Check for
updates

A framework for the fine-grained evaluation
of the instantaneous expected value of soccer possessions

Javier Fernández[1] 🄳 · Luke Bornn[2] · Daniel Cervone[3]

© The Author(s) 2021

Abstract

The expected possession value (EPV) of a soccer possession represents the likelihood of
a team scoring or conceding the next goal at any time instance. In this work, we develop
a comprehensive analysis framework for the EPV, providing soccer practitioners with the
ability to evaluate the impact of observed and potential actions, both visually and analyti-
cally. The EPV expression is decomposed into a series of subcomponents that model the
influence of passes, ball drives and shot actions on the expected outcome of a possession.
We show we can learn from spatiotemporal tracking data and obtain calibrated models for
all the components of the EPV. For the components related with passes, we produce vis-
ually-interpretable probability surfaces from a series of deep neural network architectures
built on top of flexible representations of game states. Additionally, we present a series of
novel practical applications providing coaches with an enriched interpretation of specific
game situations. This is, to our knowledge, the first EPV approach in soccer that uses this
decomposition and incorporates the dynamics of the 22 players and the ball through track-
ing data.

Keywords Deep learning · Sports analytics · Spatiotemporal statistics · Convolutional
neural networks

1 Introduction

Professional sports teams have started to gain a competitive advantage in recent decades by
using advanced data analysis. However, soccer has been a late bloomer in integrating ana-
lytics, mainly due to the difficulty of making sense of the game's complex spatiotemporal
relationships. To address the nonstop flow of questions that coaching staff deal with daily,

✉ Javier Fernández
  javier.fernandezr@fcbarcelona.cat; javier.fernandez.de.la.rosa@upc.edu

[1] FC Barcelona and Polytechnic University of Catalonia, Barcelona, Spain

[2] Simon Fraser University, Burnaby, Canada

[3] Zelus Analytics, Austin, USA

we require a flexible analysis framework. Such a framework should capture the complex spatial and contextual factors that rule the game while providing practical interpretations of real game situations. Some of these questions are: "which were the most relevant actions leading to goals?", "in which moments of the match did we have a higher chance of scoring and conceding goals?", "how should we defend against our next opponent to concede fewer spaces in the midfield?", or "which are the more frequently open players to receive valuable passes from our creative midfielder?".

This paper addresses the problem of estimating the expected value of soccer possessions (EPV) and proposes a decomposed learning approach that allows us to obtain fine-grained visual interpretations from neural network-based components. The EPV is essentially an estimate of which team will score the next goal at any given time, given all the spatiotemporal information available (e.g., the locations of the players and the ball, and observed actions). Let $G \in \{1, -1, 0\}$, where the values represent the team in control of the ball scoring next, the other team scoring next, or the match half ending, respectively; the EPV corresponds to the expected value of G. The frame-by-frame estimation of EPV constitutes a one-dimensional time series that provides an intuitive description of how the possession value changes in time, as presented in Fig. 1. While this value alone can provide precise information about the impact of observed actions, it does not provide sufficient practical insight into either the factors that make it fluctuate or which other advantageous actions could be taken to boost EPV further. To reach this granularity level, we formulate EPV as a composition of the expectations of three different on-ball actions: passes, ball drives, and shots. Each of these components is estimated separately, producing an ensemble of models whose outputs can be merged to produce a single EPV estimate. Additionally, by inspecting each model, we can obtain detailed insight on the impact that each of the components has on the final EPV estimation.

We propose two different approaches to learn each of the separated models, depending on whether we need to estimate a field-wide probability surface or producing only a single-valued prediction. We propose several deep neural architectures capable of producing full
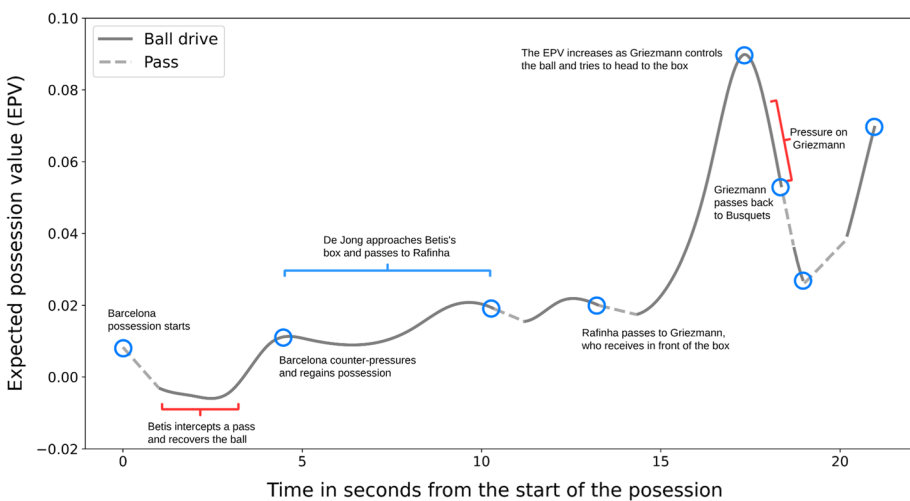

Fig. 1 Evolution of the expected possession value (EPV) from the perspective of FC Barcelona during a match against Real Betis in La Liga season 2019/2020

prediction surfaces from low-level features for the first case. We show that it is possible to learn these surfaces from very challenging learning set-ups where only a single-location ground-truth correspondence is available for estimating the whole surface. Producing these surfaces allows the components related to passes to estimate either the expected value or the probability of attempting a pass to any other location on the field. On the other hand, for the components related to ball drive or shot actions, we use shallow neural networks on top of a broad set of novel spatial and contextual features, to produce single estimations of the expected value or the success probability of these actions. From a practical standpoint, we are splitting out a complex model into more easily understandable parts so the practitioner can both understand the factors that produce the final estimate and evaluate the effect that other possible actions may have had. This type of modeling allows for easier integration of complex prediction models into the decision-making process of groups of individuals with a non-scientific background. Additionally, each of the components can be used individually, multiplying the number of potential applications. For example, we can help a coach identify situations where players took a shot with few chances of scoring but had passing opportunities with a higher scoring chance in the next seconds to show them to the players. As another example, we could identify situations where a player has passing options with high expected value but he is not taking profit from them, so the coach can analyze together with the player how to approach these situations in future matches.

The main contributions of this work are the following:

- We propose a framework for estimating the instantaneous expected outcome of any soccer possession, which allows us to provide professional soccer coaches with rich numerical and visual performance metrics.
- We show that by decomposing the target EPV expression into a series of sub-components and estimating these separately, we can obtain accurate and calibrated estimates and provide a framework with greater interpretability than single-model approaches (Cervone et al. 2016; Bransen and Van Haaren 2018).
- We develop a series of deep learning architectures to estimate the expected possession value surface of potential passes, pass success probability, pass selection probability surfaces, and show these three networks provide both accurate and calibrated surface estimates.
- We present a handful of novel practical applications in soccer that are directly derived from this framework.

2 Background

The evaluation of individual actions has been recently gaining attention in soccer analytics research. Given the relatively low frequency of soccer goals compared to match duration and the frequency of other events such as passes and turnovers, it becomes challenging to evaluate individual actions within a match. Several different approaches have been attempted to learn a valuation function for both on-ball and off-ball actions related to goal-scoring.

Handcrafted features based on the opinion of a committee of soccer experts have been used to quantify the likelihood of scoring in a continuous-time range during a match (Link et al. 2016). Another approach uses a broad set of attributes to estimate individual actions' value during the development of possessions (Decroos et al. 2019). These attributes

include the origin location, type, body part, the time where the action takes place, and its outcome. Here, the game state is represented as a finite set of consecutive observed discrete actions and, a Bernoulli distributed outcome variable is estimated through standard supervised machine learning algorithms. In a similar approach, possession sequences are clustered based on dynamic time warping distance, and an XGBoost (Chen and Guestrin 2016) model is trained to predict the expected goal value of the sequence, assuming it ends with a shot attempt (Bransen and Van Haaren 2018). Gyarmati and Stanojevic (2016) calculate the value of a pass as the difference of field value between different locations when a ball transition between these occurs. Rudd (2011) uses Markov chains to estimate the expected possession value based on individual on-ball actions and a discrete transition matrix of 39 states, including zonal location, defensive state, set pieces, and two absorbing states (goal or end of possession). A similar approach named expected threat uses Markov chains and a coarsened representation of field locations to derive the expected goal value of transitioning between discrete locations (Singh 2019). The estimation of a shot's expectation within the next 10 seconds of a given pass event has also been used to estimate a pass's reward, based on spatial and contextual information (Power et al. 2017). Beyond the quantification of on-ball actions, off-ball position quality has also been quantified, based on the goal expectation. In Spearman (2018), a physics-based statistical model is designed to quantify the quality of players' off-ball positioning based on the positional characteristics at the time of the action that precedes a goal-scoring opportunity. All of these previous attempts on quantifying action value in soccer assume a series of constraints that reduce the scope and reach of the solution. Some of the limitations of these past works include simplified representations of event data (consisting of merely the location and time of on-ball actions), using strongly handcrafted rule-based systems, or focusing exclusively on one specific type of action. However, a comprehensive EPV framework that considers both the full spatial extent of the soccer field and the space-time dynamics of the 22 players and the ball has not yet been proposed and fully validated. In this work, we provide such a framework and go one step further estimating the added value of observed actions by providing an approach for estimating the expected value of the possession at any time instance.

