Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Can Match Events Predict an Opponent's Next Goal Kick? A Binary-Classification Study for Football Coaches

Abdelhaq Elmetlini

abdelhaq.elmetlini-etu@etu.univh2c.ma

ENSEM, Hassan II University

Hicham Medromi

ENSEM, Hassan II University

Research Article

Keywords: event data, coaching decision support, match status, logistic regression, machine learning, tactical analysis

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11362986/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

A goal kick gives defending staff a short opportunity to prepare for a short build-up or a long first pass. This study tests whether events recorded before a goal kick can forecast that choice without using information from the kick itself. We replayed 346 matches from six StatsBomb Open Data tournaments. A first pass of at least 40 yards was labelled long. Logistic regression was trained on 731 Euro 2020 kicks and evaluated chronologically on 968 kicks from the 2022 men's World Cup, 1,114 from the 2023 women's World Cup, and 766 from Euro 2024. The main comparator was an online team-habit rule updated only with earlier kicks. In Euro 2024, teams used long first passes on 60.7% of kicks when ahead and 34.8% when behind. Logistic regression reached 73.5% accuracy and an area under the receiver operating characteristic curve of 0.786, compared with 62.8% and 0.663 for the habit rule. In the women's external cohort, logistic regression did not clearly improve accuracy over habit, while Random Forest reached 72.2% in an exploratory comparison. The results show that past-only event features can improve short-versus-long forecasting in one held-out competition, but transfer varies across tournaments. Retrospective event data cannot establish live timing, causal effects, or coaching benefit.

1. Introduction

When an opponent is about to take a goal kick, the defending coach must quickly decide whether to organise for a short build-up or prepare for a longer first pass and its second ball. A forecast made before the pass may provide useful context, but it cannot replace the coach's view of player positions, space, and intentions. Interviews with high-level football coaches show that information use changes with the situation and the time available [1].

The 2019 Laws of the Game changed goal-kick restarts by allowing the ball to be in play once it was kicked and clearly moved, without first leaving the penalty area [2]. This widened the practical options for short restarts. Previous studies have examined what happens after short or long goal kicks [3] or modelled goal-kick success using information from the kick and opposition [4]. Our question is different: can earlier match events forecast whether the opponent's next first pass will be short or long before that pass is recorded?

A strong operational baseline is the opponent's usual choice. Staff may already know that a team generally plays short or long, so a useful forecasting model should improve on an adaptive team tendency rather than only on a global majority. We therefore use a chronological, past-only replay and test the model in separate tournaments. The study contributes a leakage-controlled next-goal-kick prediction task, an adaptive team-habit comparator, and competition-separated evaluation with both traditional classifiers and exploratory event-graph models.

2. Related work and research question

Goal-kick studies show the tactical importance of restart choice but usually investigate subsequent outcomes rather than forecasting the choice itself [3, 4]. Pass-event research also shows that earlier

passing behaviour can characterise team style and support later-event prediction [5]. Work on football match forecasting highlights the value of chronological data, domain knowledge, and meaningful baselines [6, 7], while hybrid approaches show that human-written context can add information beyond statistical records [8]. Game interruptions can also have tactical value, but their duration varies with context [9]. Position data can describe formation and coordination more directly [10], whereas the present analysis deliberately uses event data only. Table 1 summarises how the most relevant studies relate to the present task.

Table 1
Prior studies relevant to goal-kick forecasting and their connection to the present task

Study	Measured outcome or task	Connection to present study
Mendes et al. [3]	Offensive sequence after an observed short or long goal kick	Shows tactical interest of restarts; does not forecast the kick choice.
Brooks et al. [5]	Team pass style and subsequent shots from pass events	Supports using past passes to describe behaviour.
Berrar et al. [7]	Future full-match results using football-specific histories	Motivates chronological data and substantive baselines.
Siegle and Lames [9]	Types and duration of game interruptions	Motivates future testing of the time available at a restart.

