YuShan2026: Team Description Paper

Zekai Cheng[1][0009-0002-8792-8735], Yuhang Zhang[1], Xun Shao[1], Jian She[1],
Ximing Wang[1], and Xiujun Wang[1]*

Anhui University Of Technology, Ma'anshan, China
yushan2d@gmail.com, xjwang@ahut.edu.cn

Abstract. This paper addresses the issue in traditional shot optimization methods for the RoboCup Simulation 2D project, which treat shooting as a single, homogeneous action and overlook the differences in shot formation methods. Inspired by modern football theory, we propose an optimization analysis approach centered on the key pass shot form. The study clearly defines "key pass shot" within this domain and, through empirical analysis of a total of 1200 matches (yielding 4036 valid shot data points), verifies its high success rate of 90.1%, significantly higher than the 79.2% success rate of non-key pass shots. To investigate its formation mechanism, we constructed a machine learning classification model based on XGBoost and performed analysis combined with SHAP interpretability. The research found that passing possession rate and passing frequency within the shooting zone are the most significant positive factors contributing to key pass shots. This reveals that reducing individual dribbling and improving team possession quality by increasing short-pass combinations (at least 4 or more) within the shooting zone is an effective path to creating high-efficiency shooting opportunities. Furthermore, to address the analytical shortcomings in penalty shootouts, the team developed a lightweight automated penalty success rate analysis tool capable of automatically parsing logs and evaluating goalkeeper performance, providing efficient data support for strategy optimization. This study validates the feasibility of optimizing shot forms through passing collaboration, offering new theoretical foundations and quantitative directions for offensive strategy development in RoboCup 2D.

Keywords: RoboCup 2D Simulation · Key Pass Shot · XGBoost · SHAP

1 Introduction

Since 2012, the YuShan team has participated in 11 RoboCup World Cup competitions, achieving 4 third-place finishes and securing the runner-up position at the 2025 Brazil World Cup. Over the past 13 years, we have won the China Open 8 times, placed second 4 times, and finished third once.

In recent years, the YuShan team reorganized its project based on Agent-2D 3.1.0 [1] and the new version of librcsc [2], and developed the `YuShan_Base`

* Corresponding author.

foundational layer, which is completely equivalent to the Agent-2D base code framework. In 2019, the team innovatively proposed the concept of integrating digital twin technology with the optimization and development of RoboCup 2D. Based on digital twin theory, data mining techniques are used to deeply analyze the characteristics of various teams. Through the methods and processes of the digital twin framework [3], we can identify key research and development directions. Building on this foundation, and by referencing the open-source code of teams like Gliders and CYRUS, the team has conducted in-depth analysis and improvements on team formation layout, offensive-defensive transitions, player positioning, and defensive details.

2 Related Work

This section highlights some of the YuShan team's past work:

To enhance the decision evaluation capability of simulated soccer robots in complex, dynamic adversarial environments, the YuShan team previously conducted research on evaluation models based on learning. The team introduced Long Short-Term Memory networks (LSTM) into the soccer robot decision evaluation process and combined them with genetic algorithms to optimize the evaluation network [4]. This enables the evaluator to better capture the temporal characteristics of match states, thereby supporting soccer robots in selecting better action parameters in real-time.

Addressing the core issue of shot optimization, traditional research often treats shooting as a single, homogeneous technical action, primarily improving success rates by adjusting parameters like shot angle and power, while neglecting the essential differences between various shot formation methods. However, modern football theory and practical experience indicate that shooting opportunities created through team coordination are often more threatening than individual, solo attempts. Inspired by this, the YuShan team further focuses its research on the specific shot form of key pass shots, aiming to enhance overall offensive efficiency and goal quality by optimizing the pass-shot collaborative decision-making process.

3 Research on Factors Influencing Key Pass Shots

In modern football, a key pass is defined as: a passing action that directly creates a shooting opportunity [6]. Correspondingly, a key pass shot refers to a shot form based on such a pass. In the logs generated by RoboCup 2D matches, a kick is the only type of ball contact action for a player besides a tackle. Based on the modern football definition and the characteristics of simulation matches, the RoboCup 2D key pass shot is defined as follows: Taking the period from the start of the match until the ball stops after a shot action as a complete shot cycle, if the kick action immediately preceding the shot kick is produced by a teammate who is not the shooter, then this shot is considered a key pass shot; otherwise, it is considered a non-key pass shot.