Action evaluation has also been approached in other sports such as basketball and ice-hockey by using spatiotemporal data. The expected possession value of basketball possessions was estimated through a multiresolution process combining macro-transitions (transitions between states following coarsened representation of the game state) and micro-transitions (likelihood of player-level actions), capturing the variations between actions, players, and court space (Cervone et al. 2016). Also, deep reinforcement learning has been used for estimating an action-value function from event data of professional ice-hockey games (Liu and Schulte 2018). Here, a long short-term memory deep network is trained to capture complex time-dependent contextual features from a set of low-level input information extracted from consecutive on-puck events.

This paper employs two frequently used data types in sports analytics: event data and spatiotemporal tracking data. Event data consists of a series of annotated events observed during matches, which include the location and time of the start and end of the event, the name of the player attempting and receiving the action (when it applies), as well as a large set of additional game-related labels, depending on the event. Usually, event data includes on-ball actions such as goals, passes, shots, aerial duels, crosses, set-pieces, tackles, dribbles, or many other typical soccer actions; and some specialized sources might include off-ball actions, such as ball pressure, or team-related information such as lineups and formations. On the other hand, tracking data consists of all the players' locations on the field and the ball. This data is usually provided at a frequency ranging from 10Hz to 25Hz and

captured using computer-vision algorithms on top of soccer match videos. This type of spatiotemporal data is typically obtained through a semi-automated process where, first, ball and player locations are automatically recognized and then manually verified and corrected in case of misidentification. For both data types, the locations are usually normalized according to the team in possession of the ball (e.g., left to right) and are provided in 2-dimensional (or 3-dimensional) coordinate systems that can be transformed following the length and width of the field dimensions.

3 Structured modeling

In this study, we aim to provide a model for estimating a soccer possession's expected outcome at any given time. While the single EPV estimate has practical value itself, we propose a structured modeling approach where the EPV is decomposed into a series of subcomponents. Each of these components can be estimated separately, providing the model with greater adaptability to component-specific problems and facilitating the final estimate's interpretation. In this section, we provide the theoretical framework for approaching EPV as a Markov decision process, as well as the decomposed modeling approach. Additionally, we provide a broad definition of soccer possessions in the context of the EPV.

3.1 Defining possessions in soccer

Although there is no definition of possession within soccer rules, this concept is frequently used in game analyses as units encompassing sequences of actions. The standard approach is to consider possessions the time slots in which a team controls the ball. However, this definition may vary depending on the problem and the analysis approach. We will provide a broad concept of possession, which allows us to capture the long-term expected value from any given game situation. In the context of the EPV, we only require possessions to have three well-defined elements: the starting time, the ending time, and an observed outcome. We assume that possessions start from a single initial state represented by kick-offs (i.e., the first event taking place after a half starts or a goal is scored). Regarding the outcome, we assume three possible absorbing states: one of the two teams scores a goal or that a match half ends. When we reach an absorbing state, we will say the possession resets. The starting and ending time of the possession is defined by the time either an initial or absorbing state is reached, respectively. To facilitate the learning process, an additional absorbing state could be introduced to reset the possession when a goal is not observed after a fixed amount of time ϵ. Notice that when $\epsilon = \infty$ we return to the original definition of three absorbing states. With this broad definition of possessions, we include both teams' events within the possession time range and set the possession's outcome according to the next team scoring a goal, or a match half ends. In this theoretical definition, possessions are not necessarily associated with a specific team. It provides a loose approach to the game's fluid nature that involves frequent switches of the ball control between the teams.

Some other approaches of action-value models provide different definitions for initial and absorbing states. Usually, the initial state of possessions is defined by a change in the ball control between teams. Regarding the outcome of the possession, several approaches have been proposed, such as the next shot observed within a fixed amount of time (Power et al. 2017), a goal being observed after a fixed amount of on-ball actions (Decroos et al. 2019), or the expected goals of a shot observed within the ball control of a given team

(Bransen and Van Haaren 2018). All of these alternative definitions share the main characteristics of the broader definition provided in this section. Having this, both the decomposed approach, presented in Sect. 3.3, and the inference of these components presented in Sect. 5, could be developed following any of these different definitions of possessions.

3.2 EPV as a Markov decision process

This problem can be framed as a Markov decision process (MDP). Let a player with possession of the ball be an agent that can take any action of a discrete set A from any state of the set of all possible states S; we aim to learn the state-value function $EPV(s)$, defined as the expected return from state s, based on a policy $\pi(s, a)$, which defines the probability of taking action a at state s. By approaching the EPV in this way, we are essentially focusing on the problem of estimating the long-term reward (EPV) that a team in possession of the ball might expect, according to the game situation (state) at any given time. To estimate this, we need to represent the game state with soccer spatiotemporal data, define the series of discrete actions that a player can take at any time (A), and estimate how probable it is that a player takes that action ($\pi(s, a)$), given the game state. In contrast with typical MDP applications, our aim is not to find the optimal policy π (i.e., what is the best action the player can take), but to estimate the expected possession value (EPV) from an average policy learned from historical data (i.e., which are the most likely actions).

Let Γ be the set of all possible soccer possessions, and $r \in \Gamma$ represents the full path of a specific possession. Let Ψ be a high dimensional space, including all the spatiotemporal information and a series of annotated events, $T_t(r) \in \Psi$ is a snapshot of the spatiotemporal data after t seconds from the start of the possession. And let $G(r)$ be the outcome of a possession r, where $G(r) \in \{1, -1, 0\}$, with 1 being a goal is scored by the team in control of the ball, -1 being a goal is conceded, and 0 being that the match half ends.

Definition 1 The expected possession value of a soccer possession at time t is $EPV_t = \mathbb{E}[G|T_t]$

Note that since the probability of a team scoring equals the opponent team conceding probability and vice versa, we can estimate and express the EPV from either team's perspective. Following this, G could equivalently be parameterized as the home team scoring next, the away team scoring next, or the half ends. However, we stick to the perspective of the controlling team throughout for ease of narrative.

Definition 1 shares similarities with previous approaches in other sports, such as basketball (Cervone et al. 2016) and American football (Yurko et al. 2020), from which part of the notation used in this section is inspired. Following this definition, we can observe that the EPV is an integration over all the future paths a possession can take at time t, given the available spatiotemporal information at that time, T_t. Note that T_t is essentially a subset of data from all the possible spatiotemporal information that could be available (Ψ), taken at time t. At this modeling stage, we want to express that, while $T_t(r)$ could take many different shapes depending on the implementation and data sources available, it essentially represents the available data that the outcome of the possession G will be conditioned to, when estimating $EPV_t = \mathbb{E}[G|T_t]$.

This model is designed to be applied on top of spatiotemporal tracking data and assuming the data is accompanied and synchronized with event data, consisting of annotated events observed during the match, indicating the location, time, and other possible tags.

Let Ψ be the infinite set of possible tracking data snapshots; this modeling approach defines a continuous state space, represented by Ψ.

3.3 A decomposed model

To obtain the desired structured modeling of EPV described in Sect. 3.2, we will further decompose Definition 1 following the law of total expectation and considering the set of possible actions that can be attempted at any given time. While the set of possible actions is infinite, in theory, there is a small discrete set of actions that soccer practitioners frequently refer to when describing the game, including passes, shots, ball touches, take-on, and aerial duels, among many others. In general, we consider that the smaller set of actions that encompasses most observable soccer actions are passes, ball drives, and shots. Arguably, most named soccer actions are concepts derived from these three main mentioned actions. In this work, we will consider a pass any action where a player intends to transfer the ball's control to a teammate (including successful and missed passes). Shots are all the actions where a player kicks the ball intending to score a goal. We will broadly define a ball drive as the action of a player maintaining control of the ball before the next action is observed or the game stops (e.g., dead ball or half end).

We assume that the space of possible actions $A = \{\rho, \delta, \varsigma\}$ is a discrete set where ρ, δ, and ς represent pass, ball drive, and shot attempt actions, respectively. We can rewrite Definition 1 as in Equation 1.

$$EPV_t = \sum_{a \in A} \mathbb{E}[G|A = a, T_t] \quad \overbrace{\mathbb{P}(A = a|T_t)}^{\substack{\text{Action selection} \\ \text{probability}}} \tag{1}$$

Additionally, to consider that passes can go anywhere on the field, we define D_t to be the selected pass destination location at time t and $\mathbb{P}(D_t|T_t)$ to be a transition probability model for passes. Let L be the set of all the possible locations in a soccer field, then $D_t \in L$. On the other hand, for ball drives (δ) and shots (ς) we do not consider the destination location of the action. Following this, we can rewrite Definition 1 as presented in Equation 2. This expression ensures that both the components estimating the expected value of passes and the pass selection probability are conditioned to consider every possible destination location on the field.

$$EPV_t = \Big(\sum_{l \in L} \overbrace{\mathbb{E}[G|A = \rho, D_t = l, T_t]}^{\substack{\text{Joint expected value} \\ \text{surface of passes}}} \overbrace{\mathbb{P}(D_t = l|A = \rho, T_t)}^{\substack{\text{Pass selection} \\ \text{probability}}}\Big) \mathbb{P}(A = \rho|T_t)$$

$$+ \overbrace{\mathbb{E}[G|A = \delta, T_t]}^{\substack{\text{Expected value} \\ \text{of ball drives}}} \mathbb{P}(A = \delta|T_t) \tag{2}$$

$$+ \overbrace{\mathbb{E}[G|A = \varsigma, T_t]}^{\substack{\text{Expected value} \\ \text{from shots}}} \mathbb{P}(A = \varsigma|T_t)$$

