Universität Zürich[UZH]

An Open-Source Implementation of FIFA's Enhanced Football Intelligence

- Dogan Parlak - 20743316
- Supervisor: Prof. Dr. Claudio J. Tessone
- Co-Supervisor: Prof. Dr. Ulrik Brandes

Contents

Enhanced Football Intelligence

- Recent technology and data-driven methods have revolutionized sports analytics, notably through FIFA-inspired Enhanced Football Intelligence (EFI) systems.

- FIFA has published a document named “Enhanced Football Intelligence” that includes football performance analysis and insights utilized in the FIFA World Cup 2022 .

Concepts

- Possession Control
- Phases of Play
- Ball Recovery Time
- Line Breaks
- Receptions Behind Midfield and Defensive Lines
- Defensive Line Height and Team Length
- Team Shape
- Final Third Entries
- Pressure on the Ball
- Forced Turnovers
- Expected Goal (xG)

Problem Statement

- The implementation of the methodologies in the EFI document are not publicly available through FIFA.
- Existing packages lack the incorporation of the latest methodologies from the FIFA World Cup 2022.
- Creating such a source is essential for establishing a common standard in statistics and performance analysis.

Research Question

- Are the concepts in the Enhanced Football Intelligence document sufficiently detailed to allow for the reproduction of FIFA's match reports?

Data

- The dataset comprises the games from the FIFA World Cup 2022. Each game is associated with three separate datasets:
  - Match Data
  - Event Data
  - Tracking Data

Syncronization of Event Data and Tracking Data

Implementation of the concepts require the application of both tracking data and event data

Time Conversion

Coordinate System Conversion

Implementation

- Automatic Event Detection Using Tracking Data
- Enhanced Football Intelligence Concepts
  - Possession Control
  - Phases of Play
  - Ball Recovery Time
  - Line Breaks
  - Receptions Behind Midfield and Defensive Lines
  - Defensive Line Height and Team Length
  - Team Shape
  - Final Third Entries
  - Pressure on the Ball
  - Forced Turnovers
  - Expected Goal (xG)

Automatic Event Detection Using Tracking Data

- Employs the tracking data to label each frame with the corresponding event taking place, enabling precise analysis and interpretation.

- Based on the state of the game.
  - Game in play (Ball Alive)
  - Game not in play (Dead Ball)

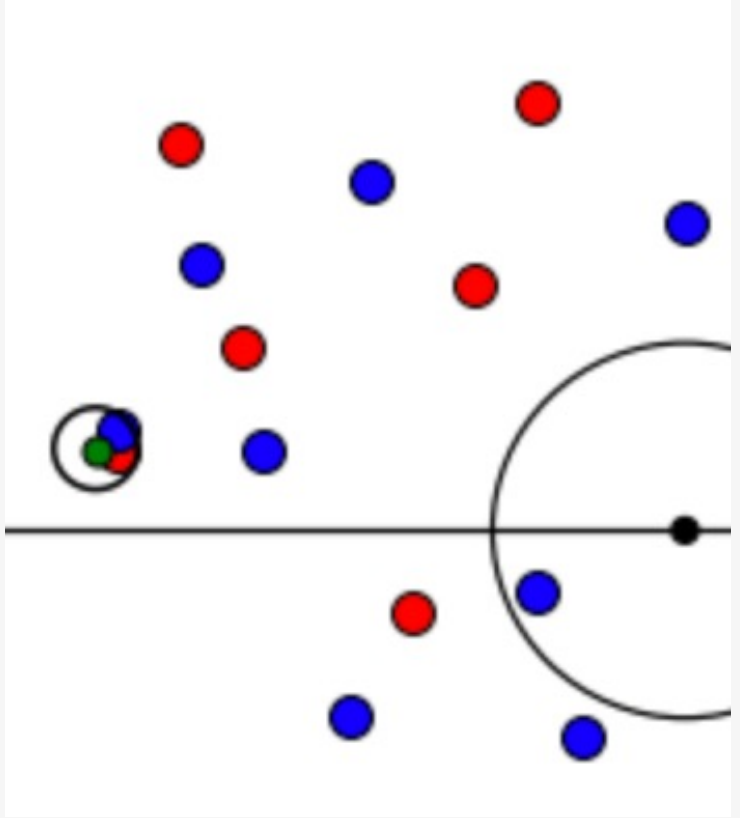
Game in Play

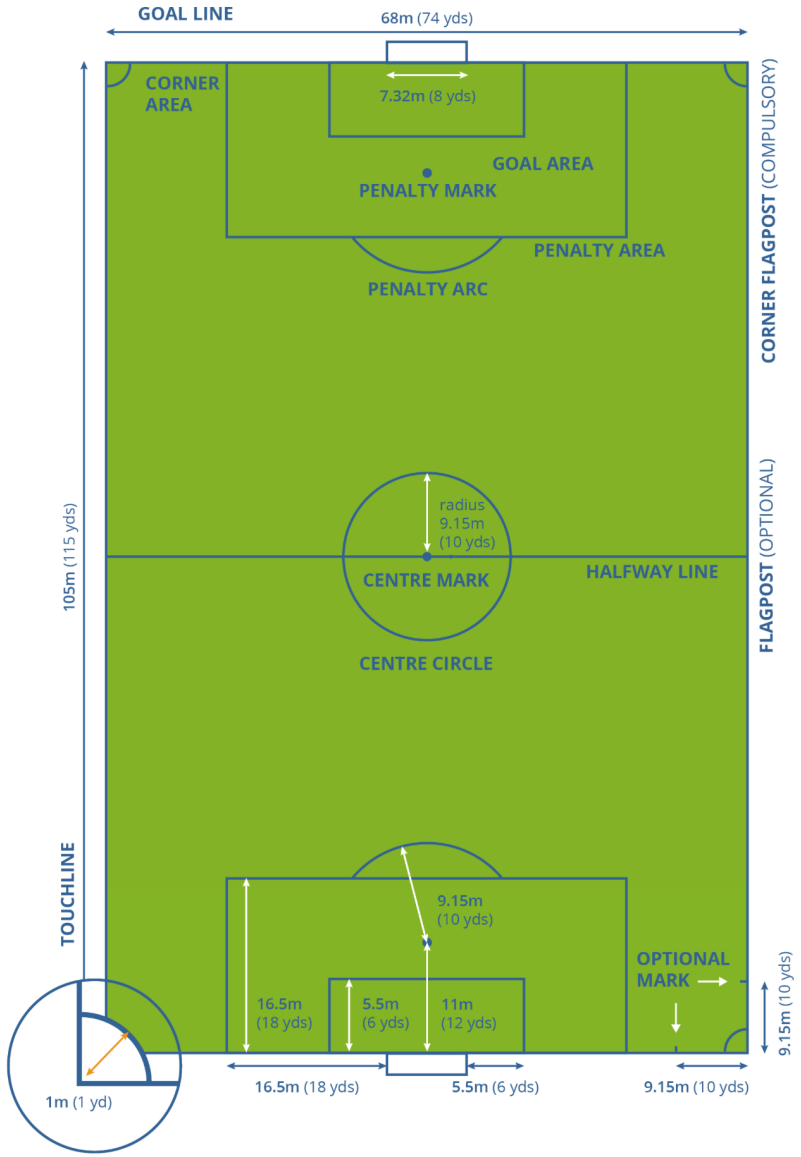
- Home team in-possession
- In contest
- Away team in-possession

In-Contest

- Consider a circle around the ball, which essentially defines the duel zone, with a radius of $\mathbf{r_{dz}}$ (2 meters).

- If there are at least two opponents within the duel zone, an in-contest situation is deemed to be occurring.



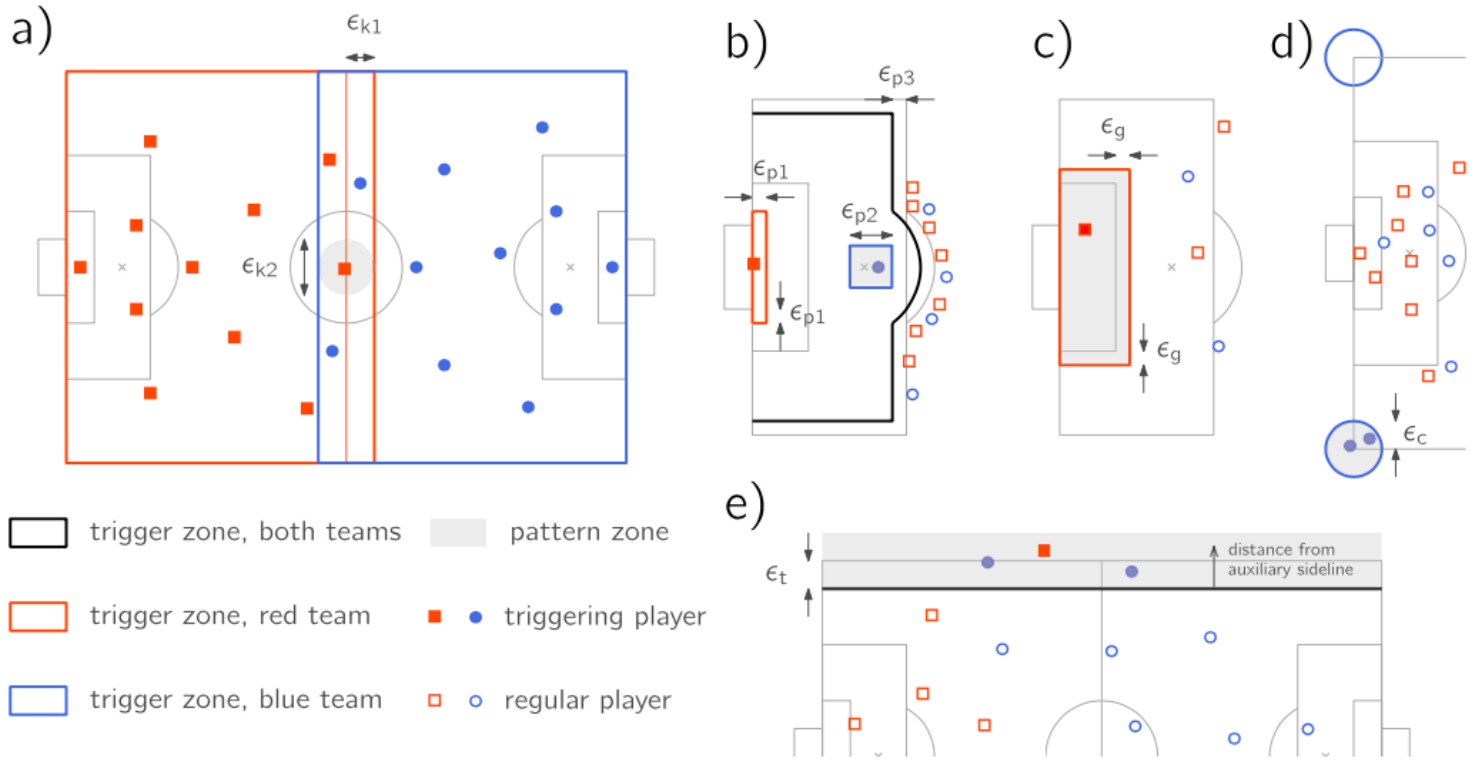


Game not in Play

- Kick-off
- Penalty
- Goal-kick
- Corner-kick
- Throw-in
- Free-kick

Set-Piece Triggers

a) Kick-Off
b) Penalty
c) Goal-Kick
d) Corner-Kick
e) Throw-In



Possession Control

EFI Definition

- The metric considers the ball alive frames for calculating possession distribution.
- Any output will generate:
  Home team in-possession (%), in-contest (%), away team in-possession (%).

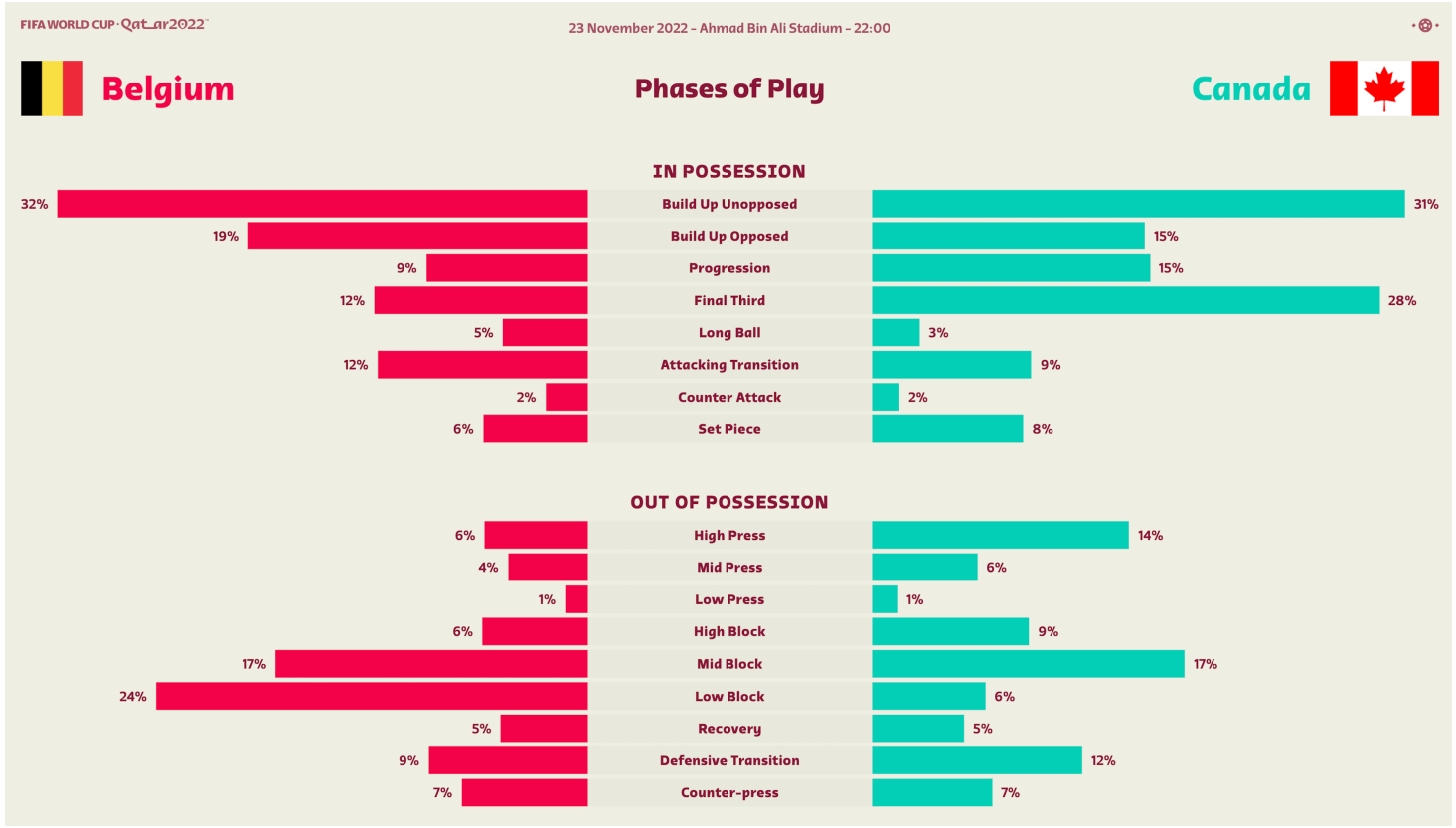
Implementation

$$team\ possessions\ (\%) = \frac{\#frames\ of\ the\ team\ in-possession}{\#ball\ alive\ frames} * 100$$

$$In\ contest\ (\%) = \frac{\#frames\ in-contest}{\#ball\ alive\ frames} * 100$$

Possession Control

Phases of Play



EFI Document

- Uses the percentage of ball-in-play time and provides insights into the strategies employed by teams throughout a match.
- For each team, the phases of play are separated into in-possession and out-of-possession segments.
- Detects these phases on a **frame-level** by analyzing:
  - Player and ball locations
  - Distances
  - Movement speeds
  - Directions

Phases of Play

Implementation

- ***SECTION 1:*** **Set-Piece**
- Automatic Event Detection algorithm to identify set-piece events when the ball is out-of-play.
  - Goal-kick
  - Corner- kick
  - Free-kick
  - Kick-off
  - Penalty
  - Throw-ins
- “set piece duration” (147 frames) is introduced to allocate after the game transitions into the in-play phase.

Phases of Play

Implementation

- ***SECTION 1:*** **Long Ball**
- Candidates: pass and cross events from event data.
- Success of the event does not affect the strategy of the team. Hence, any attempt for a potential long ball is accounted.
- Synchronize the event data with the tracking data.
- Filter out the ones that are not long ball.
  - $Vertical\ distance\ covered\ by\ the\ ball\ \geq$ "long ball distance" (45 meters)
  - $Height\ of\ the\ ball\ \geq$ "long ball height threshold" (3.24 meters) within the interval "long ball threshold" (197 frames)
- Allocate "long ball threshold" (197 frames) for each long ball event.

Phases of Play

Implementation

- ***SECTION 2: Counter-Attack***
- *Directness* and *Intensity*
- After each turnover, analyze:
  - *Directness:* Within the interval "forward threshold" (40 frames), define an edge formed by the ball locations at the beginning and end of this interval.

    $slope\ of\ the\ edge(\boldsymbol{m}) \geq$ "slope threshold" (0.51)

  - *Intensity:* Within the interval, "counter-attack threshold" (94 frames),

    $vertical\ distance\ covered\ by\ the\ ball\ (\boldsymbol{d}) \geq$ "forward distance threshold" (2.11 meters)

- Allocate "counter-attack threshold" (94 frames) for each counter-attack event.

Phases of Play

Implementation

- ***SECTION 2: Counter-Press***

- After each turnover:
  - Use the frames with pressure presence from the pressure on the ball concept output.
  - Within “counter press threshold” (21 frames), check the presence of any oppression.

- Allocate “counter press duration” (49 frames) for each counter-press event.

Phases of Play

Implementation

- ***SECTION 2: Recovery***
- The defending team should be quickly moving towards their own goal.
- For each turnover:
  - Number of frames the attacking team is in-possession of the ball is counted until another change in-possession is detected **(C)**.