3.1 Feature Analysis Model Construction

Addressing the issue that current shot optimization research does not consider shot forms, and combining relevant conclusions from modern football, we establish a shot optimization analysis approach centered on the key pass shot form. The specific research steps are as follows: first, define the criteria for key pass shots in the simulation environment; second, use actual match-generated shot data for statistics to verify their strategic value; then, select effective features and construct an XGBoost classification model; finally, use SHAP interpretability analysis to parse the association mechanism between key features and key pass shots, providing reliable theoretical support for team shot optimization.

To explore the key influencing factors and mechanisms of key pass shots, this paper employs the XGBoost algorithm to build a prediction model and performs interpretability analysis on the model results based on SHAP.

XGBoost is an ensemble algorithm based on decision trees. It improves prediction and classification performance by sequentially training multiple weak classifiers and combining them into a strong classifier. XGBoost adds a regularization term to the cost function to control model complexity, effectively reducing model variance to prevent overfitting, and possesses strong generalization capabilities. By introducing parallel computing, the gain calculations for each feature can be performed multi-threaded, greatly enhancing operational efficiency [8].

To ensure better performance of the XGBoost model, hyperparameter tuning is necessary. This paper selects the grid search method as the hyperparameter search algorithm. Grid search involves traversing every possible parameter combination within a given hyperparameter range and selecting the combination with the highest average score as the optimal choice.

The parameters set for XGBoost using grid search are shown in Table 1.

Table 1: XGBoost Parameter Settings Table

Parameter	Description	Value
learning_rate	Learning rate, controls the step size when updating weights in each iteration	0.05
n_estimators	Number of decision trees	600
max_depth	Max depth of the decision trees	8
colsample_bytree	Proportion of features used to train each tree relative to all features	0.8

After completing model training and testing, interpretability analysis of the model output is required. The SHAP method proposed by Scott M. Lundberg [9] offers stronger feature interpretability compared to other model interpretation methods and can provide visual global or local explanations for the model. Its core idea is to measure the contribution of each influencing factor to the model's prediction results by calculating the average marginal contribution when each factor is added to the model. The calculation formula for the SHAP value of the i-th influencing factor is as follows:

$$\Phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (v(S \cup \{i\}) - v(S)) \quad (1)$$

In the formula: N is the set of all features; S is the set excluding the i-th influencing factor; $v(S)$ and $v(S \cup \{i\})$ represent the contributions of the set S and the set $S \cup \{i\}$ (which includes the i-th influencing factor) to the model's prediction output, respectively.

3.2 Dataset Construction

This paper selects the 2024 China competition team YuShan2024 as the baseline testing side, conducting 300 matches each (totaling 1200 matches) against four teams: the 2023 World Cup team HELIOS2023, the 2024 World Cup team CYRUS2024, the 2024 China competition team MT2024, and the 2024 China competition team HFUTEngine2024. Based on the log data generated from the matches, the open-source log analysis tool from the Japanese Helios team [7] is used for parsing to construct a structured shot form dataset. For each shot cycle, the following are extracted within the offensive zone and shooting zone: the number of passes/dribbles, the average action distance (using Euclidean distance, i.e., calculating the displacement length between the start and end points of an action), and the pass/dribble possession rate (the ratio of possession time to the total cycle duration). Simultaneously, the shot form is labeled: key pass shots are labeled as 1, and non-key pass shots are labeled as 0.

3.3 Experimental Results Analysis

The obtained dataset was cleaned by removing missing data and extreme scenario data samples (where 9 or more out of 12 feature values were 0), resulting in 4036 valid shot data entries. Among these: there were 1,451 key pass shot samples and 2,585 non-key pass shot samples. The success rates of key pass and non-key pass shots in the shot dataset are shown in Table 2.