Graph convolutional networks (GCN) and graph isomorphism networks (GIN) learn from connected observations [11, 12]. For this task, however, a graph must contain only events that happened before the kick being predicted. The primary research question is therefore whether a past-only classifier improves on the opponent's updated usual option in a held-out tournament, and whether that improvement transfers to another competition.

3. Methods

3.1. Study design, data, and target

This was a retrospective observational study with a separate forecasting evaluation. We used public match and event JSON from Hudl StatsBomb Open Data [13, 14]. Six complete tournament cohorts contributed 346 matches. The 2018 men's World Cup and 2019 women's World Cup were descriptive cohorts only. Euro 2020 was the training competition, the 2022 men's World Cup was a chronological validation cohort, Euro 2024 was the final men's test, and the 2023 women's World Cup was treated as a separate external evaluation (Table 2).

Table 2
Tournament partition and recorded goal kicks. Regulation periods 1 and 2 are included; extra-time and shootout kicks are excluded.

Competition	Role	Matches	Kicks	Long (%)
Men's World Cup 2018	Earlier cohort, descriptive	64	684	79.4
Women's World Cup 2019	Early post-change, descriptive	52	891	66.7
UEFA Euro 2020 (played 2021)	Train	51	731	49.4
Men's World Cup 2022	Validation	64	968	48.5
Women's World Cup 2023	External test	64	1,114	61.8
UEFA Euro 2024	Final test	51	766	45.6

Each observation was an event tagged as a Goal Kick. Only regulation periods 1 and 2 were used, and passes without a positive recorded length were excluded; all 5,154 included kicks had usable lengths. The target was the first pass itself: long if its recorded length was at least 40 yards and short otherwise. A short first pass followed by a long second pass remained labelled short. Sensitivity analyses rebuilt the target at 30 and 50 yards.

3.2. Pre-kick information and chronological replay

Matches were processed by date and events by period and recorded order. For every goal kick, the model could use only completed earlier matches and events already observed in the current match. The current kick's length, destination, height, passer, recipient, and outcome were hidden until after prediction. Score was reconstructed from earlier goals, excluding shootouts. After a kick was evaluated, its observed short/long label became available for subsequent kicks. Histories from the older descriptive cohorts never entered modern-model features. The 14 inputs shared by all models are listed in Table 3.

Table 3
The 14 shared pre-kick inputs. "Earlier" means before the goal kick being predicted.

Group	Input	Meaning
Match	Minute	Recorded goal-kick time; proxy for award time.
	Second half	Whether play is in the second regulation period.
	Home team	Whether the kicking team is designated as home.
	Score difference	Kicking team goals minus opponent goals so far.
Team history	Earlier long-kick share	Smoothed long-kick share from earlier modern-data kicks.
	Earlier kick count	Earlier goal-kick count entered as log(1 + N).
	Current-match long share	Smoothed long-kick share earlier in this match.
	Current-match kick count	Earlier goal-kick count in this match, log transformed.
	Last team kick	Short or long on the team's most recent earlier kick.
	Last kick in this match	Short or long on the team's latest earlier kick in this match.
Recent play	Mean pass distance	Mean team pass length in the previous five minutes.
	Pressured own-half passes	Team passes recorded under pressure from own half.
	Opponent pressure	Opponent pressure events in the previous five minutes.
	Pressure near goal	Opponent pressure events near the kicking team goal.

3.3. Baseline and predictive models

The online team-habit baseline counted earlier long kicks L and all earlier kicks N, with a four-kick prior centred on the Euro 2020 long-kick proportion (0.494). It predicted long when the resulting probability was at least 0.5. Unlike a static pre-match baseline, the online version also used earlier goal kicks from the current match.

The primary model was logistic regression with standardised numeric predictors and fixed L2 regularisation (C = 1). It was fitted only on the 731 Euro 2020 kicks. Secondary models were Random Forest, XGBoost, a radial-basis support vector machine (SVM), GCN, GIN, and a short explanatory decision tree (Table 4). All settings were fixed across held-out cohorts; no validation, external-test, or final-test outcomes were used to fit the primary logistic model. Additional classifier comparisons were exploratory because they were introduced after the initial test competitions had been examined.