The expected value of passing actions, $\mathbb{E}[G|D, A = \rho]$, can be further extended to include the two scenarios of producing a successful or a missed pass (turnover). We model the

outcome of a pass as O_ρ, which takes a value of 1 when a pass is successful or 0 in case of a turnover. We can then rewrite this expression as in Equation 3. In this step, we are enforcing the expression to consider the impact of the action-outcome, as well as conditioning this outcome to the selected destination location.

$$
\mathbb{E}[G|A=\rho, D_t, T_t] = \overbrace{\mathbb{E}[G|A=\rho, O_\rho=1, D_t, T_t]}^{\substack{\text{Expected value of}\\\text{successful/missed passes}}} \overbrace{\mathbb{P}(O_\rho=1|A=\rho, D_t, T_t)}^{\substack{\text{Probability of}\\\text{successful/missed}\\\text{passes}}}
$$
$$
+ \mathbb{E}[G|A=\rho, O_\rho=0, D_t, T_t]\, \mathbb{P}(O_\rho=0|A=\rho, D_t, T_t) \tag{3}
$$

Equation 4 represents an analogous definition for ball drives, having O_δ be a random variable taking values 0 or 1, representing a successful ball drive or a loss of ball control following that ball drive, which we will refer as a missed ball drive.

$$
\mathbb{E}[G|A=\delta, T_t] = \overbrace{\mathbb{E}[G|A=\delta, O_\delta=1, T_t]}^{\substack{\text{Expected value of}\\\text{successful/missed}\\\text{ball drives}}} \overbrace{\mathbb{P}(O_\delta=1|A=\delta, T_t)}^{\substack{\text{Probability of}\\\text{successful/missed}\\\text{ball drives}}}
$$
$$
+ \mathbb{E}[G|A=\delta, O_\delta=0, T_t]\, \mathbb{P}(O_\delta=0|A=\delta, T_t) \tag{4}
$$

Finally, the expression $\mathbb{E}[G|A=\varsigma]$ is equivalent to an expected goals model, a popular metric in soccer analytics (Lucey et al. 2014; Eggels 2016) which models the expectation of scoring a goal based on shot attempts. In Fig. 2 we present how the outputs of the different components presented in this section are combined to produce a single EPV estimation, while also providing numerical and visual information of how each part of the model impacts the final value. All the proposed components represent concepts that are familiar to soccer practitioners. Ideas such as identifying that a particular pass might have higher scoring value or lower likelihood to be completed, that certain shot attempts are more likely to become goals than others, or that the next action to select might be impacted by the location of the 22 players and the ball, are part of the analysis mindset of professional soccer coaches. By providing coaches with a tool for both analytical and visual interpretation capabilities that considers these familiar concepts, we expect to ease the integration of data-driven analysis within professional coaching staff.

4 Spatiotemporal feature extraction

Each of the decomposed EPV formulation components presents challenging tasks and requires sufficiently comprehensive representations of the game states to produce accurate estimates. We build these state representations from a wide set of low-level and fine-grained features extracted from tracking data. While low-level features are straightforwardly
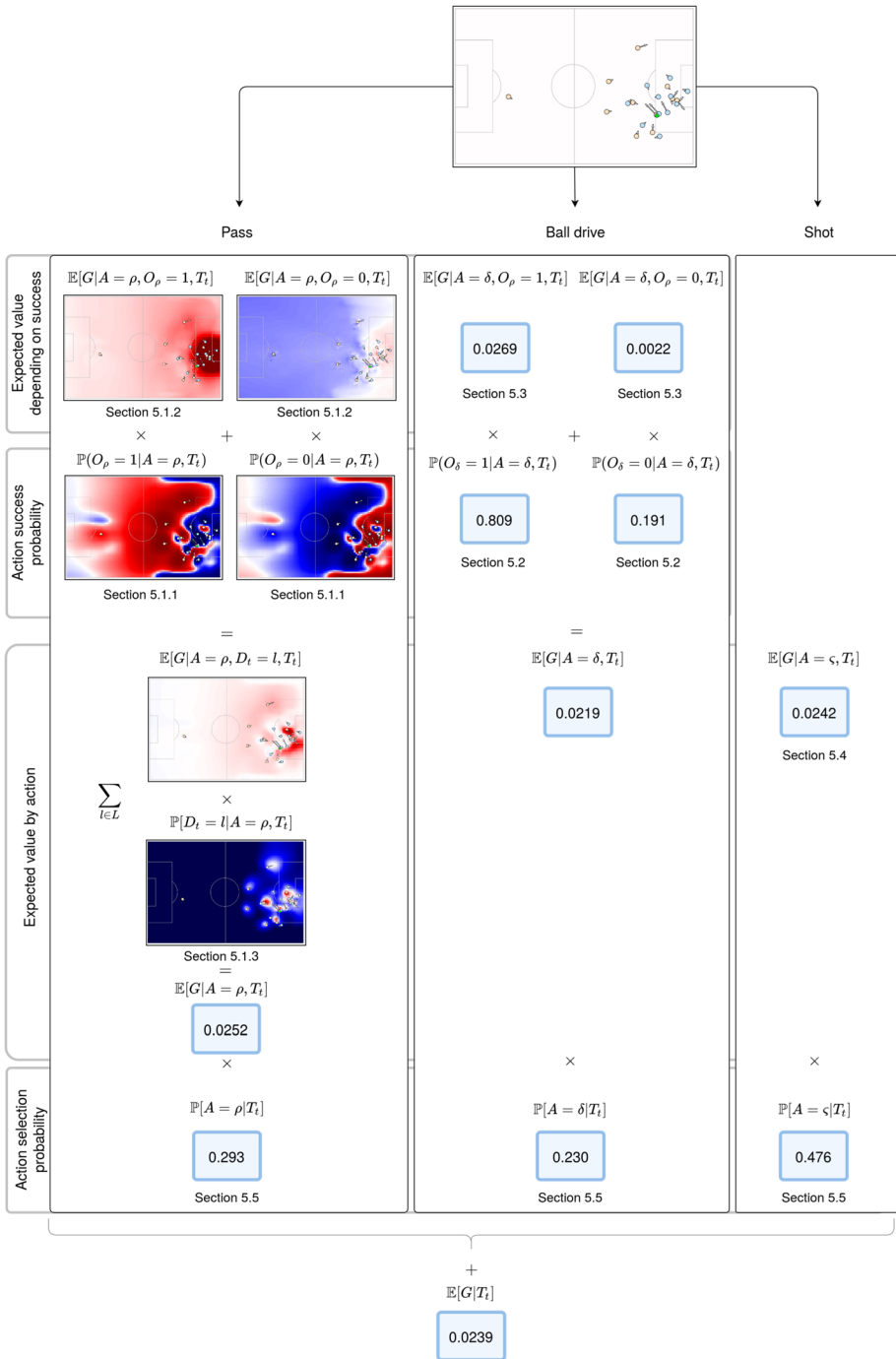
obtained from this data (i.e., a player's location and speed), fine-grained features are built through either statistical models or handcrafted algorithms developed in collaboration with a group of soccer match analysts from FC Barcelona. Figure 3 presents a visual representation of a game situation where we can observe the available players and ball locations and a subset of features derived from that tracking data snapshot. Conceptually, we split the features into two main groups: spatial features and contextual features. Both feature types are described in Sects. 4.1 and 4.2. The full set of features and their usage within the different models presented in this work are detailed in "Appendix 1".

4.1 Spatial features

We consider spatial features directly derived from the spatial location of the players and the ball in a given time range. These can be obtained for any game situation regardless of the context and comprise mainly physical and spatial information. Table 1 details a set of concepts where the specific list of features presented in "Appendix 1" are derived from. The main spatial features obtained from tracking data are related to the location of players from both teams, the velocity vector of each player, the ball's location, and the location of the opponent's goal at any time instance. From the player's spatial location, we produce
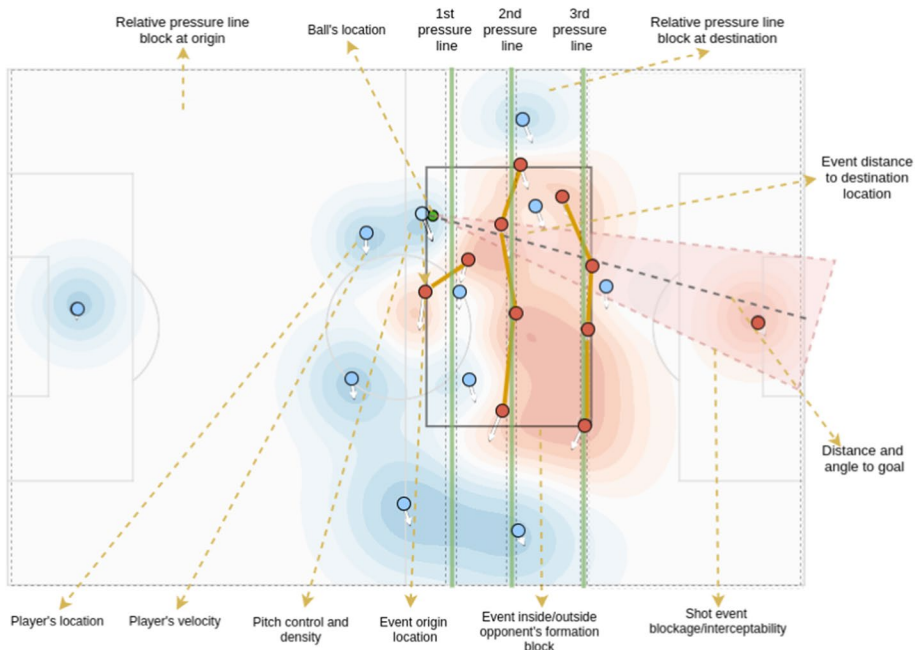

Fig. 3 Visual representation of a tracking data snapshot of spatial and contextual features in a soccer match situation. Blue and red dots represent players of the attacking and defending team, respectively, while the green dot represents the ball location. The blue and red surface represents the pitch control of each team along the field. The grey rectangle covering the red dots represents the defending team's formation block. The green vertical lines represent the defending team's vertical dynamic pressure lines, while the polygons with solid yellow lines represent the players clustered in each pressure line. The black dotted rectangles represent the relative locations between dynamic pressure lines. Dotted yellow lines and associated text describe the main extracted features (Color figure online)

Table 1 Description of a set of spatial concepts derived from tracking data

Concept type	Description
(x,y) location	Location of a player, the ball, or attempted action, normalized in the [0,1] range, relative to the pitch's dimensions. The location (0,0) corresponds to the top-left corner
Pitch control	Probability of controlling the ball in a specific location
Pitch influence	Degree of influence of a set of players in a specific location
Distance between locations	Distance in meters between two locations
Angle between locations	Angle in degrees between two locations
Player's velocity	Player's velocity vector in the last second

a series of features related to the control of space and players' density along the field. The statistical models used for pitch control and pitch influence evaluation are detailed in "Appendix 2".

