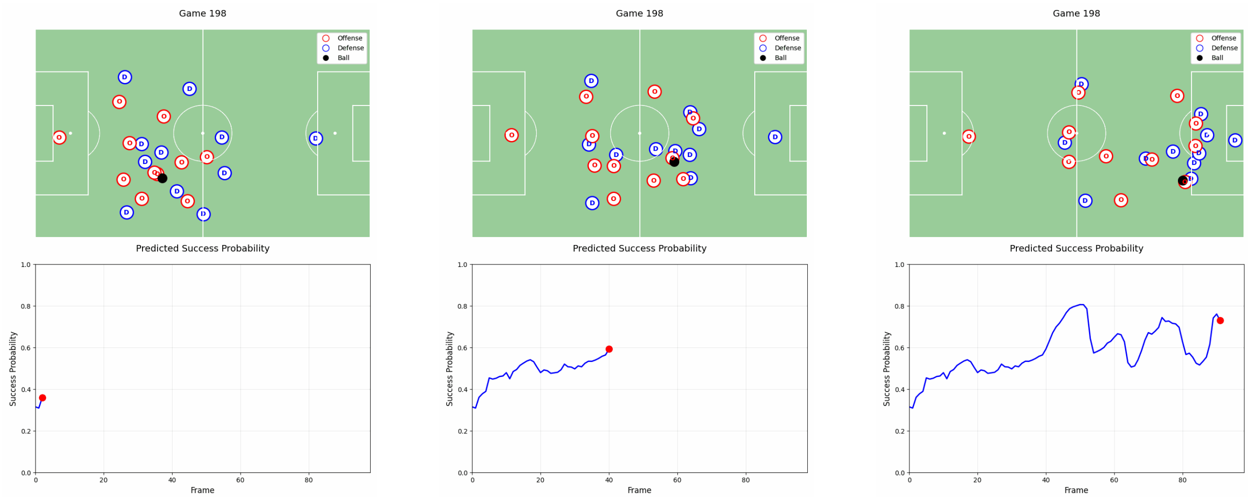

Figure 12: Tactical analysis for the probability of counterattack success

compared to their offensive counterparts. In contrast, player 14 receives a low score within the defensive team. The players 17 and 20 perform well defensively. Figure 15 presents a time series analysis of performance for players 1, 8, and 14, underscoring that player 1's score predominantly derives from his dribbling capabilities, whereas player 8's score reflects consistent off-ball running. Player 14's performance remains consistently subpar throughout the evaluation.

However, there are also some evaluation disobeying our intuition. In Figure 14, the goalkeeper player 11 also receives a high score within the offensive team but his actions clearly won't affect the counterattacks too much.

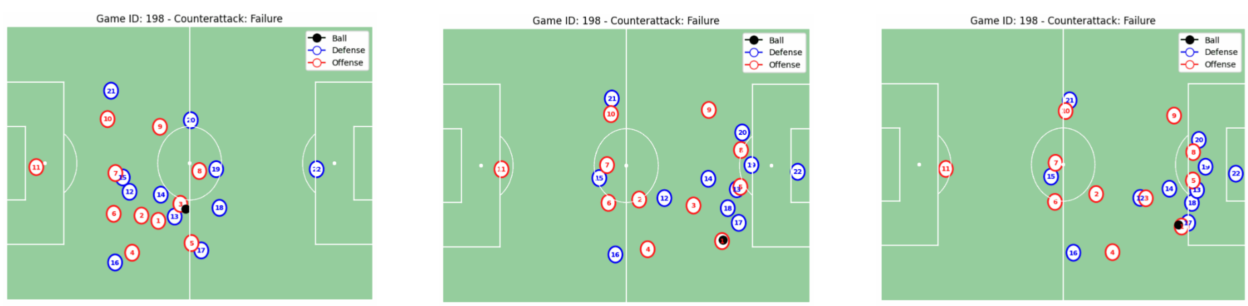

Figure 13: Game 198

6 Conclusions and recommendations

This project introduces the CounterEval framework to analyze player performance during counterattacks. By integrating a movement prediction model with a counterattack success prediction model, we can identify key contributors, such as off-ball runners and man-marking defenders. The theoretical framework of CounterEval, detailed in Section 4, is versatile and accommodates various deep learning interpretation algorithms, including permutation feature importance. In this study, we innovatively focus on the distribution of actions among average players, aiming to identify those who outperform the average rather than those who act randomly. We believe this average player model is superior to permutation feature importance. However, due to time constraints, we do not conduct a comprehensive comparison of the two methods.