Table 4
Models for the same short/long target and their fixed settings.

Model	Additional input	Fixed rule/settings
Online usual option	None	Earlier team-kick counts with training-only fallback.
Logistic regression	None	L2 regularisation, C = 1.
Short decision tree	Seven numeric inputs	Depth 2; at least 60 training kicks per leaf.
Random Forest	None	400 trees; Gini splitting; at least five samples per leaf.
XGBoost	None	120 trees; depth 3; learning rate 0.05; L2 penalty 3.
SVM (radial)	None	C = 1; training-only probability calibration.
GCN	Up to 16 earlier events	Two neighbour layers; 40 fixed epochs.
GIN	Up to 16 earlier events	Two sum-neighbour layers; 40 fixed epochs.

3.4. Evaluation and statistical analysis

We report accuracy, balanced accuracy, long-class F1, Brier score, cross entropy, area under the receiver operating characteristic curve (ROC-AUC), and confusion counts. Lower Brier score and cross entropy indicate better probability accuracy. The primary paired comparison was the logistic model minus the online team-habit rule on the same kicks. Whole matches were resampled 2,500 times to obtain percentile intervals while keeping repeated kicks and both teams within a match together. Score-state descriptions used the kicking team's score immediately before the kick (behind, level, or ahead), with match-bootstrap intervals for long-kick percentages and ahead-minus-behind differences.

4. Results

4.1. Match situation and observed choices

Across the four modern tournaments, teams taking goal kicks used a long first pass more often when ahead than when behind. In Euro 2024, 125 of 206 kicks were long when the kicking team was ahead (60.7%), compared with 170 of 405 when level (42.0%) and 54 of 155 when behind (34.8%). The ahead-minus-behind difference was 25.8 percentage points (95% match-bootstrap interval 10.5–42.1). In the women's external cohort, the corresponding rates were 73.4%, 58.6%, and 57.6% (Table 5; Fig. 1). These are observational associations and do not show that score state causes the choice.

Table 5
Observed goal-kick choice by the kicking team's score before the first pass.

Competition	Score	N	Long n	Long % [95% interval]
Euro 2020	Behind	169	60	35.5 [26.1, 44.8]
	Level	344	165	48.0 [38.7, 57.0]
	Ahead	218	136	62.4 [52.2, 71.8]
Men's WC 2022	Behind	201	77	38.3 [30.3, 45.9]
	Level	517	228	44.1 [37.1, 50.7]
	Ahead	250	164	65.6 [58.4, 72.7]
Women's WC 2023	Behind	349	201	57.6 [49.0, 65.6]
	Level	498	292	58.6 [51.6, 64.9]
	Ahead	267	196	73.4 [63.6, 82.0]
Euro 2024	Behind	155	54	34.8 [20.7, 47.0]
	Level	405	170	42.0 [34.3, 49.5]
	Ahead	206	125	60.7 [52.6, 68.2]

4.2. Held-out forecasting

Euro 2024 contained 766 kicks, 45.6% long; first-pass length distributions for all cohorts are shown in Fig. 2a. Static previous-match habit reached 61.0% accuracy, online habit 62.8%, and the last-kick rule 64.1%. The full logistic model reached 73.5% accuracy (95% match-bootstrap interval 70.4–76.7%) and ROC-AUC 0.786. Its paired accuracy gain over online habit was 10.7 percentage points (95% interval 6.3–15.3), while Brier score improved from 0.231 to 0.189 (Fig. 2b). The validation men's World Cup showed the same direction, 68.7% versus 62.3%. In Euro 2024 the model was more accurate than online habit in 30 of 51 matches, tied in 14, and less accurate in seven. The corresponding confusion matrices are shown in Fig. 3.