    If, $\boldsymbol{C} \leq$ "$recovery\ threshold$"$(139\ frames)$ then, Interval **(I)** = **C**

    else, **I** = "$recovery\ threshold$"$(139\ frames)$
  - Average speed throughout the match (when game is in play) **(v_m)**, Average speed **(v),** start **(s)** and end **(e)** points covered during **I** by the defending team is obtained.
  - If $\boldsymbol{v} > \boldsymbol{v_m}\ and\ \boldsymbol{s_y} \geq \boldsymbol{e_y}$ then recovery phase is detected.
- Allocate **I** frames for each recovery event.

Phases of Play

Implementation

- ***SECTION 2: Defensive/Attacking Transition***

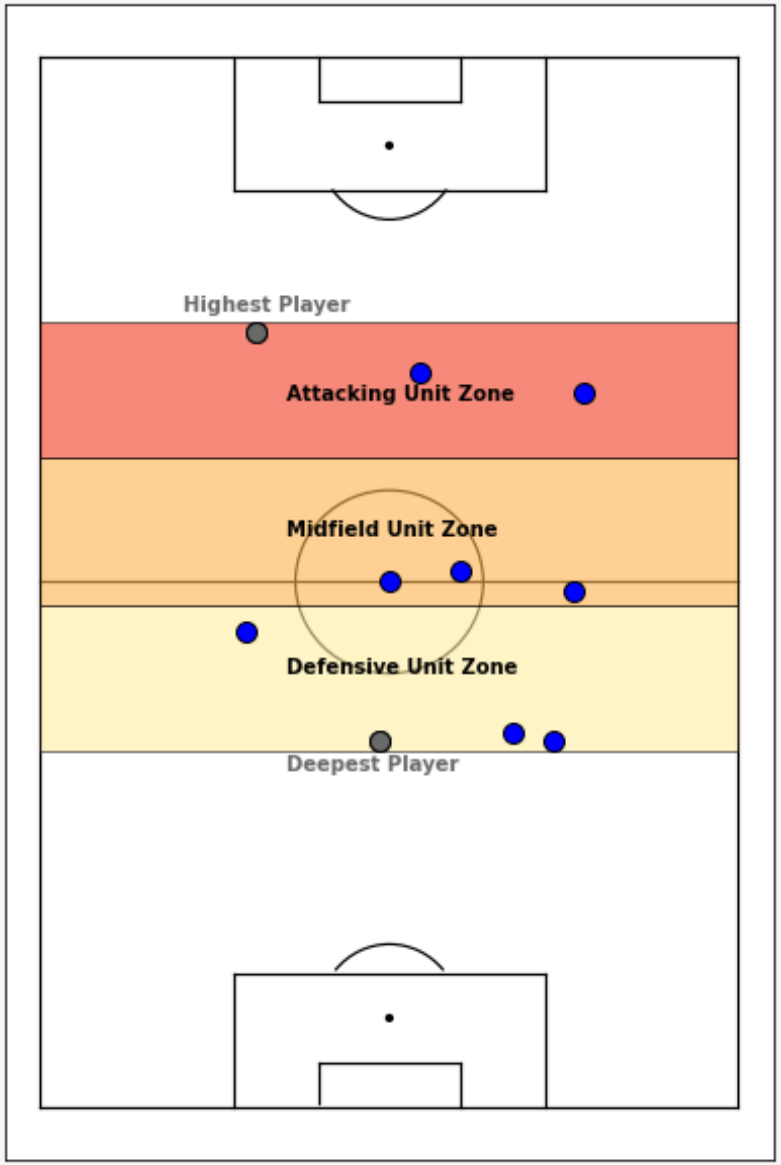
- Defensive transition phase of home/away team = Attacking transition phase of away/home team.

- Detecting only the defensive transition phase captures moments of the attacking transition phase as well.

- Defensive Transition = Counter Press + Recovery

Zonal Unit Classification

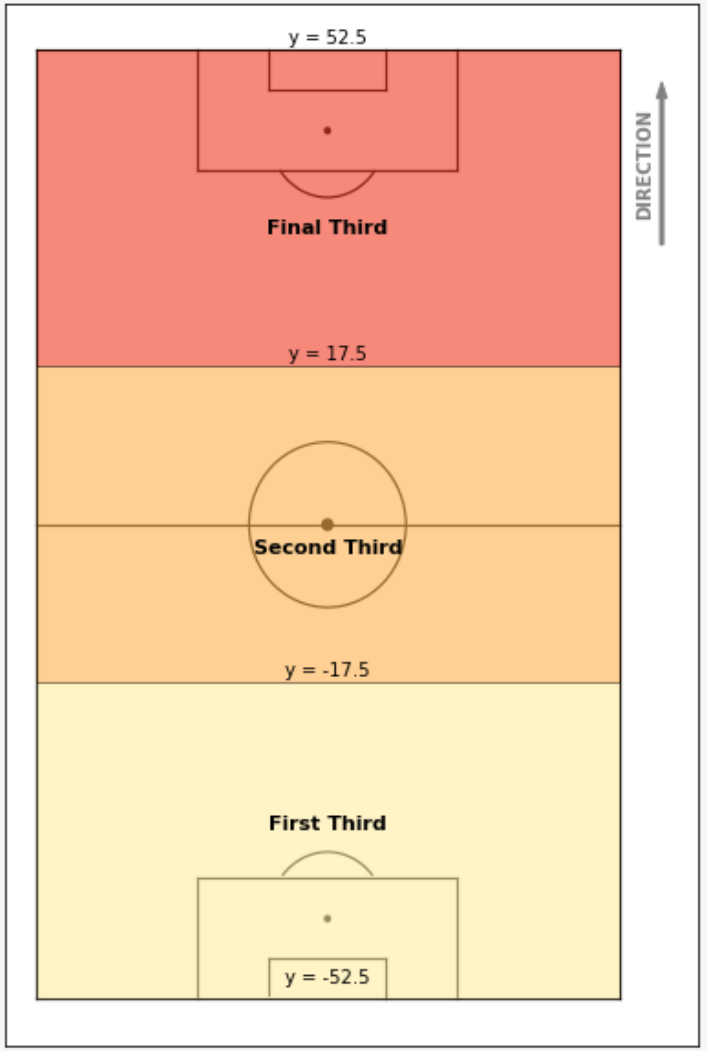
- Classifies players into units in frame-level
- Procedure:
  - The deepest **(D)** and highest **(H)** players, excluding the goalkeeper, are identified.
  - Team Length $(\mathbf{TL}) = \boldsymbol{H_y} - \boldsymbol{D_y}$
  - Unit Length **(UL) =** $\frac{\boldsymbol{TL}}{\boldsymbol{3}}$**,** length of defensive, midfield, and attacking units.
  - Each player is assigned to their corresponding unit based on their spatial location.



Phases of Play

Implementation

- ***SECTION 3: Build up opposed/unopposed***
- Observed in the first third of the pitch according to the location of the ball.
- Zonal Unit Classification is applied to the team out-of-possession to decide whether the build up phase is opposed or not.
- The attacking unit of the defending team needs to be in their final third of the pitch from their perspective.
- Opposed: A defending team player is within 5.36 meters of the ball.
- Unopposed: Otherwise.



Phases of Play

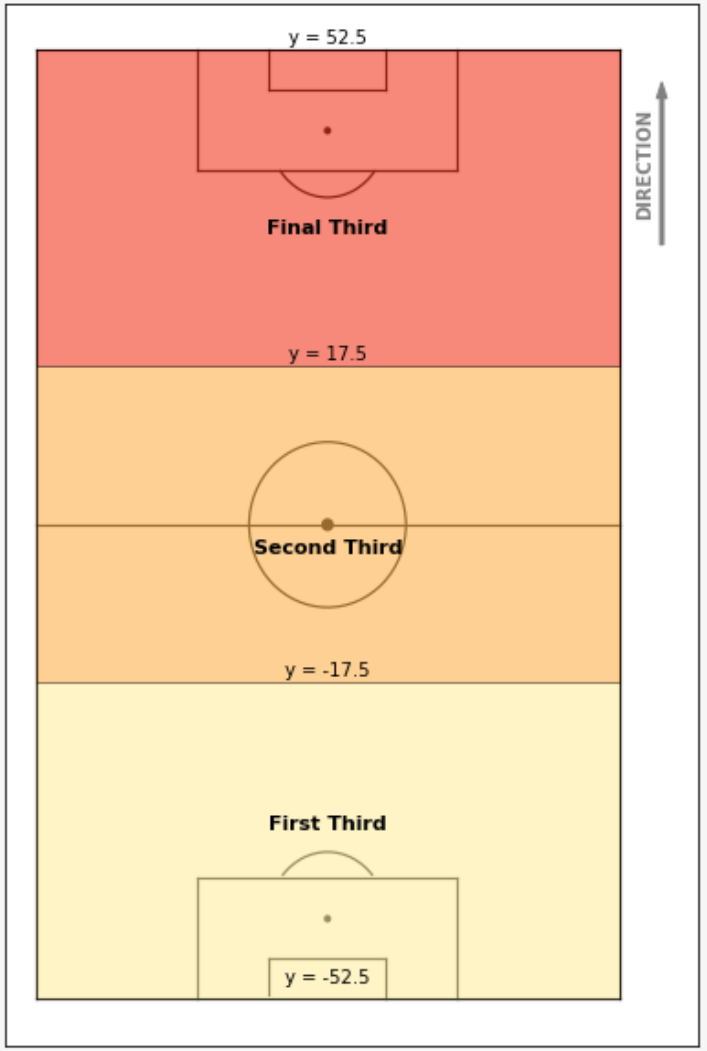
Implementation

- ***SECTION 3: Progression and Final Third***

- Detection of the progression and final third phases are directly related to the location of the ball.

- *Progression:* Ball is in the second third of the pitch.

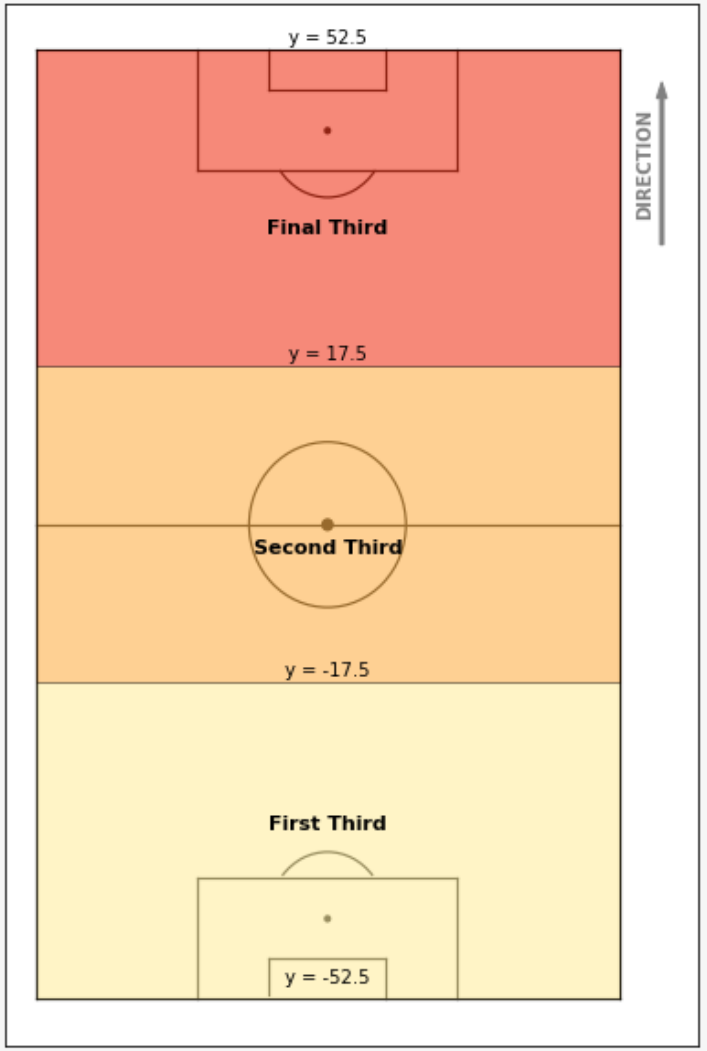
- *Final Third:* Ball is in the final third of the pitch.



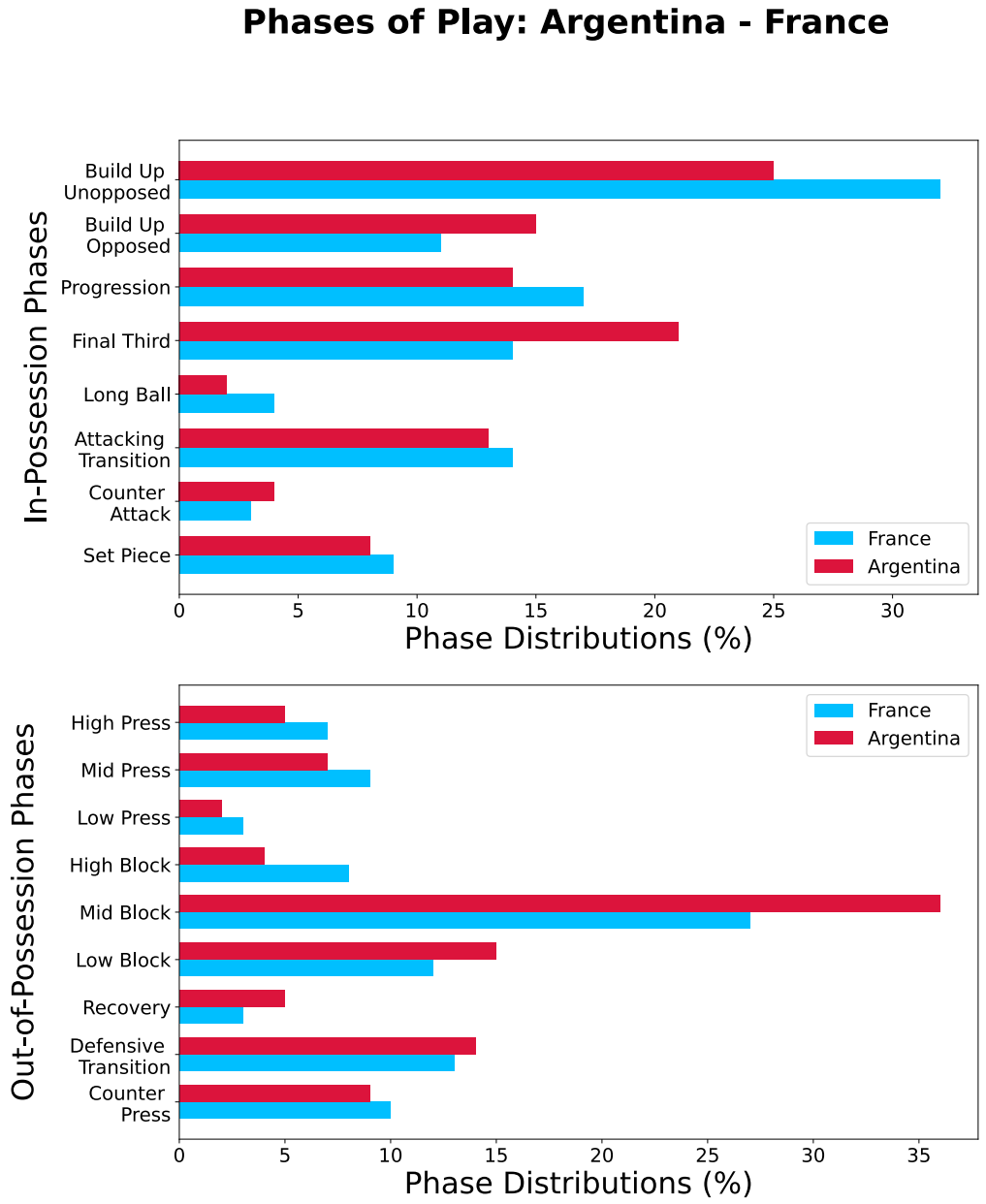
Phases of Play

Implementation

- ***SECTION 3:* *High Block/Press*, *Mid Block/Press*, *Low Block/Press***
- Ball location: **final third**, **second third**, **first third** **of the pitch** (relative to defending team).
- *Press:* A defending player is within **(5.07, 2.28, 1.23) meters** of the ball.
- *Block:*
  - The defending team's attacking unit is in their **final third**, **second third**, **first third** of the pitch.
  - Defending team's length $\leq$ **(37.70, 35.31, 31.85) meters**.



Phases of Play



Ball Recovery Time

EFI Definition

- Average amount of time it takes for a team to regain possession of the ball after losing it.