4.2 Contextual features

To provide a more comprehensive state representation, we include a series of features derived from soccer-specific knowledge, which provides contextual information to the model. Table 2 presents the main concepts from which multiple contextual features are derived.

The concept of dynamic pressure lines refers to players being aligned with their teammates within different alignment groups. For example, a typical conceptualization of pressure lines in soccer would be the groups formed by the defenders, the midfielders, and the attackers, which tend to be aligned to keep a consistent formation. The details on the calculation of dynamic pressure lines are presented in "Appendix 3". By identifying the pressure lines, we can obtain every player's opponent-relative location, which provides high-level information about players' expected behavior. For example, when a player controls the ball and is behind the opponent's first pressure line, we would expect a different pressure behavior and turnover risk than when the ball is close to the third pressure line and the goal. Also, the soccer experts that accompanied this study considered passes that break pressure lines to significantly impact the increase of the goal expectation of the possession.

Table 2 Description of a set of contextual concepts derived from tracking data

Concept type	Description
Possession	Possessions start and end times are identified to segment each match in episodes or sequences of actions
Dynamic pressure lines	Relative positioning of players according to the team's current formation or the opponents. These formations change dynamically and are calculated in time ranges of a few seconds
Outplayed players	Number of players that are surpassed after an action is attempted
Interceptability	Features related to the likelihood of intercepting the ball
Baseline event-based models	Models built on top of event data, which are used as a baseline to enrich the learning of tracking data-based models

From the concept of outplayed players, we can derive features such as the number of opponent players to overcome after a given pass is attempted or the number of teammates in front of or behind the ball, among many similar derivatives. In combination with the opponent's formation block location, we can obtain information about whether the pass is headed towards the inside or outside of the formation block and how many players are to be surpassed. Intuitively, a pass that outplays several players and that is headed towards the inside of the opponent block is more likely to produce an increase of the EPV, than a pass back directed outside the opponent's block that adds two more opponent players in front of the ball. On the other hand, the interceptability concept is expected to play an essential role in capturing opponents' spatial influence near a shooting option, allowing us to produce a more detailed expected goals model. Mainly, we derive features related to the number of players pressing the shooter closely and the number of players in the triangle formed between the shooter and the posts.

The described spatial and contextual features represent the main building blocks for deriving the set of features used for each implemented model. In Sect. 5, we describe in great detail the characteristics of these models.

5 Separated component inference

In this section we describe the approaches followed for estimating each of the components described in Equations 1, 2, 3 and 4. In general, we use function approximation methods to learn models for these components from spatiotemporal data. Specifically, we want to approximate some function f^* that maps a set of features x, to an outcome y, such that $y = f^*(x)$. To do this, we will find the mapping $y = f(x;\theta)$ to learn the values of a set of parameters θ that result in an approximation to f^*.

Customized convolutional neural network architectures are used for estimating probability surfaces for the components involving passes, such as pass success probability, the expected possession value of passes, and the field-wide pass selection surface. Standard shallow neural networks are used to estimate ball drive probability, expected possession value from ball drives and shots, and the action selection probability components. This section describes the selection of features x, observed value y, and model parameters θ for each component.

5.1 Estimating pass impact at every location on the field

One of the most significant challenges when modeling passes in soccer is that, in practice, passes can go anywhere on the field. Previous attempts on quantifying pass success probability and expected value from passes in both soccer and basketball assume that the passing options a given player has are limited to the number of teammates on the field, and centered at their location at the time of the pass (Power et al. 2017; Cervone et al. 2016; Hubáček et al. 2018). However, in order to accurately estimate the impact of passes in soccer (a key element for estimating the future pathways of a possession), we need to be able to make sense of the spatial and contextual information that influences the selection, accuracy, and potential risk and reward of passing to any other location on the field. We propose using fully convolutional neural network architectures designed to exploit spatiotemporal information at different scales. We extend it and adapt it to the three related passing action models we require to learn: pass success probability, pass

selection probability and pass expected value. While these three problems necessitate from different design considerations, we structure the proposed architectures in three main conceptual blocks: a *feature extraction block*, a *surface prediction block*, and a *loss computation block*. The proposed models for these three problems also share the following common principles in its design: a layered structure of input data, the use of fully convolutional neural networks for extracting local and global features and learning a surface mapping from single-pixel correspondence. We first detail the common aspects of these architectures and then present the specific approach for each of the mentioned problems.

Layers of low-level and field-wide input data To successfully estimate a full prediction surface, we need to make sense of the information at every single pixel. Let the set of locations L, presented in Sect. 3.3, be a discrete matrix of locations on a soccer field of width w and height h, we can construct a layered representation of the game state $Y(T_t)$, consisting of a set of slices of location-wise data of size $w \times h$. By doing this, we define a series of layers derived from the data snapshot T_t that represent both spatial and contextual low-level information for each problem. This layered structure provides a flexible approach to include all kinds of information available or extractable from the spatiotemporal data, which is considered relevant for the specific problem being addressed.

Feature extractor block The feature extractor block is fundamentally composed of fully convolutional neural networks for all three cases, based on the SoccerMap architecture (Fernández and Bornn 2020). Using fully convolutional neural networks, we leverage the combination of layers at different resolutions, allowing us to capture relevant information at both local and global levels, producing location-wise predictions that are spatially aware. Following this approach, we can produce a full prediction surface directly instead of a single prediction on the event's destination. The parameters to be learned will vary according to the input surfaces' definition and the target outcome definition. However, the neural network architecture itself remains the same across all the modeled problems. This allows us to quickly adapt the architecture to specific problems while keeping the learning principles intact. A detailed description of the SoccerMap architecture is presented in "Appendix 4".

Learning from single-pixel correspondance Usually, approaches that use fully convolutional neural networks have the ground-truth data for the full output surface. In more challenging cases, only a single classification label is available, and a weakly supervised learning approach is carried out to learn this mapping (Pathak et al. 2015). However, in soccer events, only a single pixel ground-truth information is available: for example, the destination location of a successful pass. This makes our problem highly challenging, given that there is only one single-location correspondence between input data and ground-truth. At the same time, we aim to estimate a full probability surface. Despite this extreme set-up, we show that we can successfully learn full probability surfaces for all the pass-related models. We do so by selecting a single pixel from the predicted output matrix, during training, according to the known destination location of observed passes, and back-propagating the loss at a single-pixel level.

In the following sections, we describe the design characteristics for the feature extraction, surface prediction, and loss computation blocks for the three pass-related problems: pass success probability, pass selection probability, and expected value from passes. By joining these models' output, we will obtain a single action-value estimation (EPV) for passing actions, expressed by $\mathbb{E}[G|A = \rho, T_t]$. The detailed list of features used for each model is described in "Appendix 1".

5.1.1 Pass success probability

From any given game situation where a player controls the ball, we desire to estimate the success probability of a pass attempted towards any of the other potential destination locations, expressed by $\mathbb{P}(A = \rho, D_t | T_t)$. Figure 4 presents the designed architecture for this problem. The input data at time t is conformed by 13 layers of spatiotemporal information obtained from the tracking data snapshot T_t consisting mainly of information regarding the location, velocity, distance, and angles between the both team's players and the goal. The feature extraction block is composed strictly by the SoccerMap architecture, where representative features are learned. This block's output consists of a $104 \times 68 \times 1$ pass probability predictions, one for each possible destination location in the coarsened field representation. In the surface prediction block a sigmoid activation function σ is applied to each prediction input to produce a matrix of pass probability estimations in the [0,1] continuous range, where $\sigma(x) = \frac{e^x}{e^x+1}$. Finally, at the loss computation block, we select the probability output at the known destination location of observed passes and compute the negative log loss, defined in Equation 5, between the predicted ($\hat{y}$) and observed pass outcome (y).

$$\mathcal{L}(\hat{y}, y) = -(y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})) \tag{5}$$