The external women's tournament provided a transfer check. Logistic accuracy was 68.9% versus 68.4% for online habit, a difference of 0.4 points with an interval crossing zero. A history-only logistic model reached 71.1%, and Random Forest reached 72.2% in an exploratory comparison. These results show that the match-context relationships learned in Euro 2020 did not transfer uniformly to this cohort; the difference cannot be attributed to gender from these data.

4.3. Exploratory classifier comparison

On Euro 2024, logistic regression remained the most accurate fixed seed-42 run at 73.5%, followed closely by GCN at 73.0%. Their ROC-AUC values were 0.786 and 0.784 (Fig. 4). Random Forest, XGBoost, SVM, and GIN also exceeded online habit in accuracy, but none clearly displaced logistic regression on this competition. In the women's external cohort, Random Forest reached 72.2% accuracy and ROC-AUC 0.796, compared with 68.4% and 0.723 for online habit; its paired accuracy-gain interval over habit included zero. Removing event links from GCN and GIN did not consistently worsen performance, so the experiment did not provide reliable evidence that the graph links added useful information beyond the node and tabular values.

5. Discussion

5.1. Interpretation for coaching

The score-state results answer a simple descriptive question: an opponent's current match situation is related to the observed distribution of goal-kick choices. In Euro 2024, long first passes were considerably more frequent when the kicking team was ahead than when it was behind, and the same ordering appeared in the other modern tournaments. However, team identity, opponent quality, and match context can all contribute to this pattern. The percentages therefore provide context for staff rather than a causal rule or tactical instruction.

The primary forecasting result is more operational. In Euro 2024, events available before a goal kick improved short-versus-long prediction over an adaptive baseline that followed the opponent's updated usual option. Most of the additional correct classifications came from recognising short kicks. A model probability could therefore be displayed as one input to staff preparation, for example by indicating the estimated chance of a long first pass together with the opponent's recent choices. Such an output would not specify which player to mark, whether to press, or whether the defensive response will succeed.

5.2. Transfer, model complexity, and limitations

Transfer was less convincing in the 2023 women's World Cup. The full logistic model did not clearly improve accuracy over the habit rule, while a history-only model and Random Forest performed better in exploratory comparisons. This suggests that team history may sometimes travel across competitions more reliably than contextual coefficients learned in a single training tournament. It also illustrates why a held-out result in one competition should not be treated as universal.

The more complex models did not show a consistent advantage. GCN was close to logistic regression on Euro 2024, but deleting graph links did not reliably reduce accuracy. Seed variation and the gap between training and validation loss in the original experiments also indicate that the small training cohort limits conclusions about complex architectures. The fixed 40-yard definition is a modelling choice; rebuilding the task at 30 and 50 yards still showed an advantage of logistic regression over habit within each definition, but the resulting accuracies measure different targets.

Several limitations prevent a claim of demonstrated real-time coaching benefit. The study is retrospective and uses event timestamps rather than the exact moment a live provider would announce a goal-kick award. The event feed cannot see the formation, passing lanes, goalkeeper cues, or coaching communication around the restart. Multiple kicks come from the same matches and teams; whole-match resampling addresses clustering in the reported intervals but does not represent every possible league or season. Finally, the expanded classifier comparisons were added after initial test results had been examined and should be interpreted as exploratory.

5.3. Practical next step

A prospective shadow deployment should run the frozen model beside a live event feed without sending tactical instructions. It should record the time the restart is awarded, when the required prior events become available, when the forecast is generated, and how many usable seconds remain before the kick. Only after timing reliability is established should a controlled study test whether concise forecasts improve staff decisions or defensive outcomes.

6. Conclusions

In the analysed tournaments, long first passes were more common when the kicking team was ahead, but this observational pattern is not causal. A past-only logistic model predicted short versus long goal kicks more accurately than the opponent's updated habitual choice in Euro 2024, while transfer to the 2023 women's World Cup was limited. Exploratory complex models did not provide a consistent advantage from event-graph links. The study supports further prospective testing of pre-kick forecasting, but live timing and coaching benefit remain untested.