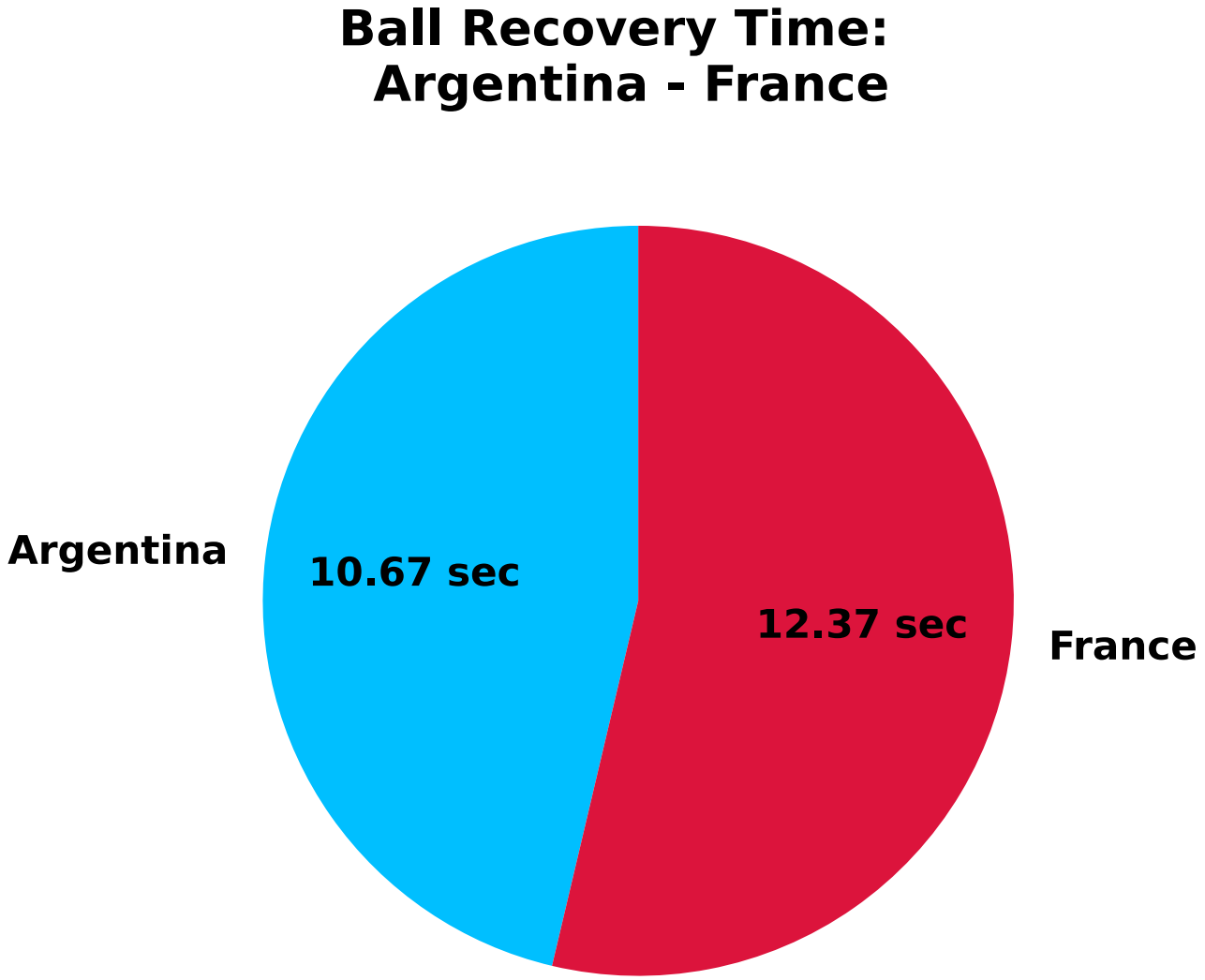
- Time difference between the last ball control of a team in-possession and the first ball control event in the following possession.

Ball Recovery Time

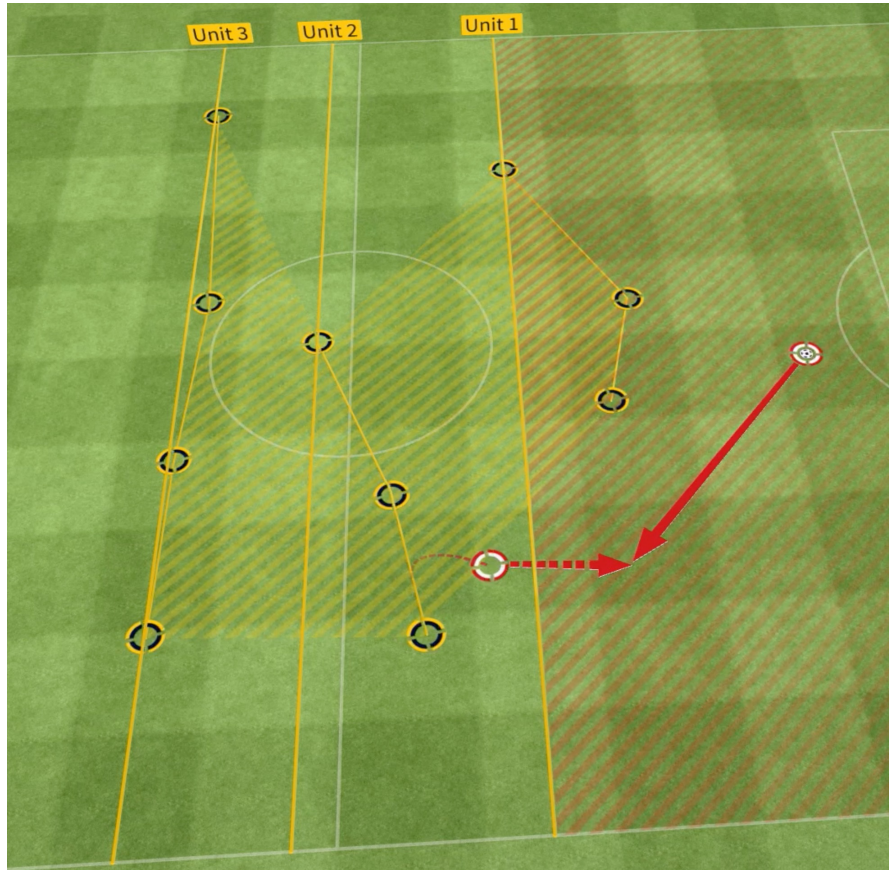
Implementation

- Trace the game when the ball is alive and note the team in-possession for each case.
- For each possession loss by one of the teams, count the number of frames until possession is regained.
- Store the counts for each possession loss scenario for each team individually.
- Compare the counts with “recovery threshold” (24 frames) and discard the values below it.
- Average the remaining values and divide by the sampling rate (25 frames/second) to obtain ball recovery time in seconds.

Ball Recovery Time



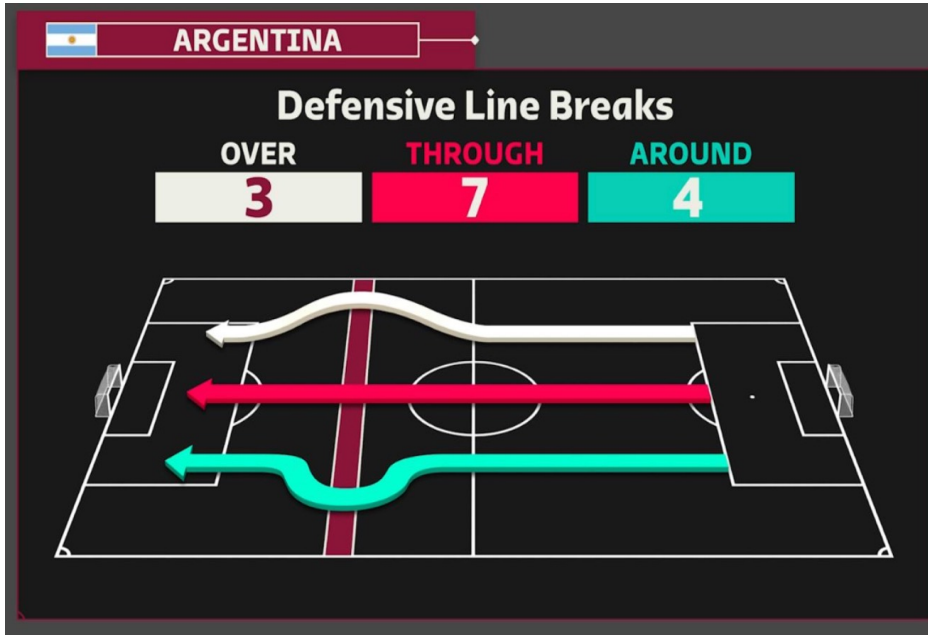
Line Breaks



EFI Definition

- An opposition line is considered broken when the attacking team successfully plays the ball beyond the deepest player in that line.
- Distribution Type
  - Pass
  - Cross
  - Ball Progession
- Direction
  - Through
  - Around
  - Over

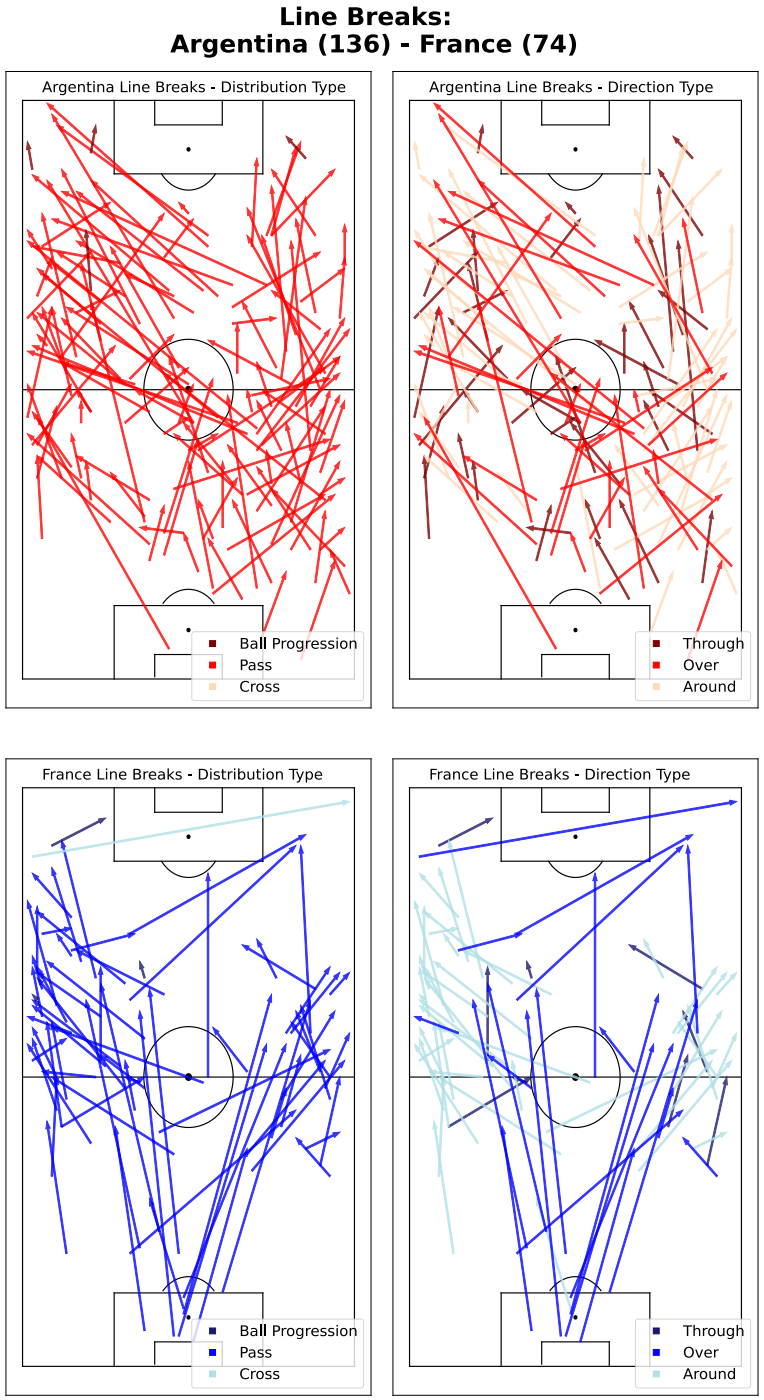
Line Breaks



Implementation

- For each attempt, consider the opposing team's player locations after "frame gap" (54 frames).
- Apply Zonal Unit Classification to determine the lines.
- For the attempts that leads to line breaks:
  - **Distribution Type:** Event type to categorize the attempt as *pass*, *cross* or *ball progression*.
  - **Direction Type:**
    - *Over:* The height of the ball should exceed "height threshold" (1.67 meters) within "over duration" (27 frames) after the attempt is made.
    - *Around:* Compare the ball's x-coordinate with the opposing team's widest players x-coordinates.
    - *Through:* If neither *over* nor *around* applies.

Line Breaks


Line Breaks:
Argentina (136) - France (74)
Argentina Line Breaks - Distribution Type
Ball Progression
Pass
Cross
Argentina Line Breaks - Direction Type
Through
Over
Around
France Line Breaks - Distribution Type
Ball Progression
Pass
Cross
France Line Breaks - Direction Type
Through
Over
Around

Receptions Behind Midfield & Defensive Lines

EFI Definition

- Count of receptions made by each team
  - Between the midfield and defensive unit
  - Behind the defensive unit

- Reception: Successful act of receiving or controlling the ball.

Implementation

- Classify the relative locations of the receptions in relation to the lines constructed by the opposing team.

- Receptions usually occur following a successful pass or cross event.

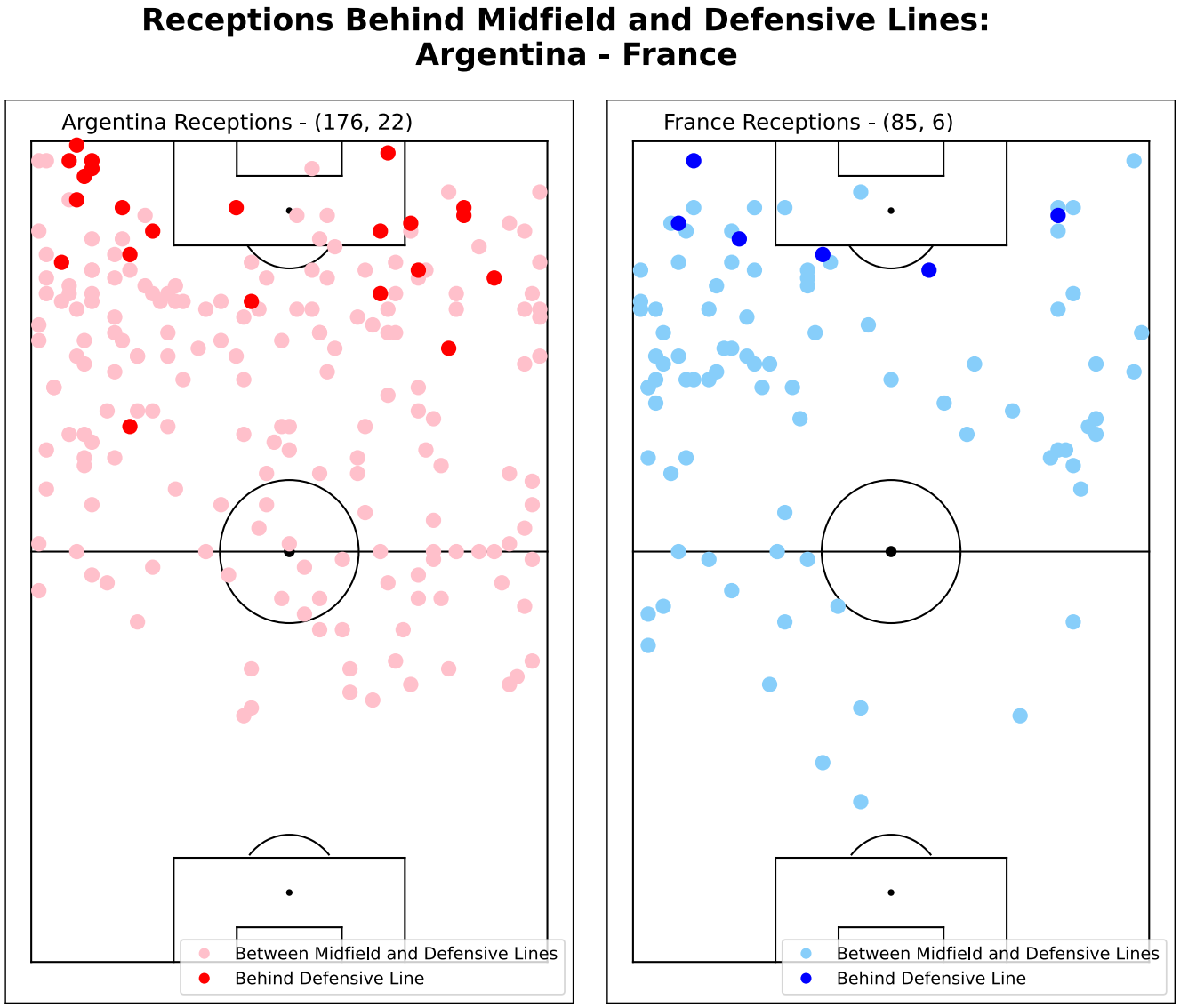
- To ensure that no other events resulting in a reception are overlooked, the reception events in the event data are directly considered instead of analyzing all potential actions that could lead to a reception.