Note that we are learning all the network parameters θ needed to produce a full surface prediction by the back-propagation of the loss value between the predicted value at that location and the observed outcome of pass success at a single location. We show in Sect. 6.6 that this learning set is sufficient to obtain remarkable results.
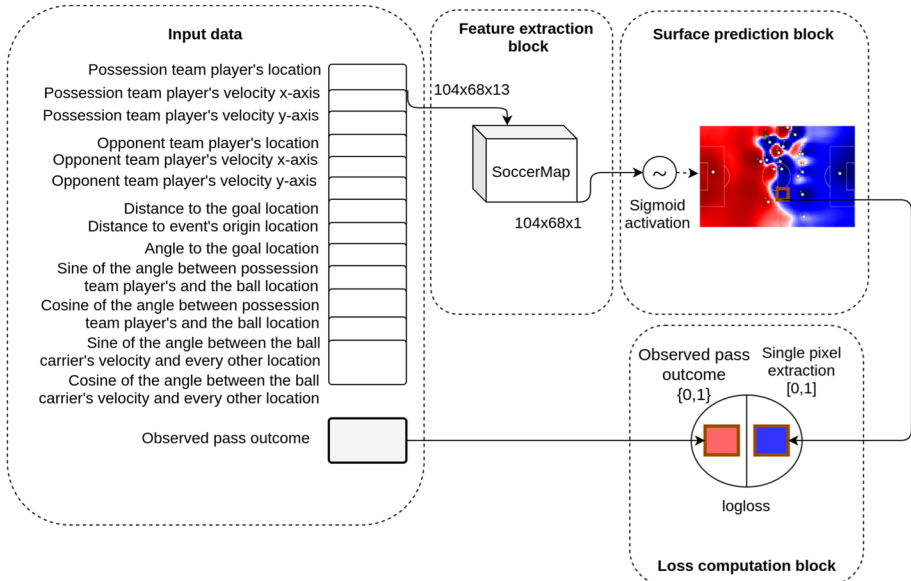

Fig. 4 Representation of the neural network architecture for the pass probability surface estimation, for a coarsened representation of size 104×68. Thirteen layers of spatial features are fed to a SoccerMap feature extraction block, which outputs a 104×68×1 prediction surface. A sigmoid activation function is applied to each output, producing a pass probability surface. The output at the destination location of an observed pass is extracted, and the log loss between this output and the observed outcome of the pass is back-propagated to learn the network parameters

5.1.2 Expected possession value from passes

Once we have a pass success probability model, we are halfway to obtaining an estimation for $\mathbb{E}[G|A = \rho, D_t, T_t]$, as expressed in Equation 3. The remaining two components, $\mathbb{E}[G|A = \rho, O_p = 1, D_t, T_t]$ and $\mathbb{E}[G|A = \rho, O_p = 0, D_t, T_t]$, correspond to the expected value of successful and unsuccessful passes, respectively. We learn a model for each expression separately; however, we use an equivalent architecture for both cases. The main difference is that one model must be learned with successful passes and the other with missed passes exclusively to obtain full surface predictions for both cases.

The input data matrix consists of 16 different layers with equivalent location, velocity, distance, and angular information to those selected for the pass success probability model. Additionally, we append a series of layers corresponding to contextual features related to outplayed players' concepts and dynamic pressure lines. Finally, we add a layer with the pass probability surface, considering that this can provide valuable information to estimate the expected value of passes. This surface is calculated by using a pre-trained version of a model for the architecture presented in Sect. 5.1.1.

The input data is fed to a SoccerMap feature extraction block to obtain a single prediction surface. In this case, we must observe that the expected value of G should reside within the $[-1, 1]$ range, as described in Sect. 3.2. To do so, in the surface prediction block, we apply a sigmoid activation function to the SoccerMap predicted surface obtaining an output within [0, 1]. We then apply a linear transformation, so the final prediction surface consists of values in the $[-1, 1]$ range. Notably, our modeling approach does not assume that a successful pass must necessarily produce a positive reward or that missed passes must produce a negative reward.

The loss computation block computes the mean squared error between the predicted values and the reward assigned to each pass, defined in Equation 6. The model design is independent of the reward choice for passes. In this work, we choose a long-term reward associated with the observed outcome of the possession, detailed in Sect. 6.2.

$$\text{MSE}(\hat{y}, y) = \frac{1}{N} \sum_{i}^{N} (y_i - \hat{y}_i)^2 \qquad (6)$$

5.1.3 Pass selection probability

Until now, we have models for estimating both the probability and expected value surfaces for both successful and missed passes. In order to produce a single-valued estimation of the expected value of the possession given a pass is selected, we model the pass selection probability $\mathbb{P}(A = \rho, D_t|T_t)$ as defined in Equation 1. The values of a pass selection probability surface must necessarily add up to 1, and will serve as a weighting matrix for obtaining the single estimate.

Both the input and feature extraction blocks of this architecture are equivalent to those designed for the pass success probability model (see Sect. 5.1.1). However, we use the softmax activation function presented in Equation 7 for the surface prediction block, instead of a sigmoid activation function. We then extract the predicted value at a given pass destination location and compute the log loss between that predicted value and 1, since only observed passes are used. With the different models presented in Sect. 5.1, we

can now provide a single estimate of the expected value given a pass action is selected, $\mathbb{E}[G|A = \rho, T_t]$.

$$softmax(v)_i = \frac{e^{v_i}}{\sum_{j=1}^{K} e^{v_i}} \text{ for } i = 0, \dots, K \qquad (7)$$

5.2 Estimating ball drive probability

We will focus now on the components needed for estimating the expected value of ball drive actions. In this work's scope, a ball drive refers to actions where a player keeps control of the ball. For this implementation, ball drives lasting more than 1 second are split into a set of individual ball drives of 1-second duration. While keeping the ball, the player might sustain the ball-possession or lose the ball (either because of bad control, an opponent interception, or by driving the ball out of the field, among others). The probability of keeping control of the ball with these conditions is modeled by the expression $\mathbb{P}(O_\delta = 1|A = \delta, T_t)$.

　　We use a standard shallow neural network architecture to learn a model for this probability, consisting of two fully-connected layers, each one followed by a layer of ReLu activation functions, with a single-neuron output preceded by a sigmoid activation function. We provide a state representation for observed ball drive actions that are composed of a set of spatial and contextual features, detailed in "Appendix 1". Among the spatial features, the level of pressure a player in possession of the ball receives from an opponent player is considered to be a critical piece of information to estimate whether the possession is maintained or lost. We model pressure through two additional features: the opponent's team density at the player's location and the overall team pitch control at that same location. Another factor that is considered to influence the ball drive probability is the player's contextual-relative location at the moment of the action. We include two features to provide this contextual information: the closest opponent's vertical pressure line and the closest possession team's vertical pressure line to the player. These two variables are expected to serve as a proxy for the opponent's pressing behavior and the player's relative risk of losing the ball. By adding features related to the spatial pressure, we can get a better insight into how pressed that player is within that context and then have better information to decide the probability of keeping the ball. We train this model by optimizing the loss between the estimated probability and observed ball drive actions that are labeled as successful or missed, depending on whether the ball carrier's team can keep the ball's possession during after the ball drive is attempted.

5.3 Estimating ball drive expectation

Finally, once we have an estimate of the ball drive probability, we still need to obtain an estimate of the expected value of ball drives, in order to model the expression $\mathbb{E}[G|A = \delta, T_t]$, presented in Equation 4. While using a different architecture for feature extraction, we will model both $\mathbb{E}[G|A = \delta, O_\delta = 1, T_t]$ and $\mathbb{E}[A = \delta, O_\delta = 0, T_t]$, following an analogous approach of that used in Sect. 5.1.2.

　　Conceptually, by keeping the ball, players might choose to continue a progressive run or dribble to gain a better spatial advantage. However, they might also wait until a teammate moves and opens up a passing line of lower risk or higher quality. By learning a model

for the expression $\mathbb{E}[G|A = \delta, T_t]$ we aim to capture the impact on the expected possession value of these possible situations, all encapsulated within the ball drive event. We use the same input data set and feature extractor architecture used in Sect. 5.2, with the addition of the ball drive probability estimation for each example. Similarly to the loss surface prediction block of the expected value of passes (see Sect. 5.1.2), we apply a sigmoid activation function to obtain a prediction in the [0, 1] range, and then apply a linear transformation to produce a prediction value in the [−1, 1] range. The loss computation block computes the mean squared loss between the observed reward value assigned to the action and the model output.

5.4 Expected goals model

Once we have a model for the expected values of passes and ball drives, we only need to model the expected value of shots to obtain a full value state-value estimation for the action set A. We want to model the expectation of scoring a goal at time t given that a shot is attempted, defined as $\mathbb{E}[G|A = \varsigma]$. This expression is typically referred to as *expected goals* (xG) and is arguably one of the most popular metrics in soccer analytics (Eggels 2016). For estimating this expected goals model we include spatial and contextual features related derived from the 22 players' and the ball's locations, to account for the nuances of shooting situations.

Intuitively, we can identify several spatial factors that influence the likelihood of scoring from shots, such as the level of defensive pressure imposed on the ball carrier, the interceptability of the shot by close opponents, or the goalkeeper's location. Specifically, we add the number of opponents that are closer than 3 meters to the ball-carrier to quantify the level of immediate pressure on the player. Additionally, we account for the interceptability of the shot (blockage count) by calculating the number of opponent players in the triangle formed by the ball-carrier location and the two posts. We include three additional features derived from the location of the goalkeeper. The goalkeeper's location can be considered an important factor influencing the scoring probability, particularly since he has the considerable advantage of being the only player that can stop the ball with his hands. In addition to this spatial information, we add a contextual feature consisting of a boolean flag indicating whether the shot is taken with the foot or the head, the latter being considered more difficult. Additionally, we add a prior estimation of expected goal as an input feature to this spatial and contextual information, produced through the baseline expected goals model described in "Appendix 5". The full set of features is detailed in "Appendix 1".

Having this feature set, we use a standard neural network architecture with the same characteristics as the one used for estimating the ball drive probability, explained in Sect. 5.2, and we optimize the mean squared error between the predicted outcome and the observed reward for shot actions. The long-term reward chosen for this work is detailed in Sect. 6.2.