Table 2: Shot Success Rate Statistics Table

Shoot type	Successful shots	Failed shots	Total shots	Successful rate
Shot from key pass	1308	143	1451	90.1%
Shot from non-key pass	2047	538	2585	79.2%

As shown in Table 2, the success rate of key pass shots is 10.9% higher than that of non-key pass shots. This result verifies that key pass shots also possess

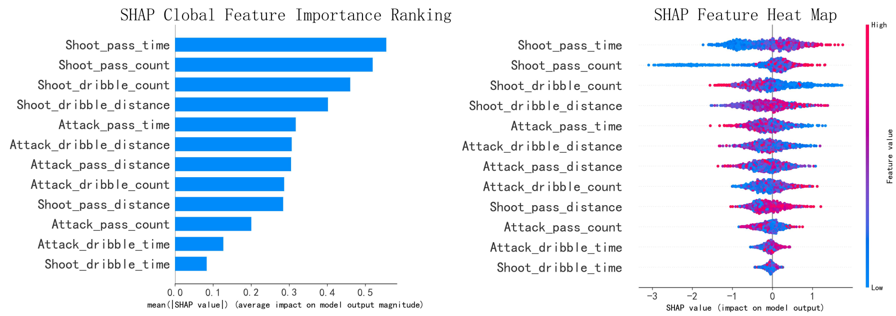


Fig. 1: Feature Importance Analysis Chart

strategic value in the RoboCup 2D domain and proves the effectiveness of the research approach aimed at shot optimization by optimizing shot forms.

From the feature importance ranking chart on the left of Figure 1, it can be seen that passing possession rate in the shooting zone and passing frequency in the shooting zone are important factors influencing key pass shots. From the SHAP heatmap on the right of Figure 1, it can be observed that for these two important features—passing possession rate in the shooting zone and passing frequency in the shooting zone—samples with high feature values (dark color) are concentrated in the SHAP > 0 region, indicating that both have a positive effect on forming key pass shots. For the feature of dribbling frequency in the shooting zone, samples with high feature values (dark color) are mainly concentrated in the SHAP < 0 region, while samples with low feature values (light color) are distributed in the SHAP > 0 region, indicating that this feature has a negative correlation with pass shots.

Figure 2 shows the variation of SHAP values for the passing frequency in the shooting zone across dataset samples. The passing frequency within the shooting zone for samples is mainly concentrated between 0-9 times. When the passing frequency is between 4 and 6 times, the SHAP values for the vast majority of samples are > 0. When the passing frequency is between 7-9 times, there is a slight decline. When the passing frequency is greater than 9 times, although there are fewer samples, all samples have SHAP values > 0. From this, it can be concluded that the passing frequency within the shooting zone should be at least greater than 4 times.

In addition to the main influencing features mentioned above, for the two features of average passing distance in the shooting zone and average dribbling distance in the shooting zone: samples with both high and low feature values are distributed on both sides of the SHAP axis, indicating that their influence is scenario-dependent. In the sample explanation chart for sample 102 shown in Figure 3(a), the average passing distance and average dribbling distance within the shooting zone are 12.834 and 1.815, respectively, and both play a negative role. The passing and dribbling counts are 5 and 29, respectively. This suggests

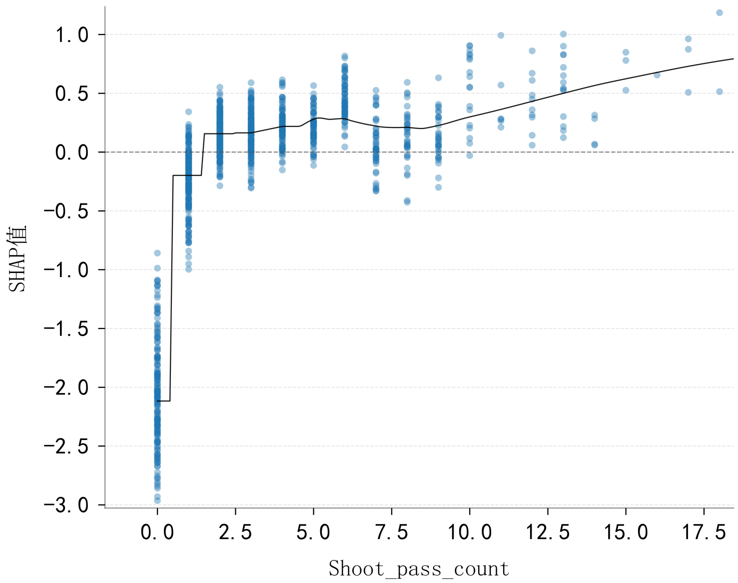
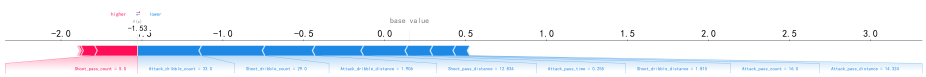
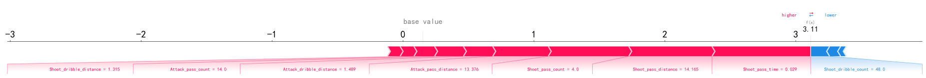