Receptions Behind Midfield & Defensive Lines

Implementation

- Synchronize event data and tracking data.
- Discard any unsuccessful attempt.
- For each reception event consider the opposing team's player locations before "frame gap" (5 frames).
- Apply Zonal Unit Classification to determine the lines.
- Record the frequency of receptions made comparing the spatial location of the reception point with the lines.
  - Between Midfield and Defensive Lines
  - Behind Defensive Lines

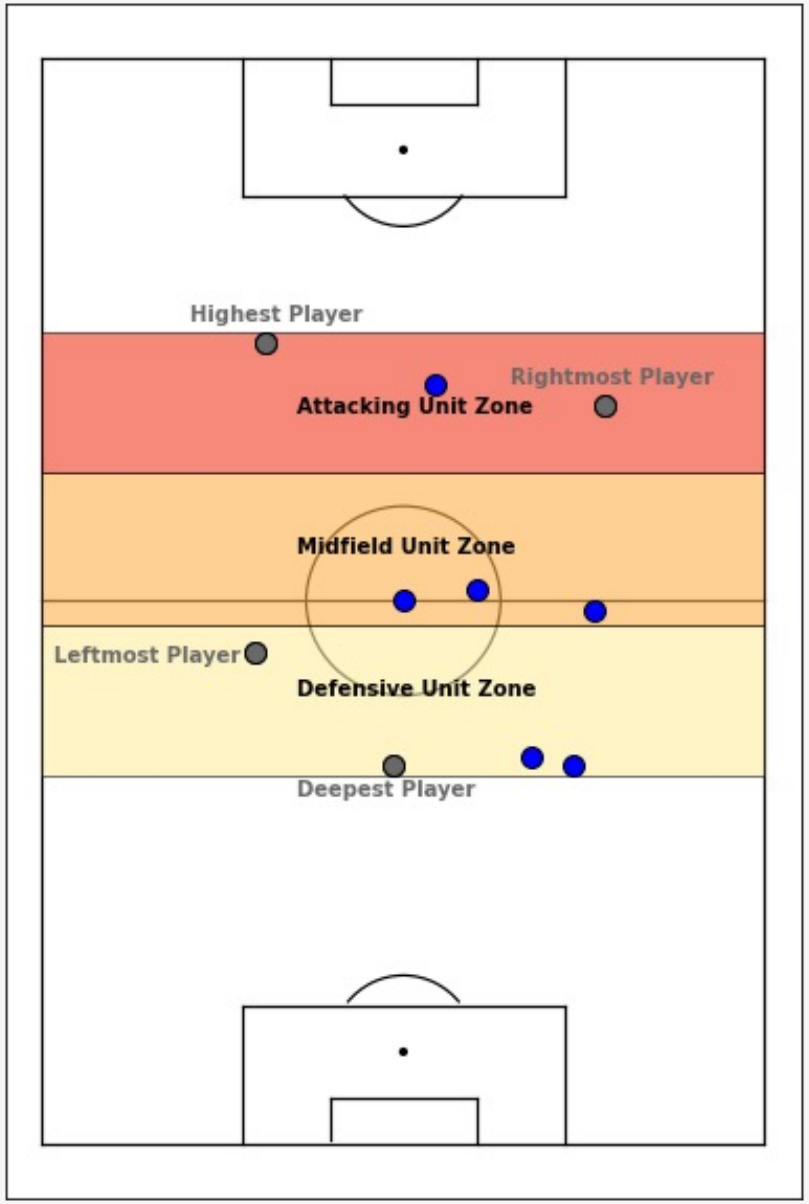
Receptions Behind Midfield & Defensive Lines

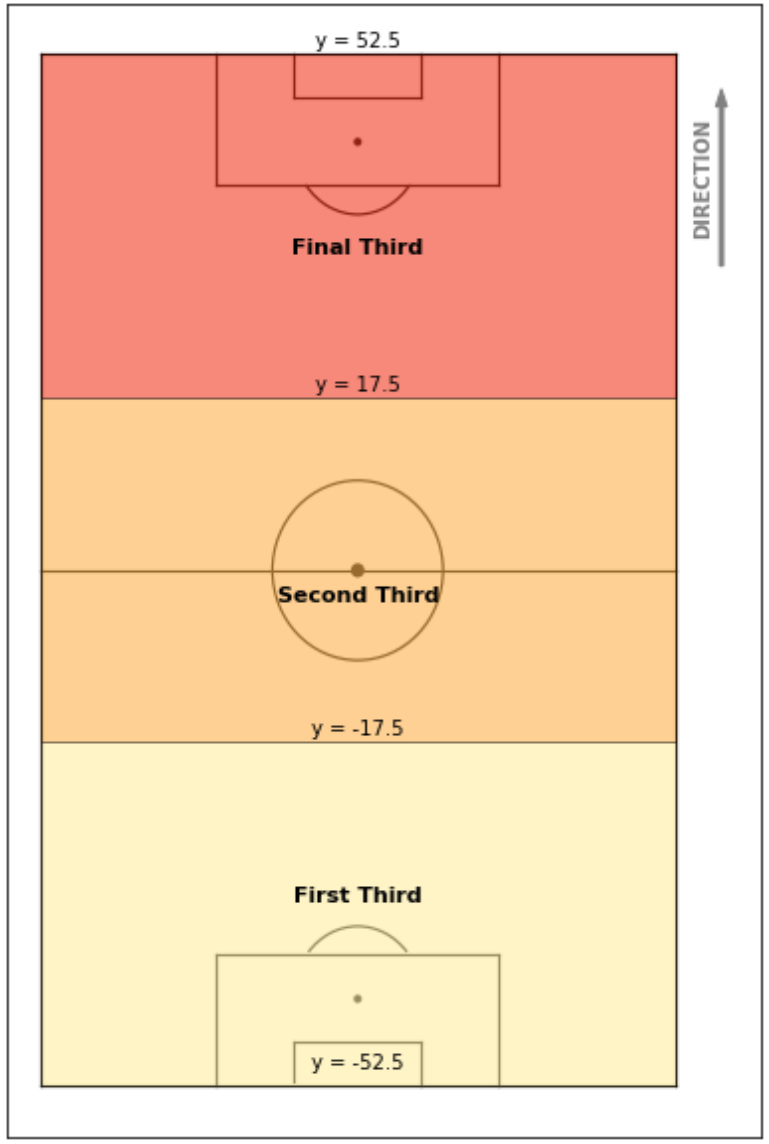


Defensive Line Height & Team Length

EFI Definition

- Line Height = $Deepest\ Player_{y-coordinate} - Goal_{y-coordinate}$
- Team Length = $Highest\ Player_{y-coordinate} - Deepest\ Player_{y-coordinate}$
- Team Width = $Rightmost\ Player_{x-coordinate} - Leftmost\ Player_{x-coordinate}$



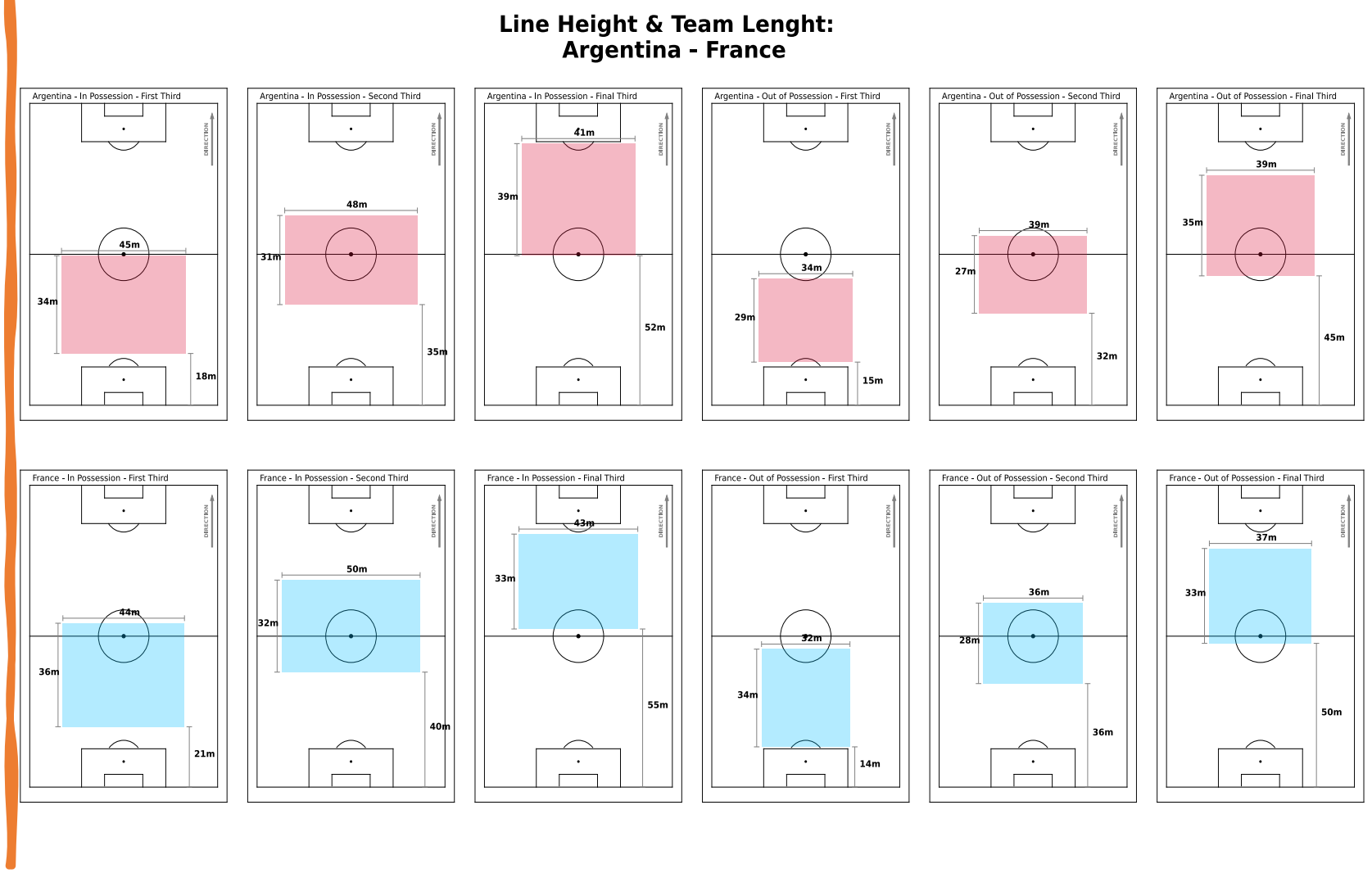


Defensive Line Height & Team Length

Implementation

- Categorize phases into segments according to spatial location of the ball.
- Categorization applies to both in-possession and out-of-possession cases.
  - *In-possession & first third*
  - *In-possession & second third*
  - *In-possession & final third*
  - *Out-of-possession & first third*
  - *Out-of-possession & second third*
  - *Out-of- possession & final third*
- Calculate both teams' **length, width,** and **line height** in frame level, then aggregate.

Defensive Line Height & Team Length



Team Shape

EFI Definition

- Each player is assigned to a specific unit based on their positioning in relation to their teammates.

- When the number of players in each unit is aggregated throughout the match, it forms the overall team shape.
  - While the game is in play.
  - Using the players' spatial locations.

- Provides the shape of the team, both in and out-of-possession cases, as well as the combination of those (i.e., without distinction of ball possession).

Team Shape

Implementation

- Extended Zonal Classification to shape assignment for individual frames:
  - 3- Level Formation: #Defensive Unit (**DU**) – #Midfield Unit (**MU**) - #Attacking Unit (**AU**)
  - 4- Level Formation: #Defensive Unit (**DU**) – #Defensive Midfield Unit (**DMU**) – #Attacking Midfield Unit (**AMU**) – #Attacking Unit (**AU**)

- Initial assignment is done according to 4- level formation using the same approach in zonal classification.

- If **#DMU + #AMU < 5** then convert to 3- Level Formation:
  - **MU = DMU + AMU**

Team Shape

- The aim is to capture the popular 4- level formations such as:
  - 4-1-4-1
  - 4-2-3-1
  - 3-4-1-2

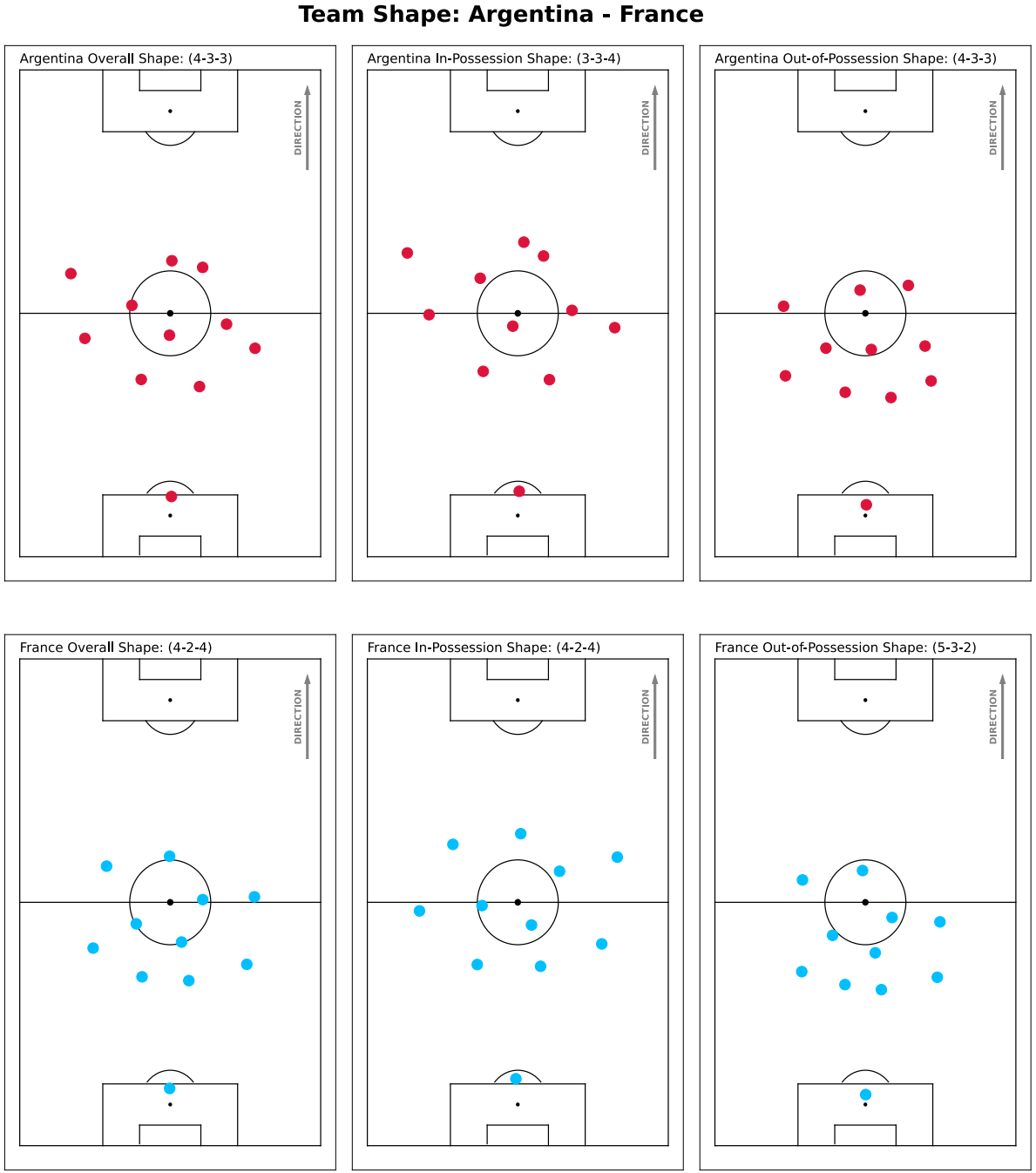
Team Shape

Implementation

- Majority count for aggregation:
  - To capture the team's shape at the frame-level.
  - Avoid overlooking strategies by averaging over the entire match.

- For each scenario (i.e., in-possession, out-of-possession, and overall) the formation with the highest frequency is selected.

Team Shape


Team Shape: Argentina - France
Argentina Overall Shape: (4-3-3)
Argentina In-Possession Shape: (3-3-4)
Argentina Out-of-Possession Shape: (4-3-3)
France Overall Shape: (4-2-4)
France In-Possession Shape: (4-2-4)
France Out-of-Possession Shape: (5-3-2)
DIRECTION

Final Third Entries

EFI Document

- Measures the amount and location of final third entries split across five different entry zones:
  - *Left channel*
  - *Left inside channel*
  - *Central channel*
  - *Right inside channel*
  - *Right channel*

- Final third entries being credited to a team if the ball is successfully **distributed** or **carried** into the final third.
- Entries starting outside the final third area and ending within it are counted.

Final Third Entries

Implementation

- Distribution: pass and cross events.
- Synchronize the event data with tracking data (i.e., time conversion).
- Add “pass reception duration” (14 frames) to each distribution's converted frames to estimate when the pass or cross ends.
- Based on end locations, store approximate end frame of each distribution to correct channel.

Final Third Entries

Implementation

- Carrying: Ball progression

- The entry is determined based on the location of the ball, specifically requiring the ball to be within the final third zone.

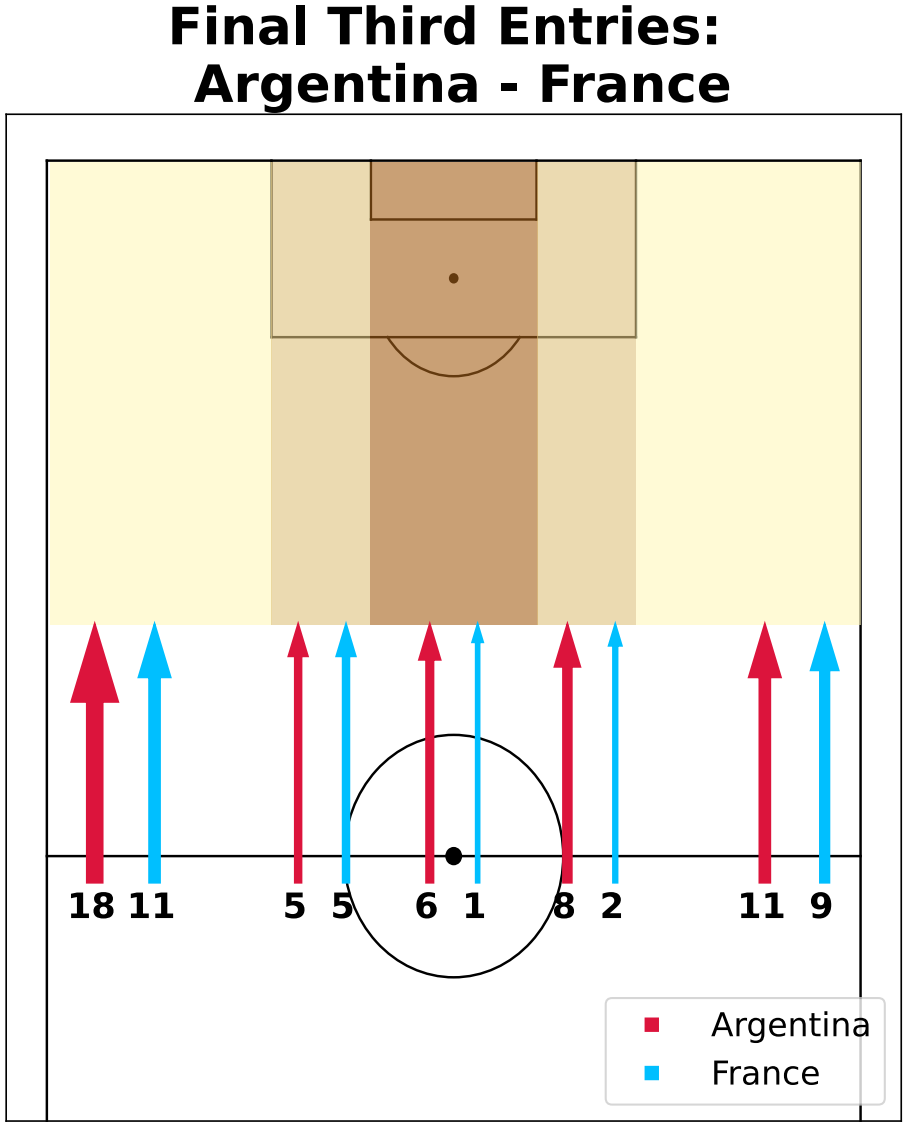
- Requires a more detailed pre-process due to the limited features compared to event data.
  - The game must be in play.
  - The team in-possession should maintain possession for a certain duration (“possession threshold”) after the entry (i.e., 107 frames).
  - To ensure ball control by the attacking team, a circle with a 1.84-meter radius is used around the ball. If an attacking player is inside this circle, they are considered to have control of the ball.
- According to the entry locations, store approximate frame of each carrying to correct channel.