Declarations

Funding:

This work was funded by the Foundation for Research, Development and Innovation in Science and Engineering, Casablanca, Morocco.

Clinical trial number: not applicable.

Consent to Publish declaration: not applicable.

Consent to Participate declaration: not applicable.

Ethics declaration: not applicable.

Competing interests: The authors declare that they have no competing interests.

Author Contribution

AE conceived the study, collected and analysed the data, developed the machine learning models, and drafted the manuscript. HM contributed to the study design, supervised the research, and critically reviewed the manuscript. Both authors read and approved the final manuscript.

Acknowledgement

The authors thank Hudl StatsBomb for making the open match and event data used in this study publicly available.

Data availability:

The event data analysed in this study are publicly available from Hudl StatsBomb Open Data (https://github.com/hudl/open-data) [13, 14].

References

1. Elstak I, Salmon P, McLean S. Naturalistic decision-making of football coaches. Int J Sports Sci Coach. 2026;21:1311–23. https://doi.org/10.1177/17479541251408666.
2. IFAB. (2019) Laws of the Game 2019/20 coming into force 1 June and new mobile application. https://www.theifab.com/news/laws-of-the-game-2019-20-coming-into-force-1-june-and-new-mobile-application/. Accessed 28 September 2026.
3. Mendes MHA, Rezende VHS, Praça GM. Association between goal-kick strategies and the offensive outcome in the UEFA Euro 2020. Hum Mov. 2023;24:64–70. https://doi.org/10.5114/hm.2023.125924.
4. Arnau Notari AR, et al. Using neural networks and hierarchical cluster analysis to study goal kicks in football. Int J Sports Sci Coach. 2024;19:1723–37. https://doi.org/10.1177/17479541231207184.
5. Brooks J, Kerr M, Guttag J. Using machine learning to draw inferences from pass location data in soccer. Stat Anal Data Min. 2016;9:338–49. https://doi.org/10.1002/sam.11318.
6. Berrar D, Lopes P, Dubitzky W. Incorporating domain knowledge in machine learning for soccer outcome prediction. Mach Learn. 2019;108:97–126. https://doi.org/10.1007/s10994-018-5747-8.
7. Berrar D, Lopes P, Dubitzky W. A data- and knowledge-driven framework for developing machine learning models to predict soccer match outcomes. Mach Learn. 2024;113:8165–204. https://doi.org/10.1007/s10994-024-06625-9.
8. Beal R, Middleton SE, Norman TJ, Ramchurn SD. (2021) Combining machine learning and human experts to predict match outcomes in football: a baseline model. In: Proceedings of the AAAI Conference on Artificial Intelligence 35:15447–15451. https://doi.org/10.1609/aaai.v35i17.17815
9. Siegle M, Lames M. Game interruptions in elite soccer. J Sports Sci. 2012;30:619–24. https://doi.org/10.1080/02640414.2012.667877.

10. Memmert D, Lemmink KAPM, Sampaio J. Current approaches to tactical performance analyses in soccer using position data. Sports Med. 2017;47:1–10. https://doi.org/10.1007/s40279-016-0562-5.
11. Kipf TN, Welling M. (2017) Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR).
12. Xu K, Hu W, Leskovec J, Jegelka S. (2019) How powerful are graph neural networks? In: International Conference on Learning Representations (ICLR).
13. Hudl StatsBomb (n.d.) StatsBomb Open Data. https://github.com/hudl/open-data. Accessed 28 September 2026.
14. Hudl StatsBomb (n.d.) StatsBomb Open Data Specification v1.1. https://github.com/hudl/open-data/blob/master/doc/StatsBomb%20Open%20Data%20Specification%20v1.1.pdf. Accessed 28 September 2026.