5.5 Action selection probability

Finally, to obtain a single-valued estimation of EPV we weigh the expected value of each possible action with the respective probability of taking that action in a given state, as expressed in Equation 1. Specifically, we estimate the action selection probability $\mathbb{P}(A|T_t)$, where A is the discrete set of actions described in Sect. 3.2. We construct a feature set composed of both spatial and contextual features. Spatial features such as the ball location

and the distance and angle to the goal provide information about the ball carrier's relative location in a given time instance. Additionally, we add spatial information related to the possession, and team's pitch control and the degree of spatial influence of the opponent team near the ball. On the other hand, the location of both teams' dynamic lines relative to the ball location provides the contextual information to the state representation. We also include the baseline estimation of expected goals at that given time, which is expected to influence the action selection decision, especially regarding shot selection. The full set of features is described in "Appendix 1". We use a shallow neural network architecture, analogous to those described in Sects. 5.2 and 5.3. This final layer of the feature extractor part of the network has size 3, to which a softmax activation function is applied to obtain the probabilities of each action. We model the observed outcome as a one-hot encoded vector of size 3, indicating the action type observed in the data, and optimize the categorical cross-entropy between this vector and the predicted probabilities, which is equivalent to the log loss.

6 Experimental setup

6.1 Datasets

We build different datasets for each of the presented models based on optical tracking data and event data from 633 English Premier League matches from the 2013/2014 and 2014/2015 season, provided by *STATS LLC*. This tracking data source consists of every player's location and the ball at a $10Hz$ sampling rate, obtained through semi-automated player and ball tracking performed on match videos. On the other hand, event data consists of human-labeled on-ball actions observed during the match, including the time and location of both the origin and destination of the action, the player who takes action, and the outcome of the event. Following our model design, we will focus exclusively on the pass, ball drive, and shot events. Table 3 presents the total count for each of these events according to the dataset split presented below in Sect. 6.3. The definition of success varies from one event to another: a pass is successful if a player of the same team receives it, a ball drive is successful if the team does not lose control of the ball after the action occurs, and a shot is labeled as successful if a goal is scored from that shot. Given this data, we can extract the tracking data snapshot, defined in Sect. 3.2, for every instance where any of these events are observed. From there, we can build the input feature sets defined for each of the presented models. For the detailed list of features used, see "Appendix 1". For each sample, the players' and the ball locations are normalized so the team taking the action is attacking from left to right (i.e., scores goals in the rightmost goal, and concedes goals in the leftmost goal of the field).

Table 3 Total count of events included within the tracking data of 633 English Premier League matches from the 2013/2014 and 2014/2015 season

Data type	# Total	# Training	# Validation	# Test	% Success
Match	633	379	127	127	–
Pass	480,670	288,619	96,500	95,551	79.64
Ball drive	413,123	284,759	82,271	82,093	90.60
Shot	13,735	8240	2800	2695	8.54

6.2 Defining the estimands

Each of the components of the EPV structured model has different estimands or outcomes. For both the pass success and ball drive success probability models, we define a binomially distributed outcome, according to the definition of success provided in 6.1. These outcomes correspond to the short-term observed success of the actions. For the pass selection probability, we define the outcome as a binomially distributed random variable. A value of 1 is given for every observed pass in its corresponding destination location. We define the action selection model's estimand as a multinomially distributed random variable that can take one of three possible values, according to whether the selected action corresponds to a pass, a ball drive, or a shot.

For the EPV estimations of passes, ball drives, and shot actions, respectively, we define the estimand as a long-term reward, corresponding to the outcome of the possession where that event occurs. We follow the definition of possession presented in Sect. 3.1, where possession starts with a kick-off event and ends when a goal is observed, or a match half ends. By doing this, we allow the ball to either go out of the field or change control between teams an undefined number of times until the next goal is observed. Once a goal is observed, all the actions between the goal and the previous one are assigned an outcome of 1 if the action is taken by the scoring team or −1 otherwise. If the match half ends before observing the next goal, the actions' outcome value is set to 0. Following this, each action gets assigned a long-term reward as an outcome.

Additionally, we will include the possession resetting state described in Sect. 3.1 to limit possessions' time extent. There is a low frequency of goals in matches (2.8 goals on average in our dataset) compared to the number of observed actions (1,433 on average). Given this, the definition of the time extent of possession is expected to influence the balance between individual actions' short-term value and the long-term expected outcome after that action is taken. Let ϵ be the constant representing the time in seconds between each action and the next goal; all the actions observed more than ϵ time from the observed goal received a reward of 0. For this work, we choose $\epsilon = 15s$, which corresponds to the average duration of standard soccer possessions in the available matches. Note this is equivalent to assuming that the current state of possession only has ϵ seconds impact.

For the implementation of this model we will use only passes, ball drives and shot actions that are observed within an open-play phase of the possession, and ignore the actions occurring during set-pieces. We will say that an action belongs to a set-piece if it is observed 5 seconds or less from the start of a direct or indirect free-kick, a corner kick, a throw-in or a penalty kick. All the other actions are considered to occur in open-play. It is important to remark that all the goals available in the dataset are used in this implementation, including those occurring within a set-piece time range. This means that if a goal is scored in a corner kick, all the actions preceding the goal will be labeled with −1, 1 or 0 (according to the definition of possession described above), except for those that are 5 seconds or less closer to the goal. By doing this, our implementation focuses on learning the expected value of open-play actions, and leaves for future work the modeling of set-pieces, since these involve different spatiotemporal dynamics.

6.3 Model setting

We randomly sample the available matches and split them into training (379), validation (127), and test sets (127). From each of these matches, we obtain the observed on-ball actions and the tracking data snapshots to construct the set of input features corresponding to each model, detailed in "Appendix 1". The events are randomly shuffled in the training dataset to avoid bias from the correlation between events that occur close in time. We use the validation set for model selection and leave the test set as a hold-out dataset for testing purposes. We train the models using the adaptive moment estimation algorithm (Kingma and Ba 2014), and set the β_1 and β_2 parameters to 0.9 and 0.999 respectively. For all the models we perform a grid search on the learning rate ($\{1e{-}3, 1e{-}4, 1e{-}5, 1e{-}6\}$), and batch size parameters ($\{16, 32\}$). We use early stopping with a delta of $1e{-}3$ for the pass success probability, ball drive success probability, and action selection probability models, and $1e{-}5$ for the rest of the models.

6.4 Model calibration

We include an after-training calibration procedure within the processing pipeline for the pass success probability and pass selection probability models, which presented slight calibration imbalances on the validation set. We use the temperature scaling calibration method for both models, a useful approach for calibrating neural networks (Guo et al. 2017). Temperature scaling consists of dividing the vector of logits passed to a softmax function by a constant *temperature* value T_p. This product modifies the scale of the probability vector produced by the softmax function. However, it preserves each element's ranking, impacting only the distribution of probabilities and leaving the classification prediction unmodified. We apply these post-calibration procedures exclusively on the validation set.

6.5 Evaluation metrics

For the pass success probability, keep ball success probability, pass selection probability, and action selection models, we use the cross-entropy loss. Let M be the number of classes, N the number of examples, y_{ij} the estimated outcome, and $\hat{y}_{ij}$ the expected outcome, we define the cross-entropy loss function as in Equation 8. For the first three models, where the outcome is binary, we set $M = 2$. We can directly observe that for this set-up, the cross-entropy is equivalent to the negative log-loss defined in Equation 5. For the action selection model, we set $M = 3$. For the rest of the models, corresponding to EPV estimations, we can observe the outcome takes continuous values in the $[-1, 1]$ range. For these cases, we use the mean squared error (MSE) as a loss function, defined in Equation 6, by first normalizing both the estimated and observed outcomes into the $[0, 1]$ range.

$$\mathrm{CE}(\hat{y}, y) = -\sum_{j}^{M} \sum_{i}^{N} (y_{ij} \cdot \log(\hat{y}_{ij})) \tag{8}$$

We are interested in obtaining calibrated predictions for all of the models, as well as for the joint EPV estimation. Having the models calibrated allows us to perform a fine-grained interpretation of the variations of EPV within subsets of actions, as shown in Sect. 7. We validate the model's calibration using a variation of the expected calibration error (ECE)

presented in Guo et al. (2017). For obtaining this metric, we distribute the predicted outcomes into K bins and compute the difference between the average prediction in each bin and the average expected outcome for the examples in each bin. Equation 9 presents the ECE metric, where K is the number of bins, and B_k corresponds to the set of examples in the k-th bin. Essentially, we are calculating the average difference between predicted and expected outcomes, weighted by the number of examples in each bin. In these experiments, we use quantile binning to obtain K equally-sized bins in ascending order.

$$\text{ECE} = \sum_{k=1}^{K} \frac{|B_k|}{N} \left| \left(\frac{1}{|B_k|} \sum_{i \in B_k} y_i \right) - \left(\frac{1}{|B_k|} \sum_{i \in B_k} \hat{y}_i \right) \right| \tag{9}$$

6.6 Results

Table 4 presents the results obtained in the test set for each of the proposed models. The loss value corresponds to either the cross-entropy or the mean squared loss, as detailed in Sect. 6.5. The table includes the optimal values for the batch size and learning rate parameters, the number of parameters of each model, and the number of examples per second that each model can predict.

We can observe that the loss value reported for the final joint model is equivalent to the losses obtained for the EPV estimations of each of the three types of action types, showing stability in the model composition. The shot EPV loss is higher than the ball drive EPV and pass EPV losses, arguably due to the considerably lower amount of observed events available in comparison with the rest, as described in Sect. 6.1. While the number of examples per second is directly dependent on the models' complexity, we can observe that we can predict 899 examples per second in the worst case. This value is 89 times higher than the sampling rate of the available tracking data (10Hz), showing that this approach can be applied for the real-time estimation of EPV and its components.