Final Third Entries

Implementation

- Counting
- Use the frames stored for each channel from the distribution (pass, cross) and carrying (ball progression) events.
- Introduce the parameter “out of final third threshold” (128 frames) to determine the gap necessary to distinguish two separate entries.

Final Third Entries

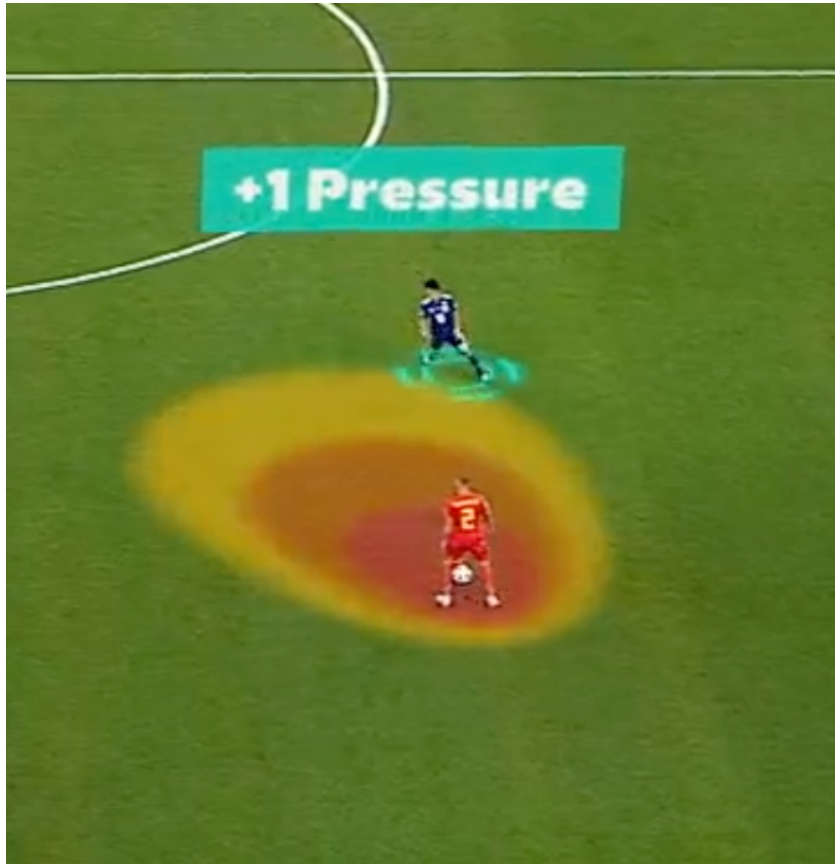


Pressure on the Ball

EFI Document

- Defensive action of a player, reducing the distance between themselves and the opponent in-possession of the ball.

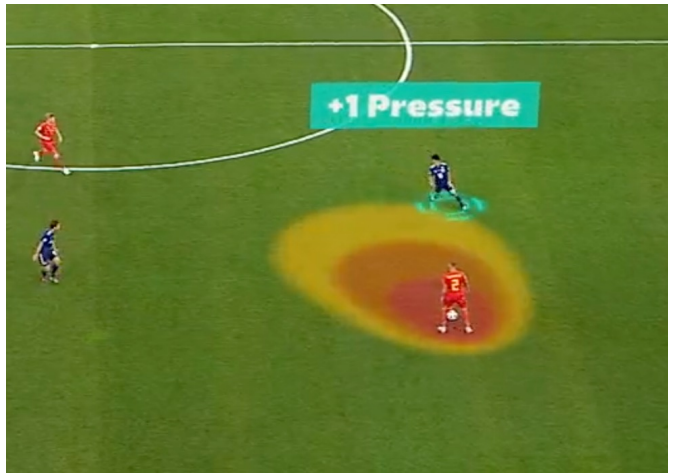
- Following information are considered:
  - Defender’s distance to the ball.
  - The angles of the defender to the ball carrier.
  - Proximity of defenders towards the ball carrier.



Pressure on the Ball

Implementation

- Pressure Detection
- The geometrical relationship between the attacker (**A**), defender (**D**), and the ball (**B**) is examined.
- Key factors to include:
  - Edge (**AD**)
  - Edge (**AB**)
  - Angle ($\boldsymbol{\theta}$) between the edges **AD** and **AB**
  - Distance ($\boldsymbol{d}$) between **D** and **B**
- There are four cases that classify a frame as pressure presence:
  - $0^{\circ} \leq \boldsymbol{\theta} \leq 30^{\circ}$ and $\boldsymbol{d} \leq 3\ meters$
  - $30^{\circ} \leq \boldsymbol{\theta} \leq 60^{\circ}$ and $\boldsymbol{d} \leq 2.5\ meters$
  - $60^{\circ} \leq \boldsymbol{\theta} \leq 90^{\circ}$ and $\boldsymbol{d} \leq 2\ meters$
  - $90^{\circ} \leq \boldsymbol{\theta}$ and $\boldsymbol{d} \leq 1\ meters$



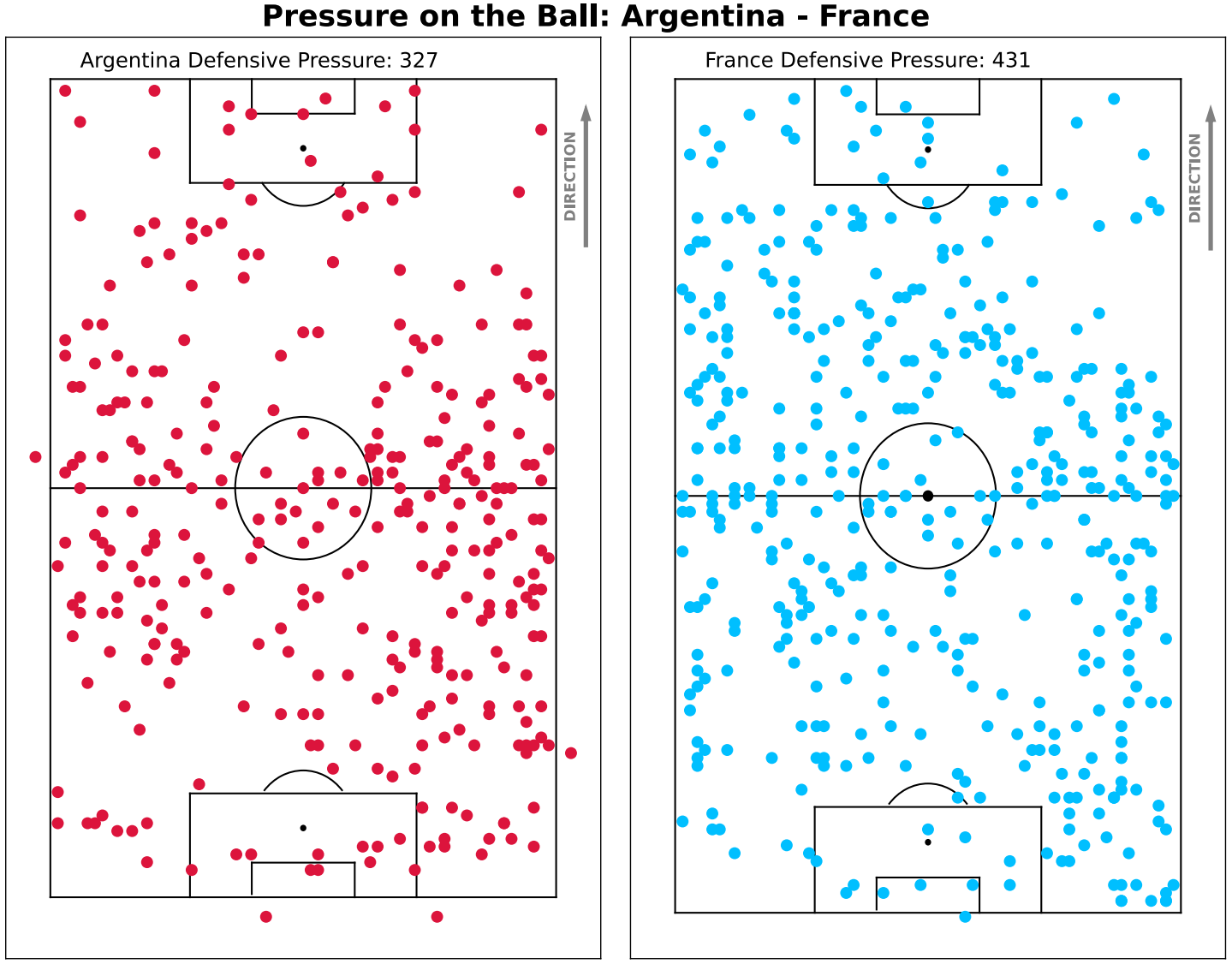
Pressure on the Ball

Implementation

- Counting

- Two parameters are introduced to count each group of consecutive pressure frames individually.
  - “pressure threshold” (27 frames): minimum gap required between groups of consecutive pressure frames.
  - “continuous pressure threshold” (5 frames): minimum group size of consecutive pressure frames.

Pressure on the Ball



Forced Turnovers

EFI Document

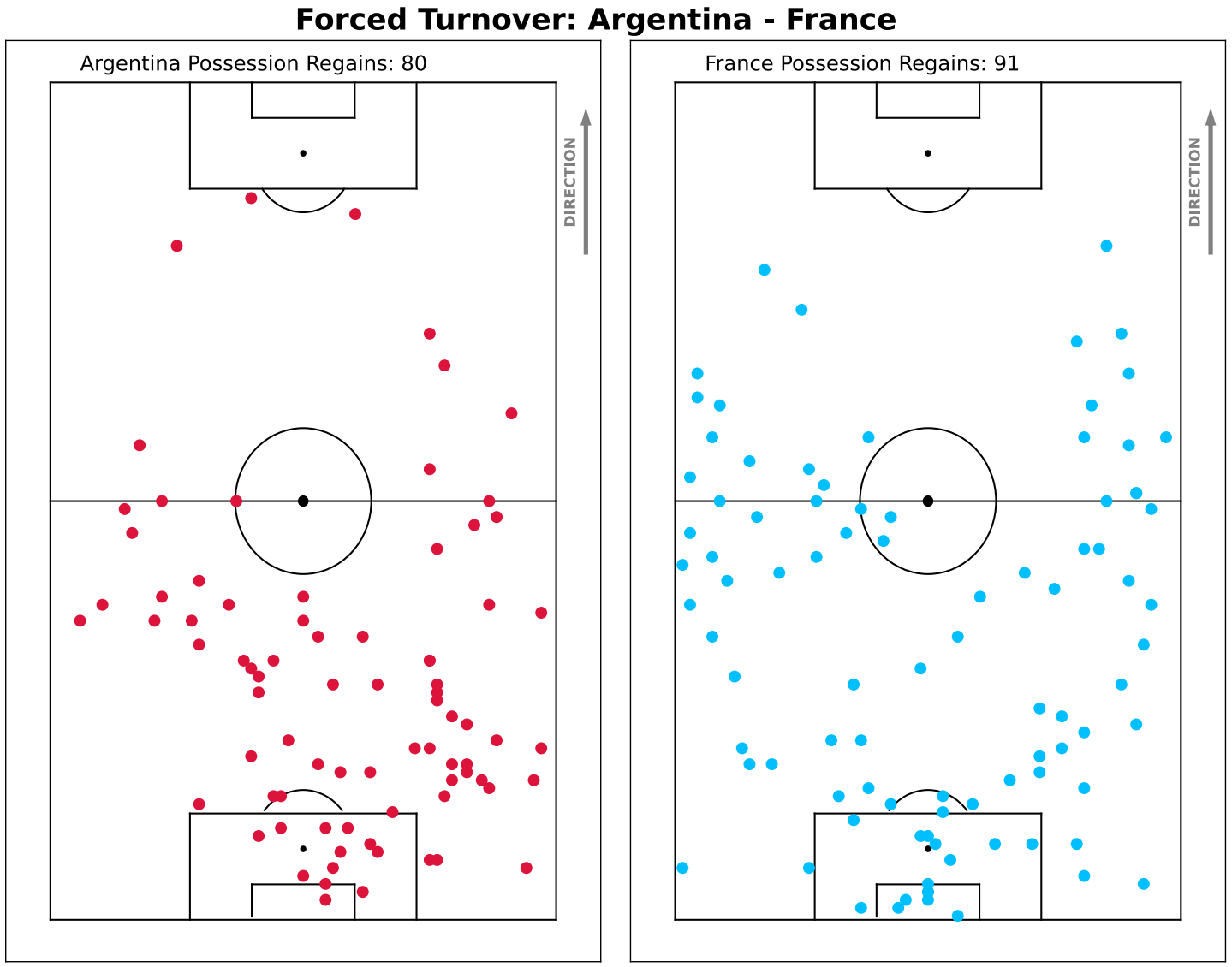
- Forced turnovers is a defensive metric awarded to the defending team.
- Considers the events when the attacking team lose possession of the ball due to pressure being applied by the defending team.
- A team is credited with a forced turnover if they exert pressure on the ball and then gains possession of the ball.

Forced Turnovers

Implementation

- Based on the outcomes observed from the pressure on the ball (i.e., indices of pressure frames)
- Check for potential pressure event whenever possession information changes.
- If the detected turnover is due to pressure on the ball, verify the possession of attacking team.
- Examine the consecutive possession frames of the attacking team and compare them to the “possession threshold” (97 frames).
  - If consecutive possession frames drop below the threshold, it implies the attacking team wasn't truly in possession, and the turnover isn't considered forced.
  - Otherwise, the turnover is recorded as forced.

Forced Turnovers



Expected Goal (xG)

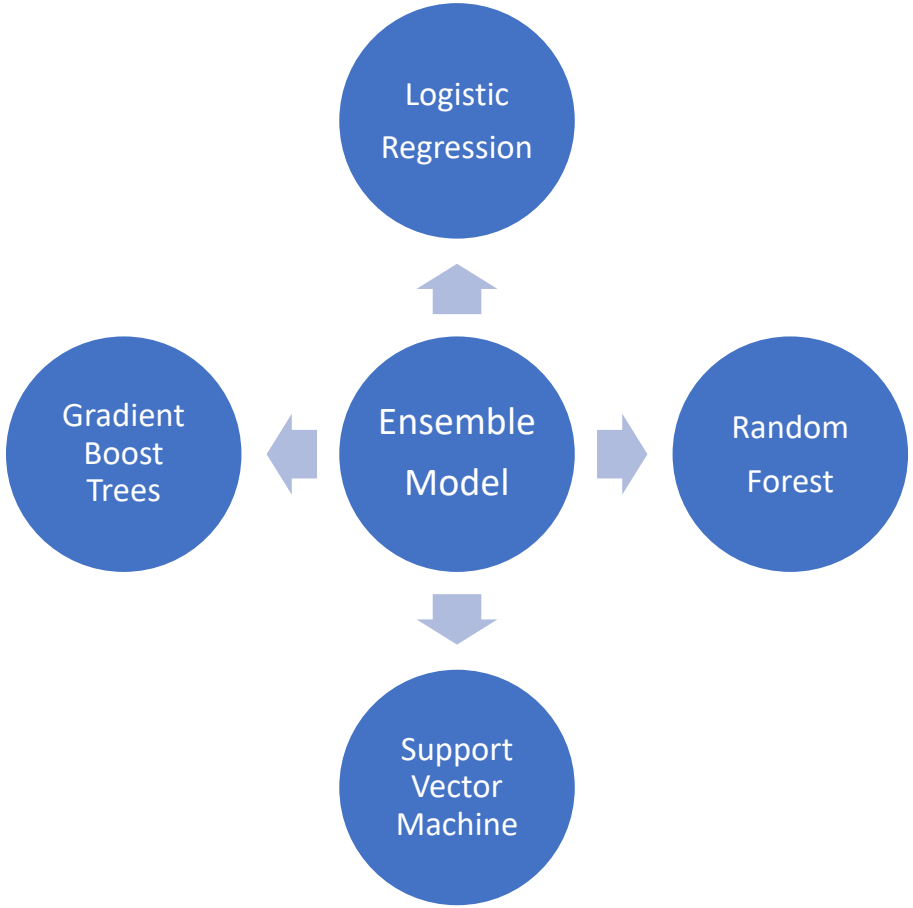
EFI Document

- Probability of scoring from a given attempt at goal, based on a statistical model developed over a historical database of shots.
- Dependent on several factors from before the attempt at goal was taken.
- The major features are:
  - Distance
  - Angle
  - Number of players obstructing the goalmouth
  - Pressure on the shooter
  - Whether the attempt is with the head, foot or body
  - The goalkeeper position at the time of the attempt

Expected Goal (xG)

Implementation

- Ensemble Model: Combination of the outputs of individual machine learning models.
- Having different angles to the same concept increases the robustness of the results.