Figures

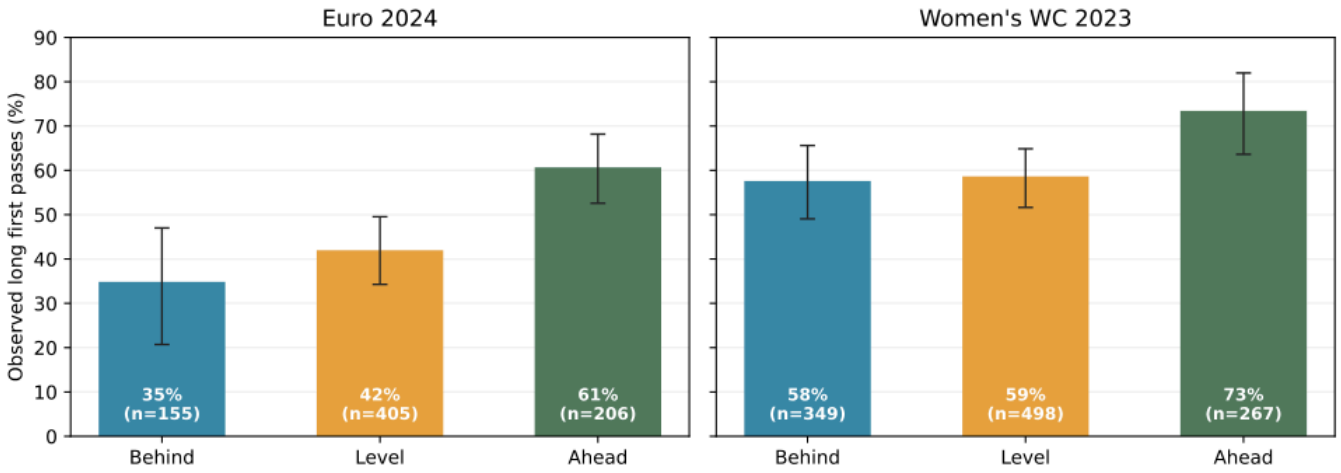


Figure 1

Observed share of long first passes by the kicking team's score before the kick. Error bars are 95% intervals from whole-match resampling

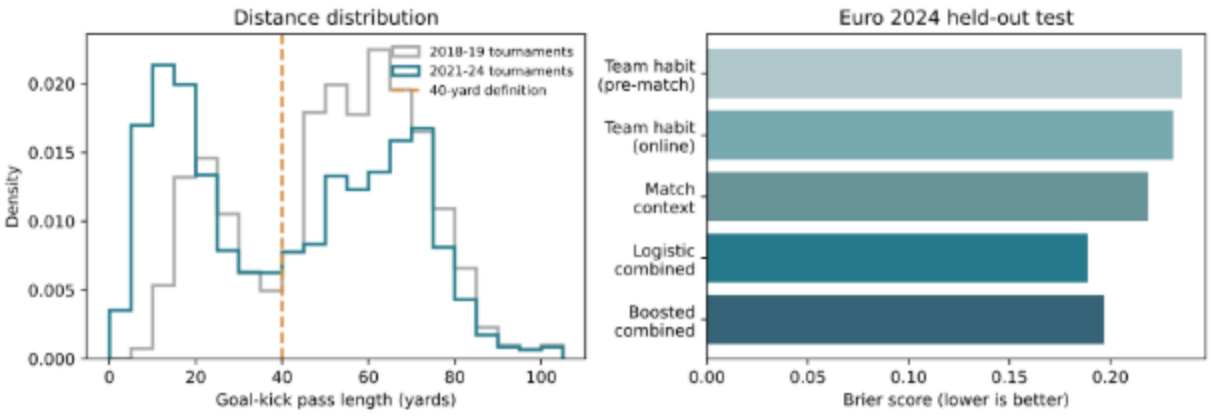


Figure 2

(a) Measured first-pass lengths in earlier versus modern cohorts. (b) Brier scores in the held-out Euro 2024 tournament