There are several areas for future improvement. First, we should implement a quantitative evaluation metric instead of relying on manual visual checks, which may require a labeled dataset with key players identified by experts. Second, our current datasets lack a time attribute, potentially affecting the accuracy of players' velocity and acceleration calculations. Additionally, the definition of a counterattacking event may be flawed; the labeling algorithms identify counterattacks as sequences starting in the defensive half, excluding set pieces, where the ball moves at least 10 meters at a velocity of 4 m/s or higher, and label them as successful if the ball enters the 18-yard box and ends within a defined ellipse. However,

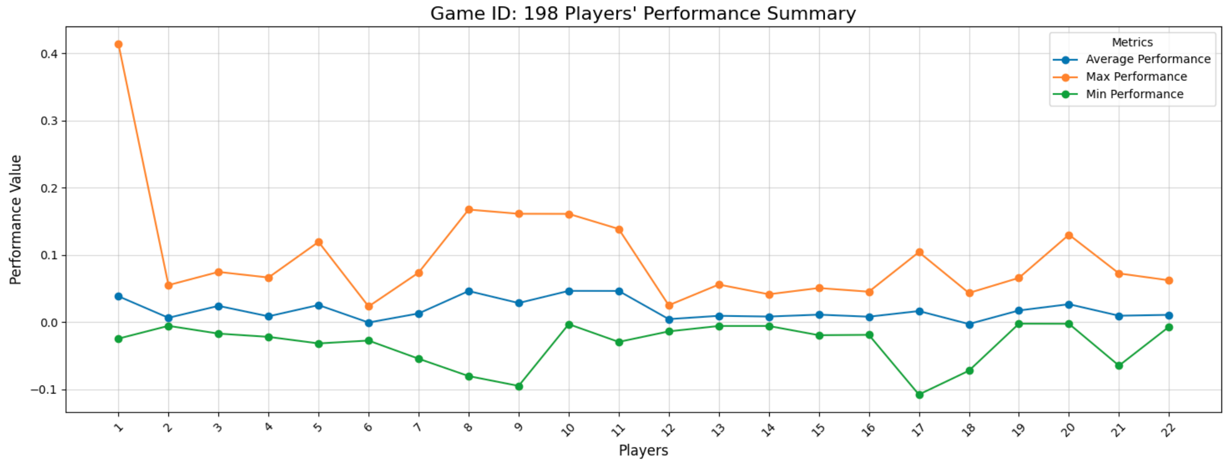

Figure 14: Player's performance in game 198

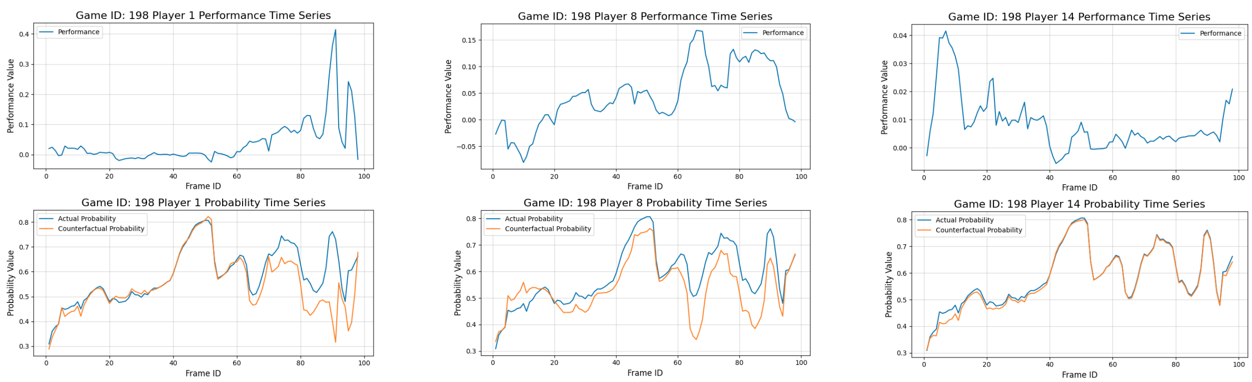

Figure 15: Evaluation on player 1, 8 and 14

about 50% of samples are found to be mislabeled upon reviewing data quality. Third, rather than focusing solely on players' next locations, we could analyze their movement paths and integrate them with passing actions. A more detailed modeling of player actions could better distinguish key players, although it may complicate the modeling process and require high-quality datasets. We encourage researchers in sports analytics to experiment with this framework across different event types and datasets.

7 Reproducibility

To ensure reproducibility of the results, we provide a complete GitHub repository containing all the code, configurations, and instructions required to replicate the experiments. You can access the repository at the following link: GitHub Repository for CounterEval.

References

[1] Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.

[2] Seokpyo Hong, Daseong Han, Kyungmin Cho, Joseph S Shin, and Junyong Noh. Physics-based full-body soccer motion control for dribbling and shooting. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019.

[3] Amod Sahasrabudhe and Joris Bekkers. A graph neural network deep-dive into successful counterattacks. In *17th Annual MIT Sloan Sports Analytics Conference. Boston, MA, USA: MIT*, page 15, 2023.

[4] William Spearman, Austin Basye, Greg Dick, Ryan Hotovy, and Paul Pop. Physics-based modeling of pass probabilities in soccer. In *Proceeding of the 11th MIT Sloan Sports Analytics Conference*, volume 1, 2017.

[5] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

[6] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. *arXiv*, October 2017.