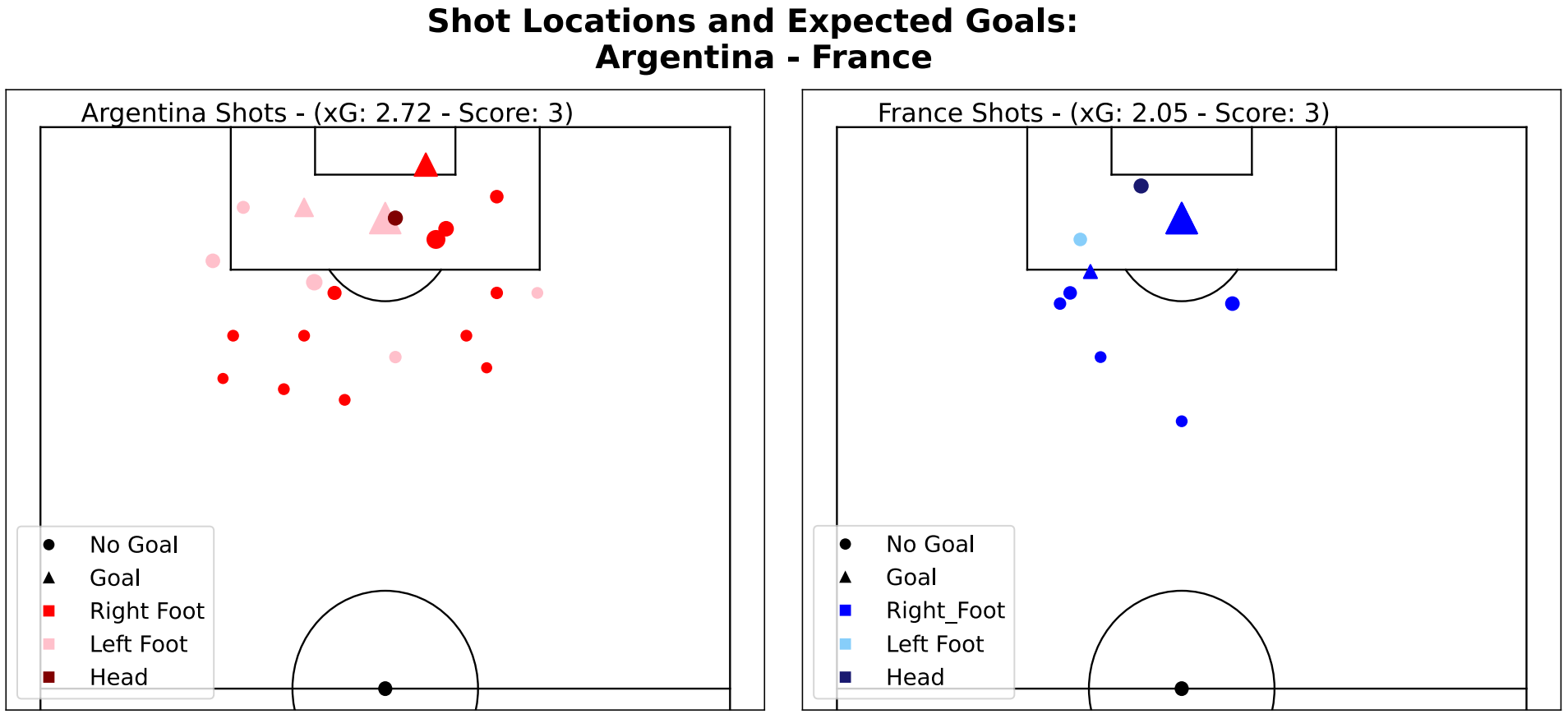


Expected Goal (xG)

Implementation

- To calculate xG values for shots in a match (test set), the remaining matches are utilized as the training set.
- Features in the dataset:
  - Shot location
  - Pressure presence information
  - Body type used to take the shot
  - Origin (A special indicator of the location where the recorded event took place, such as a penalty or corner)
  - Distance to goal
  - Angle
  - Goalkeeper location
  - Number of players obstructing the goalmouth
- Goal indicator of each shot forms the label.
- k-fold cross validation to optimize parameters of each model.
- Calculate the accuracy of each model's predictions using the provided test labels.
- These accuracy values are used as weights for each model's probability outcomes to determine the final ensemble probability result.

Expected Goal (xG)



Performance & Validation

- The performance of the concepts is analyzed in conjunction with the FIFA results.
- Comparisons are made for the best, median, and worst performing matches in each concept.

Performance & Validation

- Mean Absolute Percentage Error (**MAPE**) and Mean Squared Error (**MSE**) are the two metrics used to optimize hyperparameters and assess the performance using FIFA reports.
- N= Total Number of Observations, y_i= Actual Value, $\hat{y}_i$ = Predicted Value
- $MAPE(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i - \hat{y}_i}{y_i}$,
- $MSE(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
- MAPE is chosen because it is scale invariant, easily interpretable, and focuses on relative errors rather than absolute errors. It can be replaced by MSE when actual values are zero.
- MSE is a suitable alternative as it is a continuous and differentiable function, making it well-suited for optimization algorithms like Optuna.

Performance & Validation

- **OPTUNA**

- Automated machine learning hyperparameter optimization software framework.

- Imperative, define-by-run style user API, allowing for modular code and dynamic construction of hyperparameter search spaces.

- Requires:
  - Loss function to be minimized (e.g., MAPE or MSE)
  - Parameters to be tuned with their search space
  - Method that provides the predictions
  - True values (FIFA reports)

Performance & Validation

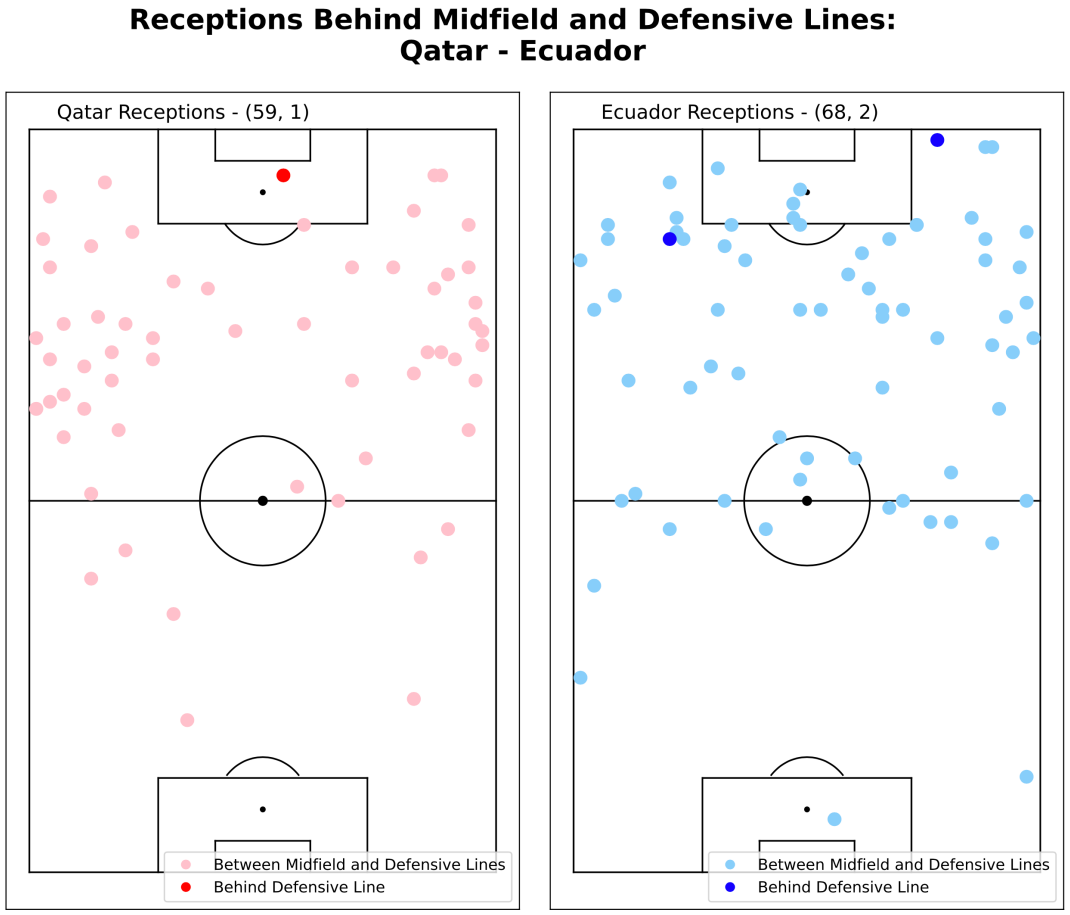
MAPE	Best	Median	Worst
Possession Control	0.10	0.17	0.37
Ball Recovery Time	**0.02**	0.13	0.39
Line Breaks	0.07	0.16	0.33
Receptions	**0.02**	0.14	**0.46**
Line Height & Team Length	0.04	0.07	0.09
Pressure on the Ball	0.04	0.13	0.45
Forced Turnovers	**0.02**	0.10	0.28

MSE	Best	Median	Worst
Final Third Entries	0.90	6.10	21.00

Team Shape & Expected Goal: Do not have certain true values.

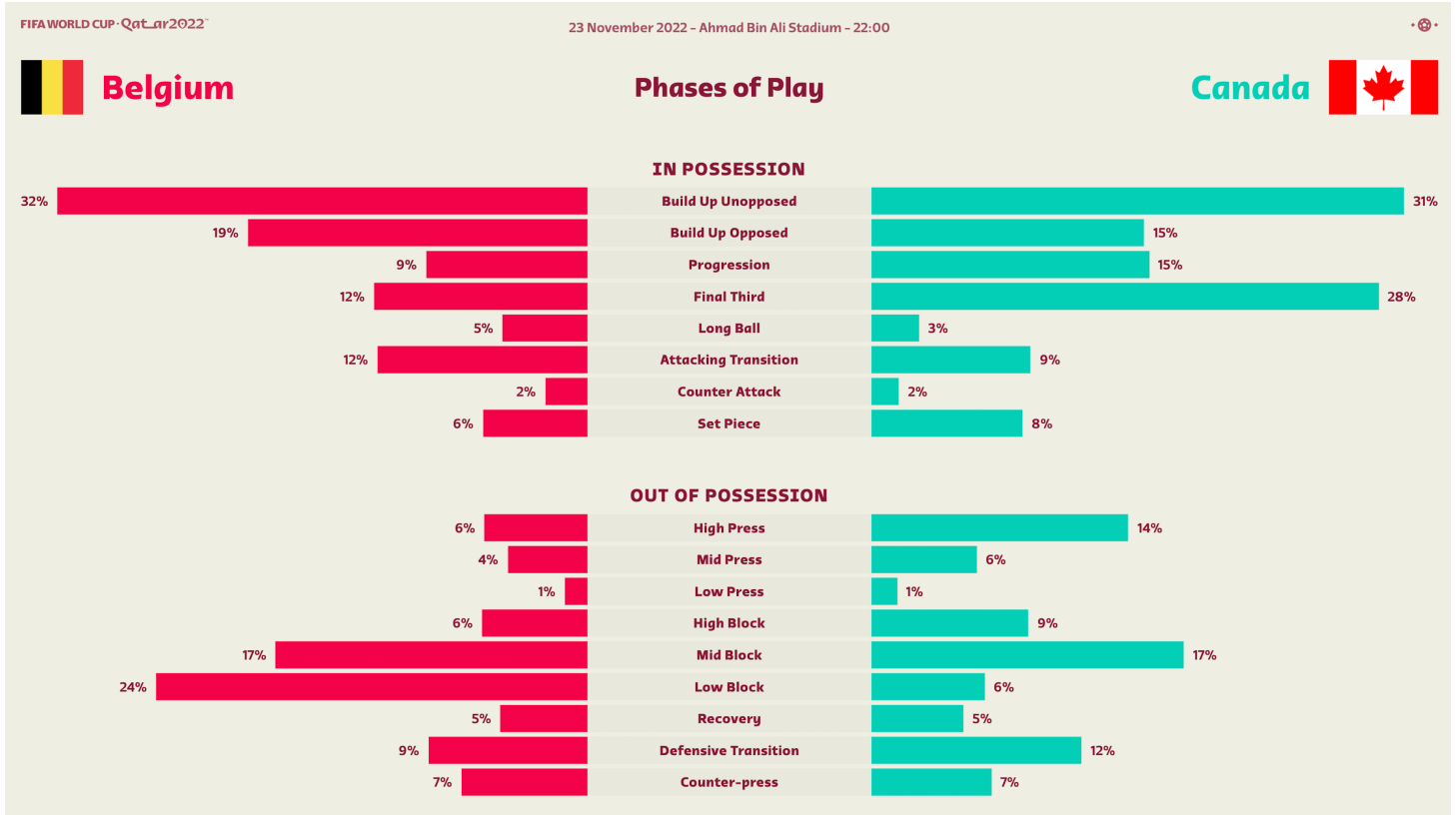
Phases of Play: Includes refinements hence not comparable.

Receptions Behind Midfield & Defensive Lines
Worst Performance



FIFA results	Between midfield & defensive lines	Behind defensive lines
Qatar	52	5
Ecuador	58	8

Phases of Play



FIFA's Sum of Phases (%)	In-Possession	Out-of-Possession
Belgium	97%	79%
Canada	111%	77%

Limitations & Futurework

- Lack of implementation details
- Heuristics
- Dataset size
- Data syncronization

Ex: FIFA's way of detecting mid-block phase

- Players are not moving forward or backward at a speed exceeding 10 km/h.
- Players (excluding the highest positioned player) forming horizontal lines and maintaining steadiness for at least 1.6 seconds within the previous 2.0 seconds .
- There should be at least 8 outfield defending players positioned in the middle third of the pitch .
- The end of the phase is detected when the block criteria are not met for 0.4 seconds within a 2.0-second interval.

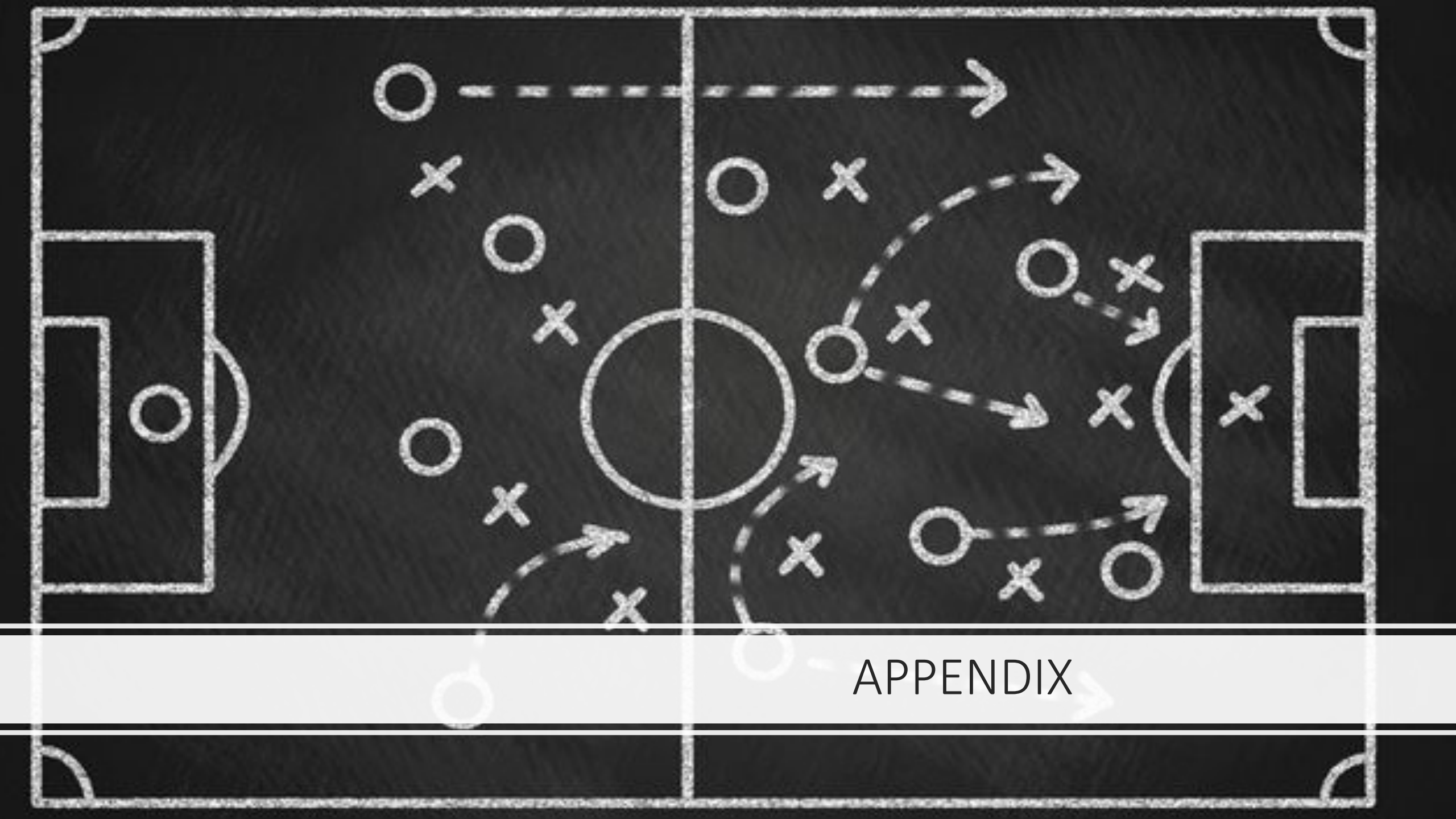
Conclusion

- Concepts from FIFA's Enhanced Football Intelligence document were implemented in an open-source library, following the provided definitions and incorporating specific parameters for refinements.

- FIFA reports were used as labels for parameter tuning and assessment, with heuristic-free or less heuristic-involved concepts yielding the most stable results.

- Overall, median performances across all concepts closely approached the best performance, indicating a satisfactory level of performance consistent with FIFA's outcomes.

Q&A



APPENDIX

Match Data

- Fundamental details about the game and the participating teams.
- In terms of the game, the data encompasses the match ID, start and end times of each phase.
- At the team level, the data provides the IDs and names of both teams, the lineups for each team are included.

Event Data

- Contains all the events that took place during a game.
- Incidents were meticulously recorded by teams of dedicated analysts, with one analyst assigned per player, to cover and document the events involving or completed by that player, partly to facilitate the computation of the concepts.
- The compilation of these events from each analyst results in the event data, which presents a chronological sequence of incidents that took place throughout the game.
- In general, information about the events includes the time and location of occurrence, ID of the player(s) involved in the action, and the outcome .

Tracking Data

- “ChyronHego's TRACAB system”.