Table 4 The average loss and calibration value for each of the components of the EPV model, as well as for the joint EPV estimation, on the corresponding test datasets

Model	Loss	ECE	Batch size	Learning rate	# Params.	Ex. (s)
Pass probability	0.1900	0.0047	32	1e−4	401,259	942
Ball drive probability	0.2803	0.0051	32	1e−3	128	67,230
Pass successful EPV	0.0075	0.0011	16	1e−6	403,659	899
Pass missed EPV	0.0085	0.0015	16	1e−6	403,659	899
Pass selection probability	5.7134	–	32	1e−5	401,259	984
Pass EPV * Pass selection	0.0067	0.0011	–	–	–	–
Ball drive successful EPV	0.0128	0.0022	16	1e−4	153	57,441
Ball drive missed EPV	0.0072	0.0025	16	1e−4	153	57,441
Shot EPV	0.2421	0.0095	16	1e−3	231	72,455
Action selection probability	0.6454	–	32	1e−3	171	23,709
EPV	0.0078	0.0023	–	–	–	–

Additionally, the table presents the optimal value of the hyper-parameters, total number of parameters, and the number of predicted examples by second, for each of the models
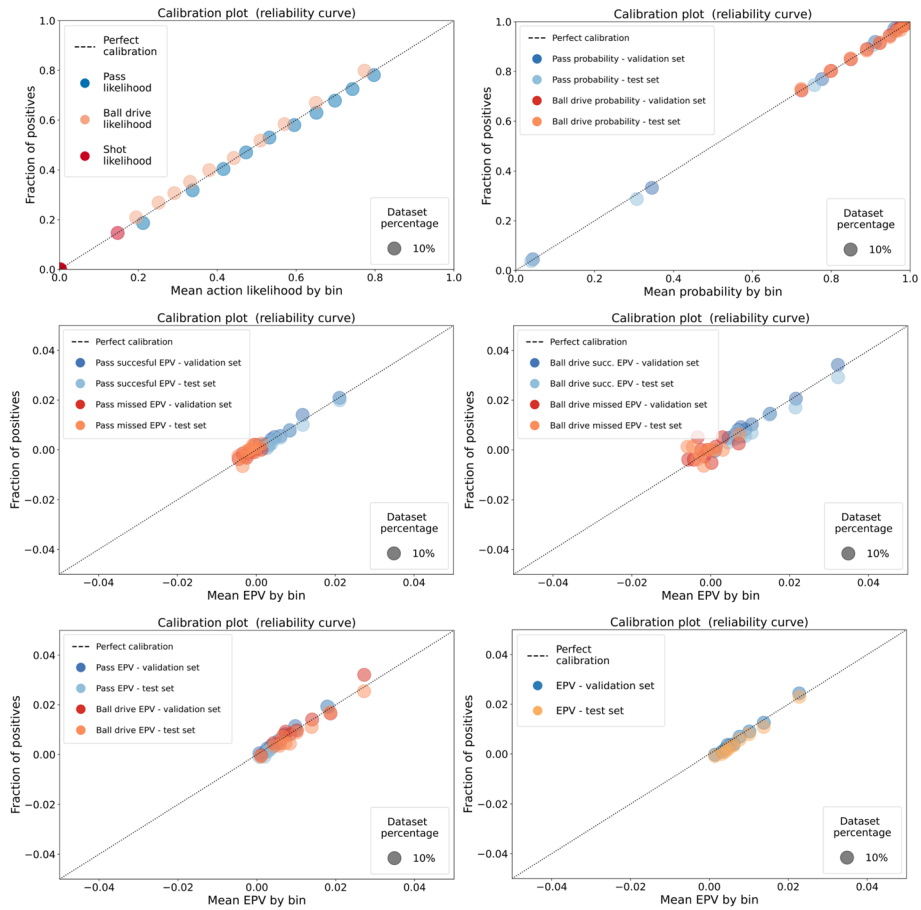
Fig. 5 Probability calibration plots for the action selection (top-left), pass and ball drive probability (top-right), pass (successful and missed) EPV (mid-left), ball drive (successful and missed) EPV (mid-right), pass and ball drive EPV joint estimation (bottom-left), and the joint EPV estimation (bottom-right). Values in the x-axis represent the mean value by bin, among 10 equally-sized bins. The y-axis represents the mean observed outcome by bin. The circle size represents the percentage of examples in each bin relative to the total examples for each model

Regarding the models' calibration, we can observe low ECE values along with all the models. Figure 5 presents a fine-grained representation of the probability calibration of each of the models. The x-axis represents the mean predicted value for a set of $K = 10$ equal-sized bins, while the y-axis represents the mean observed outcome among the examples within each corresponding bin. The circle represents the percentage of examples in the bin relative to the total number of examples. In these plots, we can observe that the different models provide calibrated probability estimations along their full range of predictions, which is a critical factor for allowing a fine-grained inspection of the impact that specific actions have on the expected possession value estimation. Additionally, we can observe the different ranges of prediction values that each model produces. For example, ball drive success probabilities are distributed more often above 0.5, while pass success probabilities cover a wide range between 0 and 1, showing that it is harder for a player to lose the ball

when keeping the ball than it is to lose the ball by attempting a pass towards another location on the field. The action selection probability distribution is heavily influenced by each action type's frequency, showing a higher frequency and broader distribution on ball drive and pass actions compared with shots. The joint EPV model's calibration plot shows that the proposed approach of estimating the different components separately and then merging them back into a single EPV estimation provides calibrated estimations. We applied post-training calibration exclusively to the pass success probability and the pass selection probability models, obtaining a temperature value of 0.82 and 0.5, respectively.

Having this, we have obtained a framework of analysis that provides accurate estimations of the long-term reward expectation of the possession, while also allowing for a fine-grained evaluation of the different components comprising the model.

7 Practical applications

In this section, we present a series of novel practical applications derived from the proposed EPV framework. We show how the different components of our EPV representation can be used to obtain direct insight in specific game situations at any frame during a match. We present the value distribution of different soccer actions and the contextual features developed in this work and analyze the risk and reward comprised by these actions. Additionally, we leverage the pass EPV surfaces, and the contextual variables developed in this work to analyze different off-ball pressing scenarios for breaking Liverpool's organized buildup. Finally, we inspect the on-ball and off-ball value-added between every Manchester City player (season 14/15) and the legendary attacking midfielder David Silva, to derive an optimal team that would maximize Silva's contribution to the team.

7.1 A real-time control room

In most team sports, coaches make heavy use of video to analyze player performance, show players their correctly or incorrectly performed actions, and even point out other possible decisions the player may have taken in a given game situation. The presented structured modeling approach of the EPV provides the advantage of obtaining numerical estimations for a set of game-related components, allowing us to understand the impact that each of them has on the development of each possession. Based on this, we can build a control room-like tool like the one shown in Figure 6, to help coaches analyze game situations and communicate effectively with players.

The control room tool presented in Figure 6 shows the frame-by-frame development of each of the EPV components. Coaches can observe the match's evolution in real-time and use a series of widgets to inspect into specific game situations. For instance, in this situation, coaches can see that passing the ball has a better overall expected value than keeping the ball or shooting. Additionally, they can visualize in which passing locations there is a higher expected value. The EPV evolution plot on the right shows that while the overall EPV is 0.032, the best possible passing option is expected to increase this value up to 0.112. The pass EPV added surface overlay shows that an increase of value can be expected by passing to the teammates inside the box or passing to the teammate outside the box. With this information and their knowledge on their team, coaches can decide whether to instruct the player to take immediate advantage of these kinds of passing opportunities or wait until better opportunities develop. Additionally, the player can gain a more visual
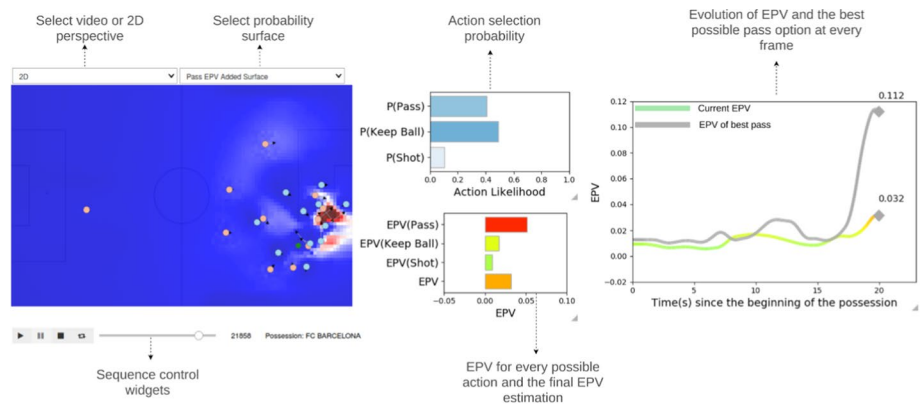
Fig. 6 A visual control room tool based on the EPV components. On the left, a 2D representation of the game state at a given frame during the match, with an overlay of the pass EPV added surface and selection menus to change between 2D and video perspective, and to modify the surface overlay. On the bottom-left corner, a set of video sequence control widgets. On the center, the instantaneous value of selection probability of each on-ball action, and the expected value of each action, as well as the overall EPV value. On the right, the evolution of the EPV value during the possession and the expected EPV value of the optimal passing option at every frame

understanding of the potential value of passing to specific locations in this situation instead of taking a shot. If the player tends to shoot in these kinds of situations, the coach could show that keeping the ball or passing to an open teammate has a better goal expectancy than shooting from that location.