Fig. 2: SHAP Variation Chart for Passing Frequency in Shooting Zone



(a) Sample 102 (Non-Key Pass Shot Sample)



(b) Sample 817 (Key Pass Shot Sample)

Fig. 3: Single Sample SHAP Explanation Charts

that the long-distance pass in this sample may not have broken the opponent's defense but merely completed a simple change of possession, ultimately relying on individual dribbling to complete the shot.

In contrast, in the sample explanation chart for sample 817 shown in Figure 3(b), the average passing distance and average dribbling distance within the shooting zone are 14.165 and 1.315, respectively, and both play a positive role. The passing and dribbling counts are 4 and 48, respectively. This suggests that this sample might have attracted opponent attention through multiple short-distance dribbles and broken the opponent's defense with a few high-quality long-distance passes, completing a key pass shot.

In summary, passing possession rate and passing frequency are the main features influencing key passes. Reducing dribbling frequency within the shooting zone and improving team coordination to form high-quality passes is an effective path to creating key passes.

4 Automated Penalty Success Rate Analysis Tool

Penalty shootout technique is a crucial aspect determining match outcomes in RoboCup 2D. However, existing log analysis tools mostly focus on data statistics from regular match phases and lack specialized analysis functions for penalty shootout phases. The YuShan team has developed a lightweight automated penalty success rate analysis tool aimed at providing data support for optimizing team penalty strategies.

This tool parses based on the RoboCup 2D standard log formats (.rcg and .rcl files). By extracting key event information from the penalty phase, it automatically tallies the goalkeeper's defensive performance.

The core functions of the tool include:

- **Log Parsing and Event Extraction**: Automatically identifies the penalty phase within match logs, extracting key events such as shots, saves, and goals, along with detailed information like timestamps, player numbers, and action types.
- **Goalkeeper Performance Statistics**: Calculates basic metrics such as the number of shots faced by the goalkeeper, number of successful saves, and number of goals conceded, and generates key evaluation indicators like save success rate and concession rate.
- **Visual Report Generation**: Outputs analysis results in a structured format, including a timeline of events, a summary of statistical data, and a goalkeeper performance rating, facilitating a quick understanding of the strengths and weaknesses in the penalty phase.

The tool is implemented using Python scripts, characterized by being lightweight, easy to deploy, and efficient in operation. It can quickly complete batch log analysis without complex environment configuration. Through this tool, the team can systematically evaluate the penalty characteristics of different opponents, analyze the defensive patterns of its own goalkeeper, and provide quantitative basis for targeted training and strategy adjustments.

5 Summary and Future Work

This study addresses the issue of traditional shot optimization neglecting different shot forms by proposing a new optimization approach centered on "key pass shots," defining its form and verifying its relatively high success rate and tactical value. Through analysis with XGBoost and SHAP models, it is concluded that passing possession rate and passing frequency in the shooting zone are the core