- Frame: observation at a specific moment in the game with a sampling rate of 25 frames per second.
  - Frame number
  - Team ID
  - Player jersey numbers
  - Player and ball speeds
  - Player and ball coordinates on the pitch
  - Ball state
  - The team in-possession

Time Conversion

- Match run time, denoted as t_{match} represents the elapsed time in milliseconds since the start of the match. With a sampling rate of 25 frames per second, each frame corresponds to a duration of 40 milliseconds.
  - $\text{frame level time} = \frac{t_{match}}{40}$
- By subtracting the frames corresponding to halftime intervals, denoted as $f_{halftime}$, the approximate start frame of the events can be determined.
  - start frame of events = frame level time – $f_{halftime}$

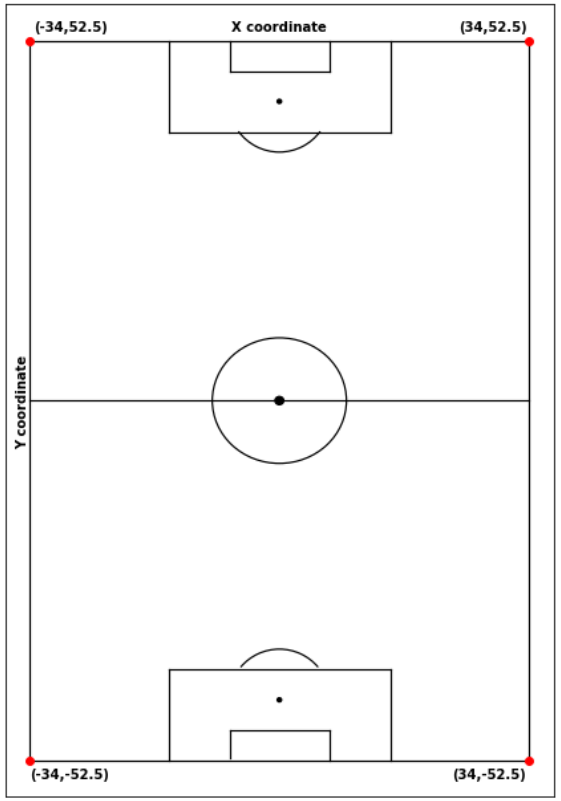
Coordinate System Conversion

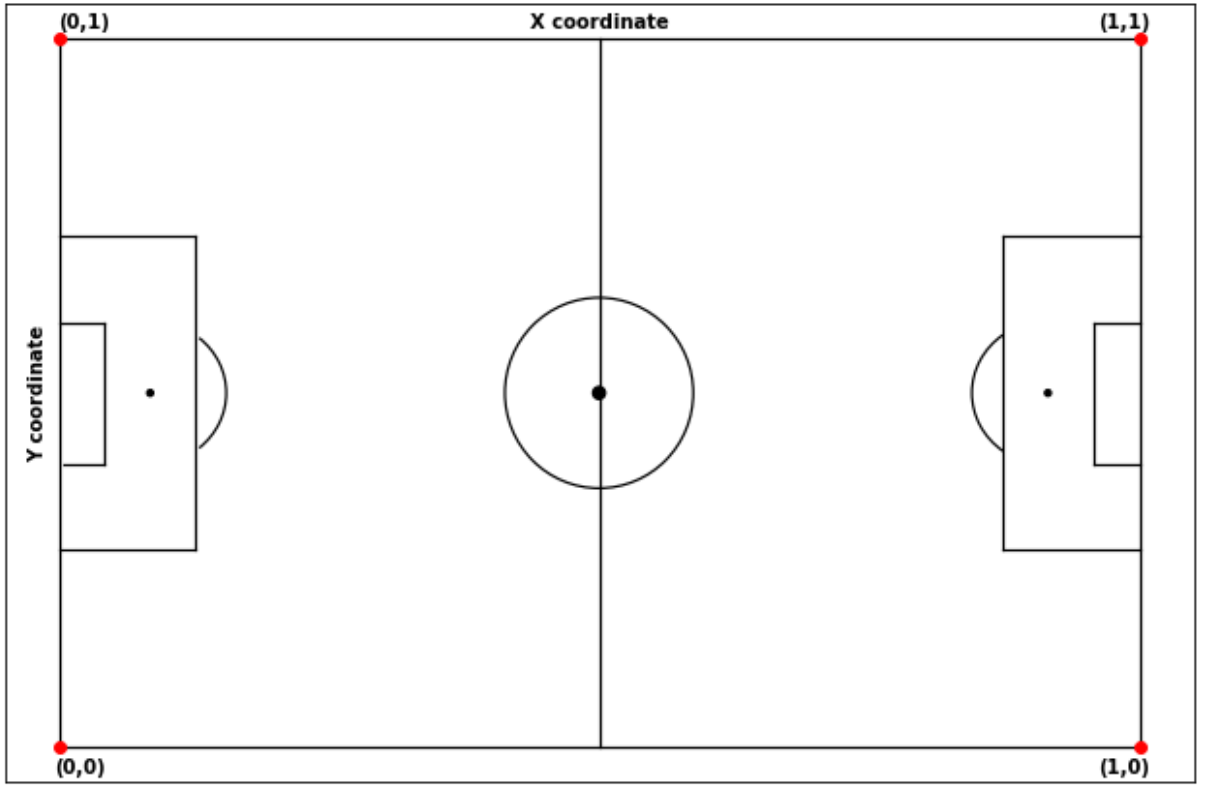
- The event data and tracking data coordinate systems differ from each other in several aspects. They have distinct centers, dimensions defined within different intervals, and varying alignments.

  - $x_{tracking}$ = (yevent $_{- 0.5) * 6}8$
  - $y_{tracking} = - (x_{event} - 0.5) * 105$

Coordinate Systems

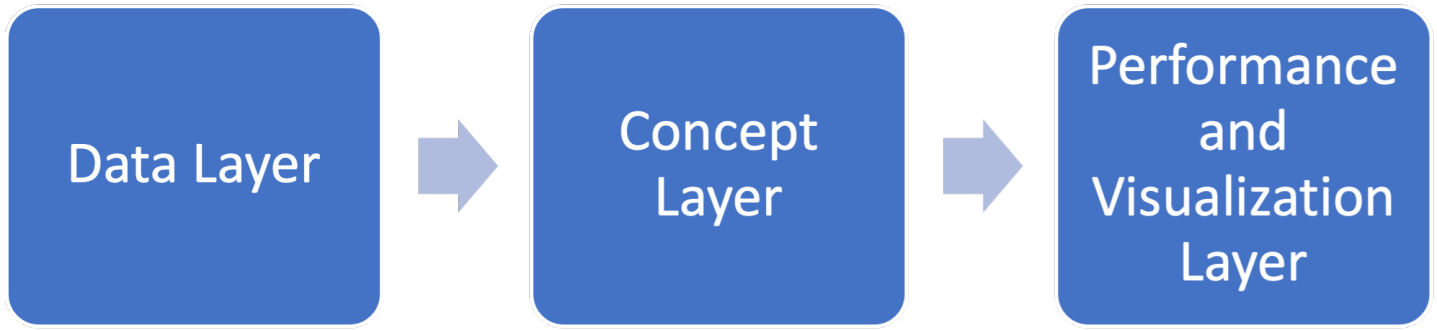
- Tracking data and event data coordinate systems





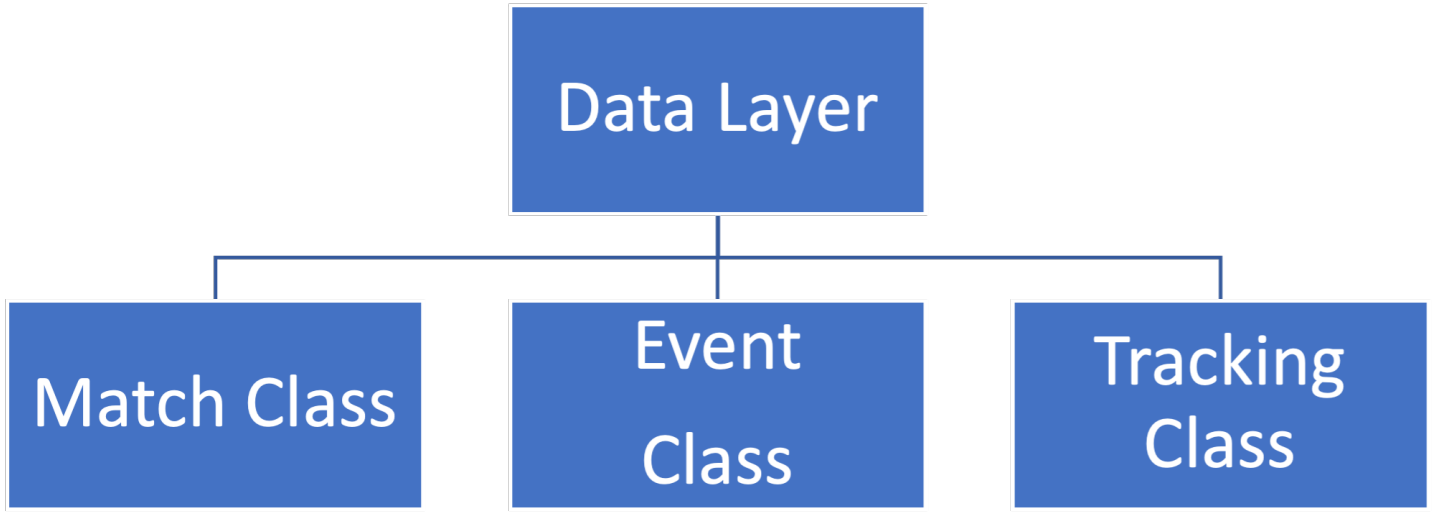
Library Design

- High Level Overview



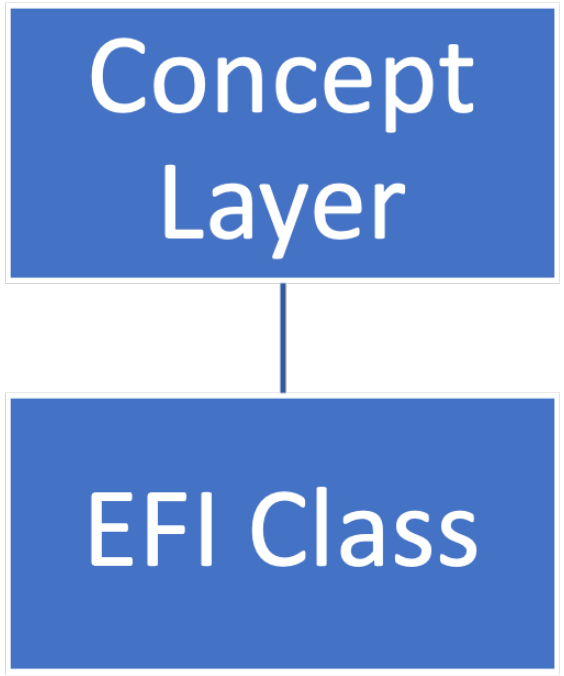
Library Design

- Data Layer
- Initialize match, event, and tracking objects.



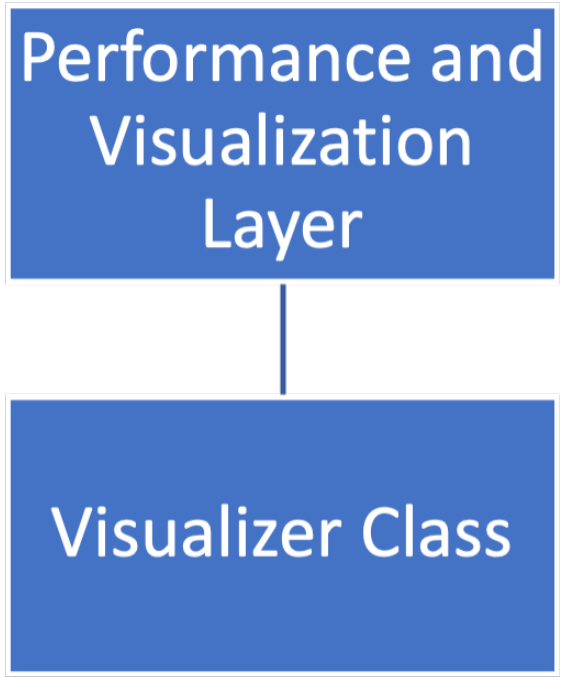
Library Design

- Concept Layer
- Using the Data Layer Objects, initialize the EFI object.
- Employ EFI object to generate outputs of the desired concepts.



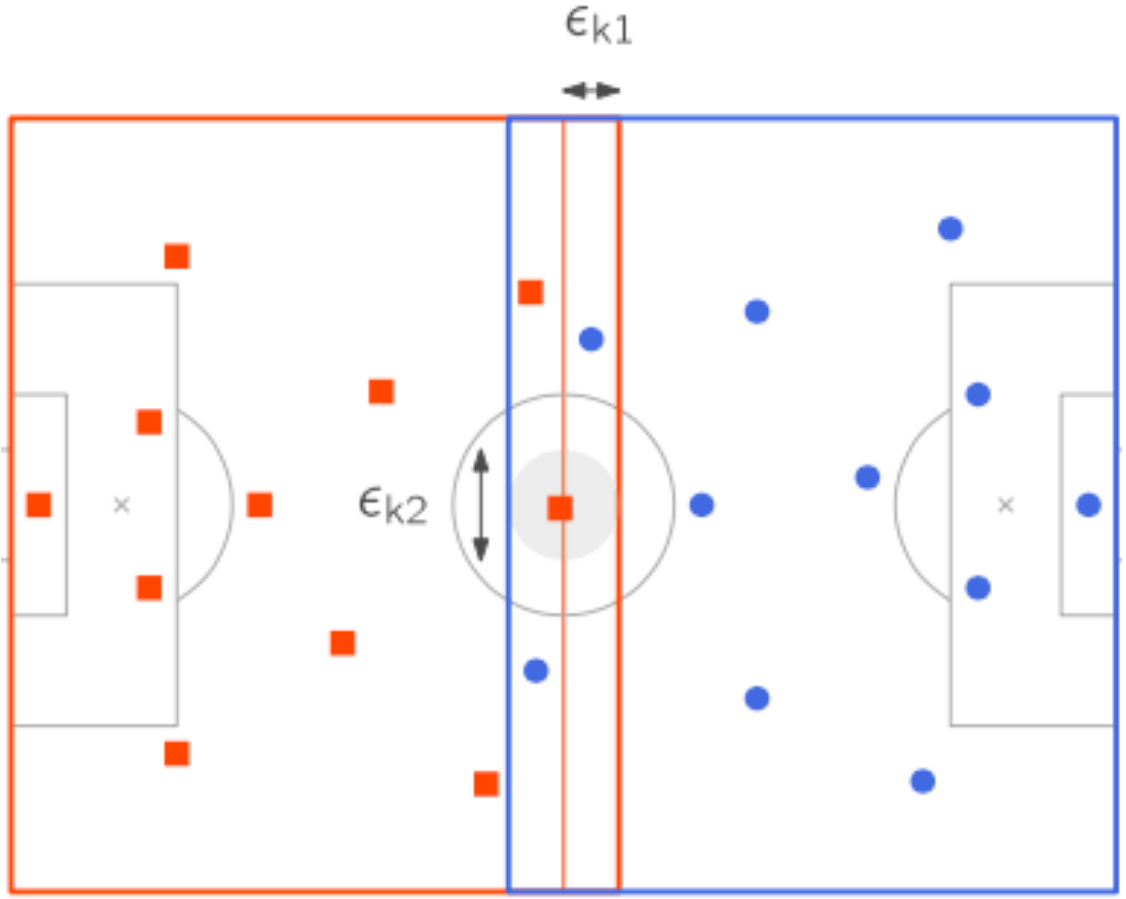
Library Design

- Performance and Visualization Layer
- Initialize the Visualization object.
- Display outputs with visual representations that are comparable with the ones in FIFA reports.



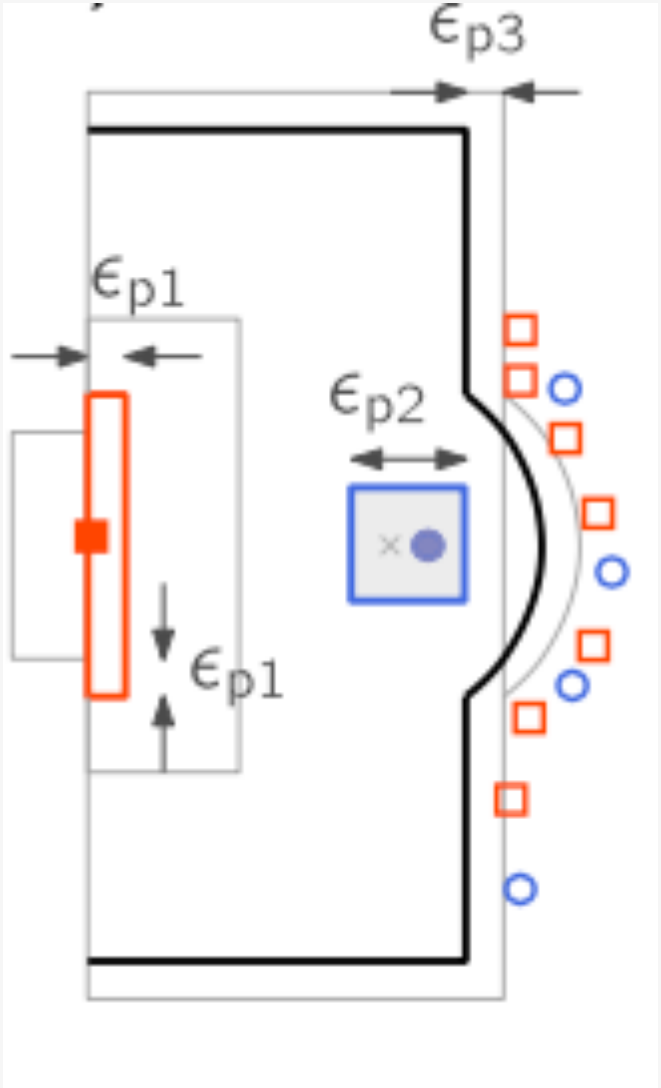
Kick-Off

- All players should be within their own halves, with a tolerance of ϵ_{k1}(4 meters).
- There should be at least one player within ϵ_{k2}(12 meters) of the center mark.