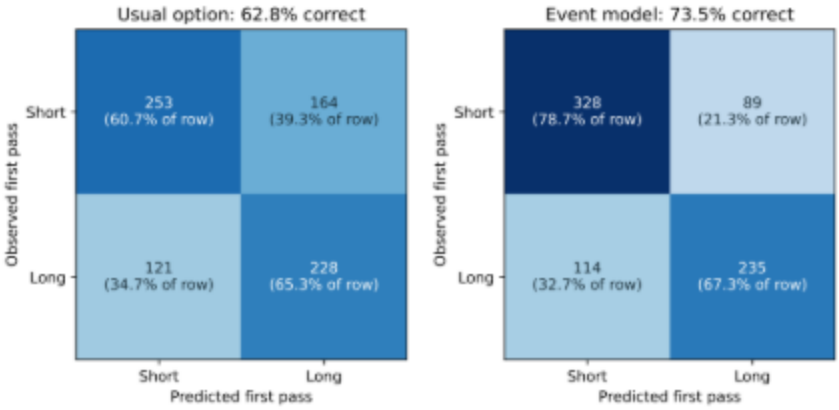


Figure 3

Euro 2024 confusion matrices for the online team-habit rule and full logistic model. Rows are observed choices and columns are predictions

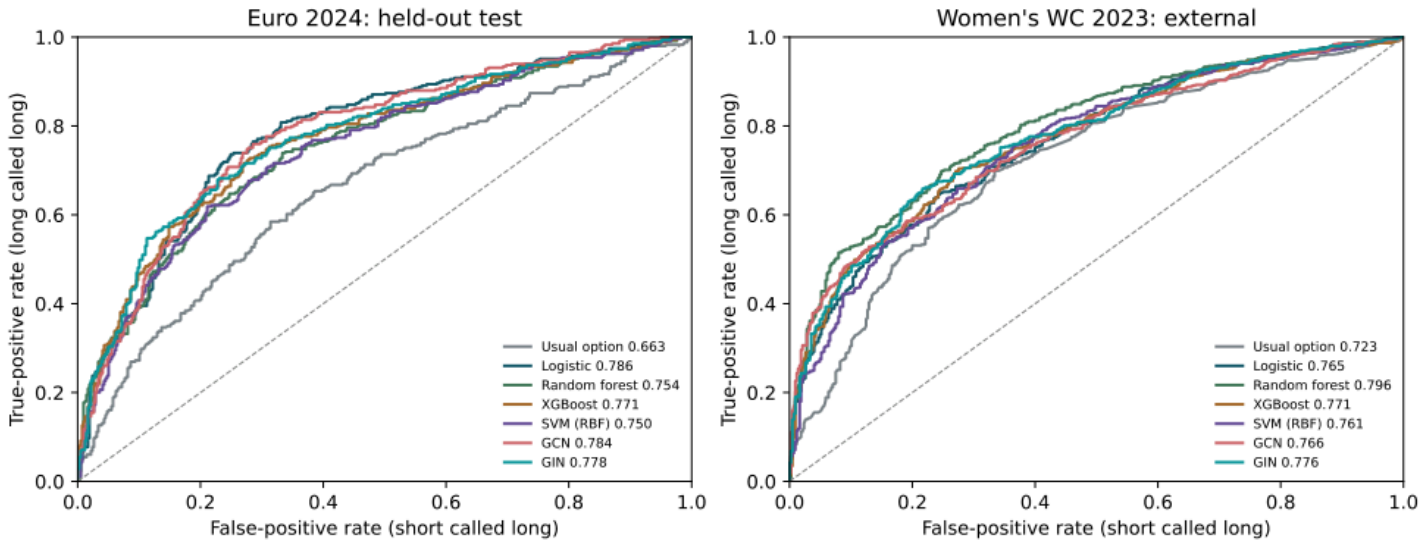


Figure 4

ROC curves for the team-habit and logistic benchmarks and five exploratory classifiers on Euro 2024 and the 2023 women's external cohort