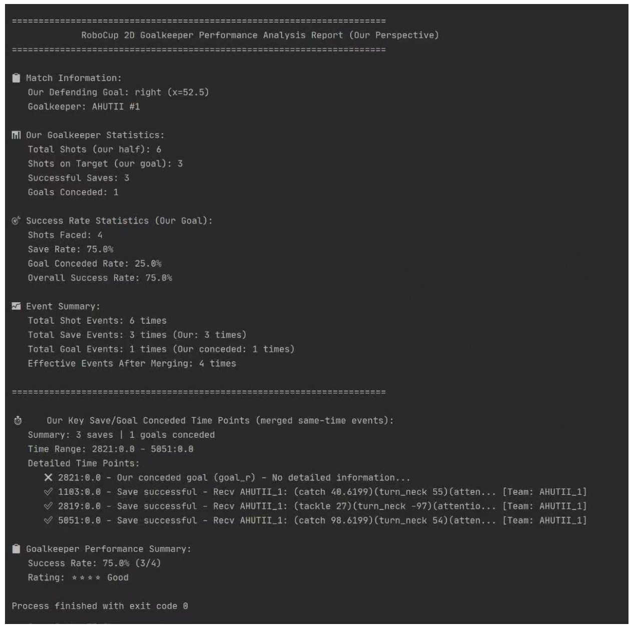


Fig. 4: Penalty Success Rate Analysis Chart

influencing factors. Simultaneously, the team developed an automated penalty success rate analysis tool for rapid strategy evaluation. Future work will refine the characteristics of different teams and conduct more in-depth regional and strategic analysis.

In our future work, we hope to learn lessons from the Brazil World Cup experience, comprehend and absorb the essence from CYRUS's open-source code. We also aim to modify the dribbling logic of penalty-taking players and the positioning logic of the goalkeeper's movement to achieve better results in penalty shootouts. The YuShan team has also noted that humanoid robots will be fully launched after the 2027 World Cup. We consider this a positive signal and a push towards the ultimate goal of RoboCup 2050. We also hope the RoboCup Federation will retain the 2D simulation project.

We thank the HELIOS, CYRUS, WrightEagle, and Glider teams for their open-source technical support. We also thank the excellent competition teams for their support of this event and hope the 2D league continues to improve. Furthermore, we also plan to explore advanced sensing and secure data sampling methods, such as RFID and differential privacy techniques [10,11], to enhance our experimental validation framework in the future.

References

1. Hidehisa Akiyama, Tomoharu Nakashima, HELIOS Base: An Open-Source Package for the RoboCup Soccer 2D Simulation, In Sven Behnke, Manuela Veloso, Arnoud Visser, and Rong Xiong editors, RoboCup 2013: Robot World Cup XVII, Lecture Notes in Artificial Intelligence, Springer Verlag, Berlin, 2014. http://dx.doi.org/10.1007/978-3-662-44468-9_46
2. Hidehisa Akiyama, Itsuki Noda, Multi-Agent Positioning Mechanism in the Dynamic Environment, In Ubbo Visser, Fernando Ribeiro, Takeshi Ohashi, and Frank Dellaert, editors, RoboCup 2007: Robot Soccer World Cup XI Lecture Notes in Artificial Intelligence, vol. 5001, Springer, pp.377-384, July 2008. https://doi.org/10.1007/978-3-540-68847-1_38
3. Zekai Cheng, Ningyu Xie, Sichun Yang, & Xingxing She. (2020). Application of Digital Twin Framework Based on RoboCup 2D Simulation. Journal of Suzhou University of Science and Technology (Natural Science Edition), 04, 64-67.
4. Anhui University of Technology. (2025). Evaluation Method of Simulated Soccer Robots Based on Genetic Algorithm Optimized LSTM Network (Patent No. 202411517958.8).
5. Mei Chen. (2016). Formation Study of RoboCup 2D Simulation Based on Density Peak Clustering (Master's Thesis, Anhui University of Technology).
6. Hao Yang, Lu Pang, Qin Liu, et al. (2023). Research on the Mechanism of Passing Influence on Shooting in Modern Professional Football: A Chain Mediation Effect Model. Journal of Sichuan Vocational and Technical College, 33(05), 97–103.
7. Defu Long. (2016). Mining and Application of RoboCup 2D Simulation Match Logs (Master's Thesis, Guangdong University of Technology).
8. BIAU G, CADRE B, ROUVIERE L. Accelerated gradient boosting[J]. Machine learning, 2019, 108(6): 971-992.
9. LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[J]. Advances in neural processing systems, 2017, 30(1):4675-4774.
10. Wang, X., Mo, L., Guo, L., Lu, Z., Liu, Z., Xue, M.: Online Streaming Sampling Publication Method Over Sliding Windows With Differential Privacy. IEEE Transactions on Dependable and Secure Computing **22**(6), 6896–6912 (2025)
11. Wang, X., Liu, Z., Zhou, X., Liao, Y., Hu, H., Zheng, X., Li, J.: A Near-Optimal Category Information Sampling in RFID Systems. IEEE Transactions on Mobile Computing (2025)