Penalty

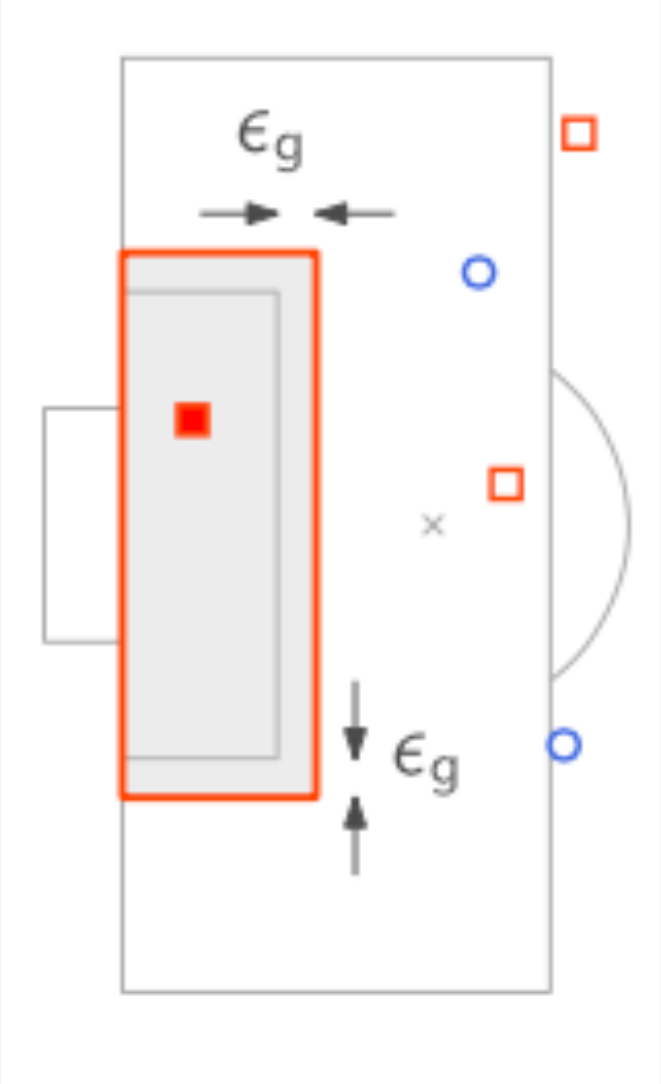
- Only one player is at their goal line between the posts (with a tolerance bounding box of ϵ_{p1}, i.e., 3 meters).
- Only one opponent is within a square bounding box from $\frac{\epsilon_{p2}}{4}$ (1 meter) in front of $\frac{3\epsilon_{p2}}{4}$ (3 meters) behind the active penalty mark.
- Other players are neither within the penalty area nor within 9.15 meters from the penalty mark (with a tolerance of εp3, i.e., 4 meters).



Goal-Kick

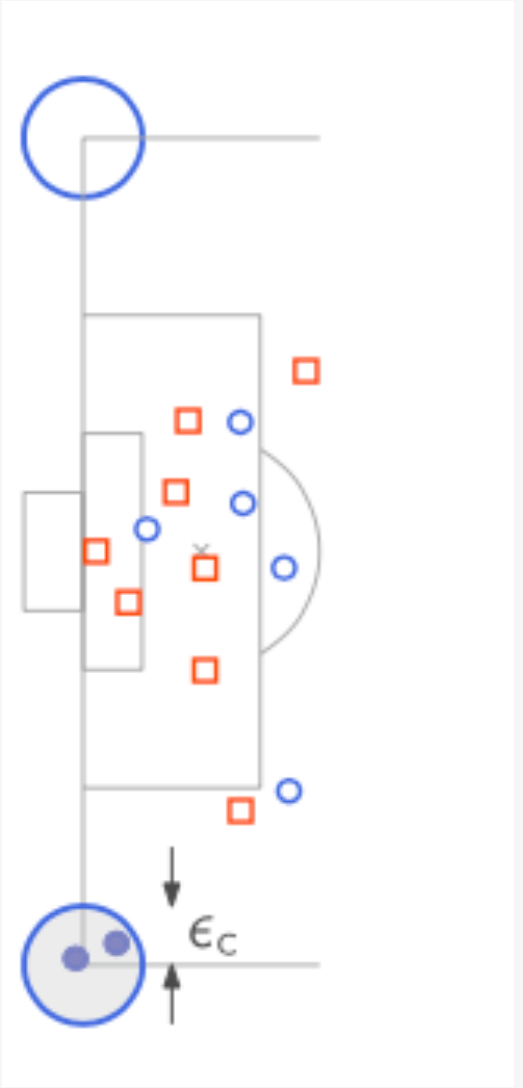
- At least one player from the team taking the goal kick within their own goal area (with a tolerance bounding box of ϵ_g, i.e., 0.5 meters).

- No players from the opposing team should be in the penalty area, and the ball must be present inside the goal area.



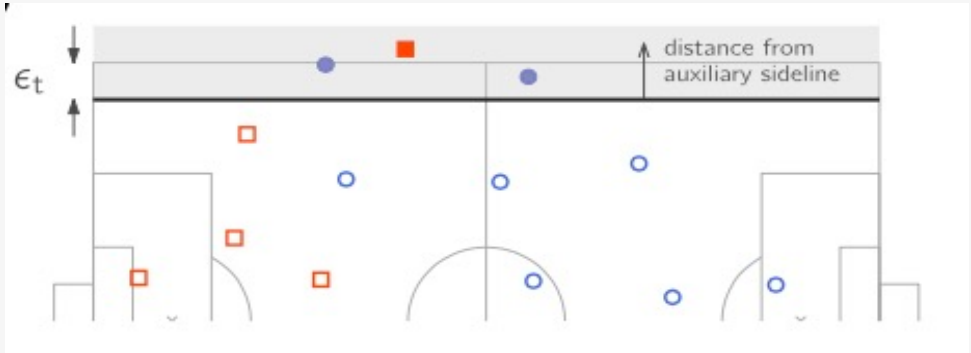
Corner-Kick

- At least one player is within ϵ_C (2 meters) of one of their active corner marks.



Throw-In

- At least one player and the ball is beyond the auxiliary side- line (with a tolerance ϵ_C , i.e., 1 meter).



Free-Kick

- Lacks distinct trigger configurations.
- If none of the mentioned dead ball events apply, the frame is classified as a free-kick.

Phases of Play

EFI Document

- Tracking data.
- Detects these phases on a **frame-level** by analyzing:
  - Player and ball locations
  - Distances
  - Movement speeds
  - Directions
- By identifying the same phase over consecutive frames, temporal sequences are formed and classified accordingly.

Phases of Play

Implementation

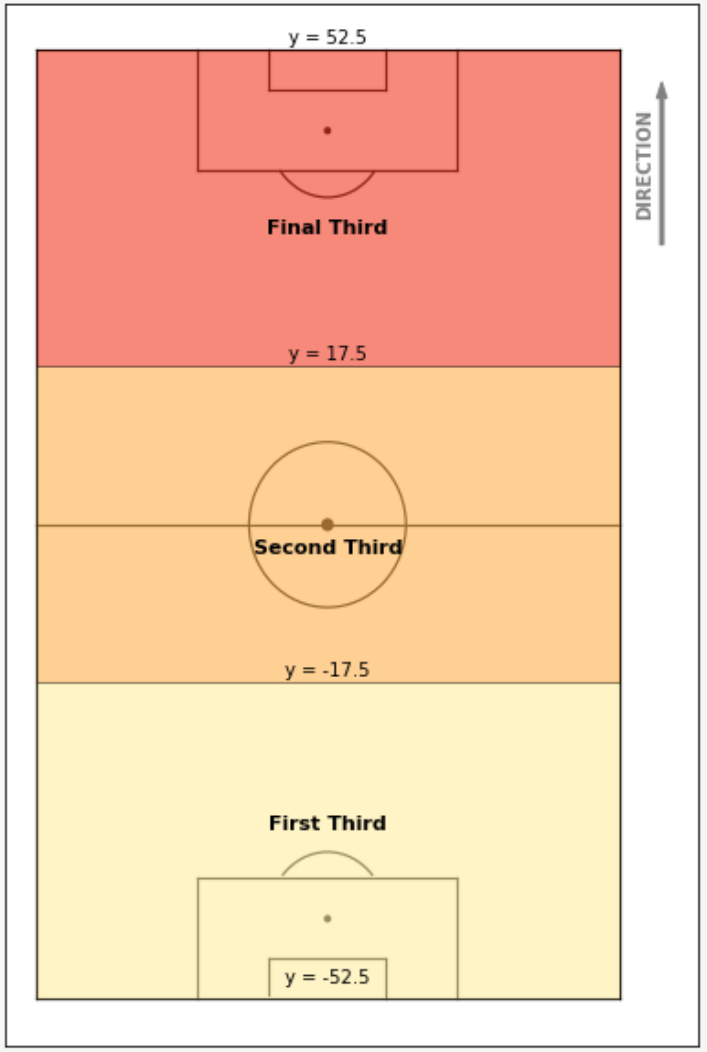
- Sections of analysis:
  - Set-piece and long ball
  - Attacking transition, defensive transition, recovery, counter-attack, and counter-press
  - Build up opposed/unopposed, progression, final third, high press/block, mid-press/block, and low press/block

- After completing each section of the analysis, verifications are conducted to ensure the correct classification of frames.
  - Assignment of the frame to the correct team.
  - Utilization of ball alive frames only.
  - Prevent the double inclusion of the same frame.

Phases of Play

Implementation

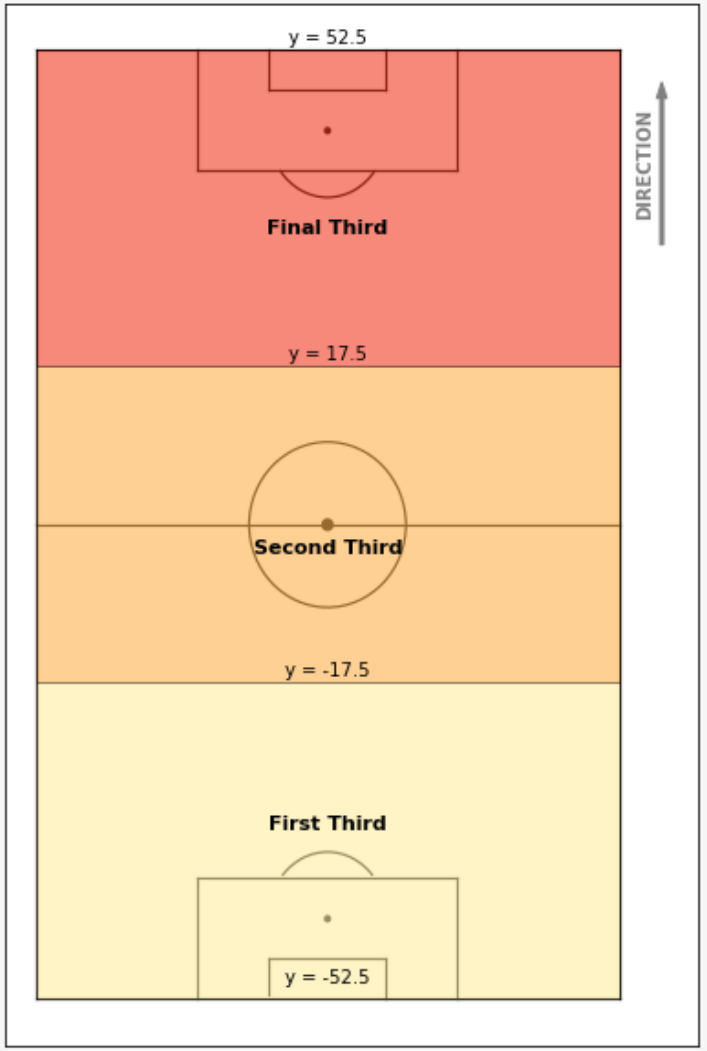
- ***SECTION 3: High Block/Press***
- Ball is in the attacking team’s first third of the pitch.
- *High Press:* A defending player is within “high pressure distance” (5.07 meters) of the ball.
- *High Block:*
  - The defending team's attacking unit is in their final third of the pitch.
  - Defending team’s length $\leq$ “high block threshold” (37.70 meters).



Phases of Play

Implementation

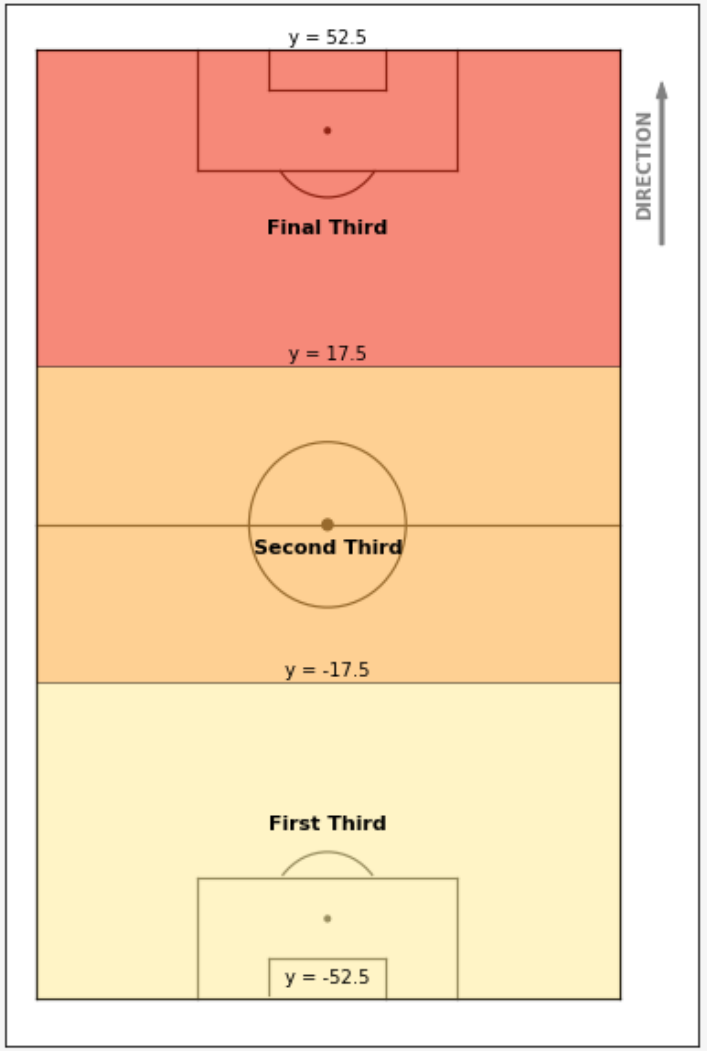
- ***SECTION 3: Mid Block/Press***
- Ball is in the second third of the pitch.
- *Mid Press:* A defending player is within “mid pressure distance” (2.28 meters) of the ball.
- *Mid Block:*
  - The defending team's attacking unit is in the second third of the pitch.
  - Defending team’s length $\leq$ “mid block threshold” (35.31 meters).



Phases of Play

Implementation

- ***SECTION 3: Low Block/Press***
- Ball is in the attacking team’s final third of the pitch.
- *Low Press:* A defending player is within “low pressure distance” (1.23 meters) of the ball.
- *Low Block:*
  - The defending team's attacking unit is in their first third of the pitch.
  - Defending team’s length $\leq$ “low block threshold” (31.85 meters).



Expected Goal (xG)

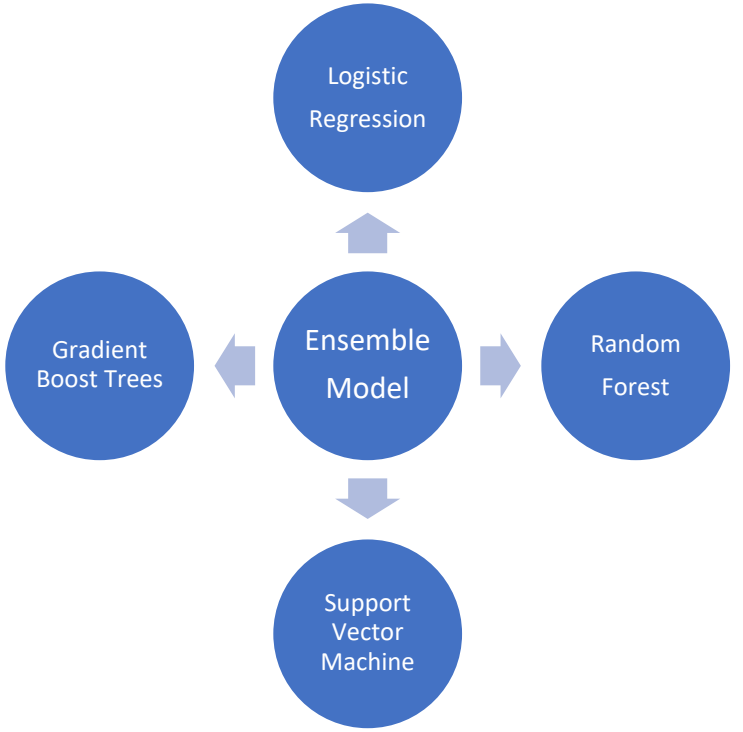
Implementation

- To determine the xG value of shots that take place in the i^{th} match out of 64 matches:
  - Training set (match indices): 1,….,i-1,i+1,....,64
  - Validation set: Randomly selected 20% of the matches from training set (i.e., 5-fold cross validation).
  - Test set: i
- Features of each model:
  - Shot location
  - Pressure presence information
  - Body type used to take the shot
  - Origin (A special indicator of the location where the recorded event took place, such as a penalty or corner)
  - Distance to goal
  - Angle
  - Goalkeeper location
  - Number of players obstructing the goalmouth
- Goal indicator of each shot is used as the label.
- To determine the final ensemble probability result, the accuracy of these predictions is calculated using the provided test labels.
- These accuracy values are used as weights for each model’s probability outcomes.

Expected Goal (xG)

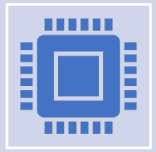
Implementation

- Ensemble Model: Combination of the outputs of individual machine learning models.
  - Having different angles to the same concept increases the robustness of the results.
- Logistic Regression: Simple and interpretable.
- Random Forest: Robust, handles high-dimensional data, resistant to overfitting, capture complex, non-linear relationships between features and the target variable
- Support Vector Machine: Capture the cases when decision boundary between shots that result in goals and those that don’t is non-linear. Which can assist in recognizing critical regions for scoring goals.
- Gradient Boost Trees: Process complicated interactions between features and capture non-linear relationships.



Data Syncronization

Using FIFA's data without their technology, only specific sections of the tracking data can be matched with the event data.



FIFA uses an algorithm for precise event data matching with tracking data, simplifying concept implementation.

FIFA's provision of synchronized data or matching methodology would significantly enhance the implementation of the concepts.

Lack of Implementation Details

- Enhancements in the implementation would benefit from a more detailed explanation from FIFA, especially regarding terminology, technical details, and data sections.

Heuristics

- Replacing heuristics with efficient techniques (e.g., geometric approaches) that do not significantly increase the complexity would enhance the stability and consistency of the methodologies.

Dataset Extension

- Improving heuristic value determination involves using more data and FIFA reports for guidance.
- Report outcomes can act as labels, aiding in parameter estimation and leading to enhanced robustness and precision for each concept.
- For the expected goals concept, a historical database necessitates augmenting the dataset with more shots to expand the available data.