This visual approach could provide a smoother way to introduce advanced statistics into a coaching staff analysis process. Instead of evaluating actions beforehand or only delivering hard-to-digest numerical data, we provide a mechanism to enhance coaches' interpretation and player understanding of the game situations without interfering with the analysis process.

7.2 Not all value is created (or lost) equal

There is a wide range of playing strategies that can be observed in modern professional soccer. There is no single best strategy found in successful teams from Guardiola's creative and highly attacking FC Barcelona to Mourinho's defensive and counter-attacking Inter Milan. We could argue that a critical element for selecting a playing strategy lies in managing the risk and reward balance of actions, or more specifically, which actions a team will prefer in each game situation. While professional coaches intuitively understand which actions are riskier and more valuable, there is no quantification of the actual distribution of the value of the most common actions in soccer.

From all the passes and ball drive actions described in Sect. 6.1, and the spatial and contextual features described in Sect. 4 we derived a series of context-specific actions to compare their value distribution. We identify passes and ball drives that break the first, second, or third line from the concept of dynamic pressure lines. We define an action (pass or ball drive) to be under-pressure if the player's pitch control value at the beginning of the action is below 0.4 and without pressure otherwise. A long pass is defined as a pass action that covers a distance above 30 meters. We define a pass back as passes where the destination location is closer to the team's goal than the ball's origin location. We count with

manually labeled tags indicating when a pass is a cross and when the pass is missed, from the available data. We identify lost balls as missed passes and ball drives ending in recovery by the opponent. For all of these action types, we calculate the added value of each observed action (EPV added) as the difference between the EPV at the end and the start of the action. We perform a kernel density estimation on the EPV added of each action type to obtain a probability density function. In Figure 7 we compare the density between all the action types. The density function value is normalized in the [0, 1] range by dividing by the maximum density value in order to ease the visual comparison between the distributions.

From Figure 7, we can gain a deeper understanding of the value distribution of different types of actions. From passes that break lines, we can observe that the higher the line, the broader the distribution, and the higher the extreme values. While passes breaking the first line are centered around 0 with most values ranging in [−0.01, 0.015], the distribution of passes breaking the third line is centered around 0.005, and most passes fall in the interval [−0.025, 0.05]. Similarly, ball drives that break lines present a similar distribution as passes breaking the first line. Regarding the level of spatial pressure on actions, we can see that actions without pressure present an approximately zero-centered distribution, with most values falling in a [−0.01, 0.01] range. On the other hand, actions under pressure present a broader distribution and a higher density on negative values. This shows both that there is more tendency to lose the ball under pressure, hence losing value, and a higher tendency to increase the value if the pressure is overcome with successful actions. Whether
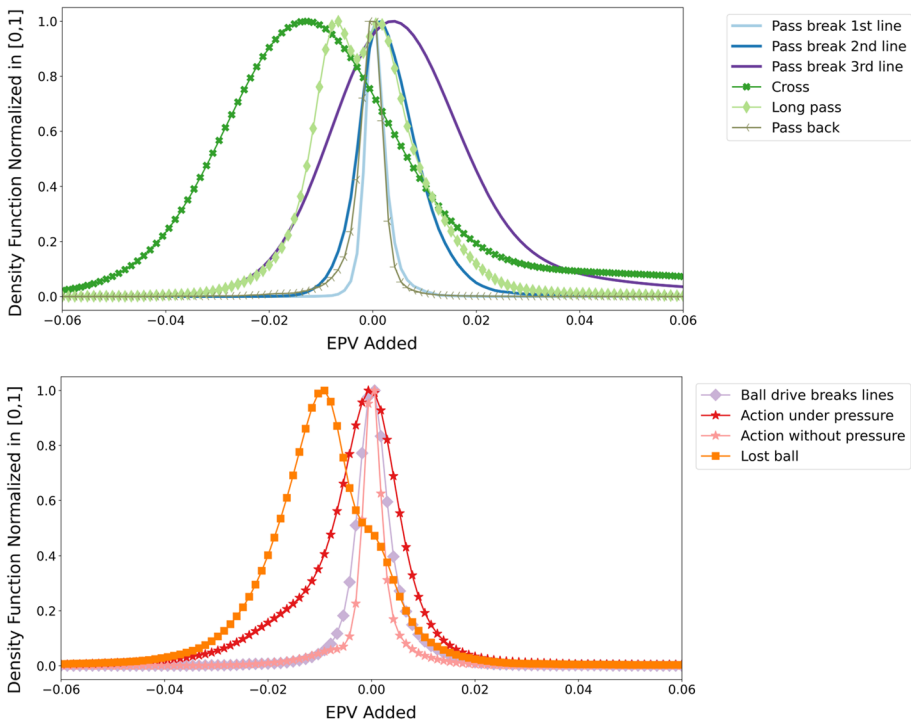

Fig. 7 Comparison of the probability density function of ten different actions in soccer. The density function values are normalized into the [0, 1] range. The normalization is obtained by dividing each density value by the maximum observed density value

crosses are a successful way for reaching the goal or not has been a long-term debate in soccer strategy. We can observe that crosses constitute the type of action with a higher tendency to lose significant amounts of value; however, it does provide a higher probability of high value increases in case of succeeding, compared to other actions. Long passes share a similar situation, where they can add a high amount of value in case of success but have a higher tendency to produce high EPV losses. For years, soccer enthusiasts have argued about whether passing backward provides value or not. We can observe that, while the EPV added distribution of passing back is the narrowest, near half of the probability lies on the positive side of the x-axis, showing the potential value to be obtained from this type of action. Finally, losing the ball often produces a loss of value. However, in situations such as being close to the opponent's box and with pressure on the ball carrier, losing the ball with a pass to the box might provide an increment in the expected value of the possession, given the increased chance of rebound.

7.3 Pressing liverpool

A prevalent and challenging decision that coaches face in modern professional soccer is how to defend an organized buildup by the opponent. We consider an organized buildup as a game situation where a team has the ball behind the first pressure line. When deciding how to press, a coach needs to decide first in which zones they want to avoid the opponent receiving passes. Second, how to cluster their players in order to minimize the chances of the opponent moving forward. This section uses EPV passing components and dynamic pressure lines to analyze how to press Brendan Rodgers' Liverpool (season 14/15).

We identify the formation being used every time by counting the number of players in each pressure line. We assume there are only three pressure lines, so all formations are presented as the number of defenders followed by the number of midfielders and forwards. For every formation faced by Liverpool during buildups, we calculate both the mean off-ball and on-ball advantage in every location on the field. The on-ball advantage is calculated as the sum of the EPV added of passes with positive EPV added. On the other hand, the off-ball advantage is calculated as the sum of positive potential EPV added. We then say that a player has an off-ball advantage if he is located in a position where, in case of receiving a pass, the EPV would increase. Figure 8 presents two heatmaps for every of the top 5 formations used against Liverpool during buildups, showing the distribution where Liverpool obtained on-ball and off-ball advantages, respectively. The heatmaps are presented as the difference with the mean heatmap in all of Liverpool's buildups during the season.

We will assume that the coach wants to avoid Liverpool playing inside its team block during buildups. We can see that when facing a 3-4-3 formation, Liverpool can create higher off-ball advantages before the second pressure line and manages to break the first line of pressure by the inside successfully. Against the 4-4-2, Liverpool has more difficulties in breaking the first line but still manages to do it successfully while also generating spaces between the defenders and midfielders, facilitating long balls to the sides. If the coaches' team does not have a good aerial game, this would be a harmful way of pressing. We can see the 4-3-3 is an ideal pressing formation for avoiding Liverpool playing inside the pressing block. This pressing style pushes the team to create spaces on the outside, before the first pressure line and after the second pressure line. In the second row, we can observe that Liverpool struggles to add value by the inside and is pushed towards the sides when passing. The 4-2-4 is the formation that avoids playing inside the block the most; however, it also allows more space on the sides of the midfielders. We can see that
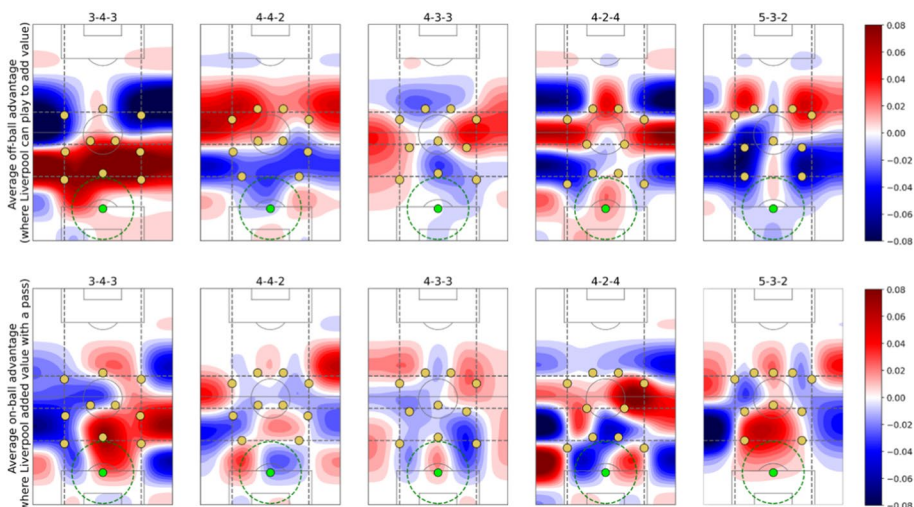
Fig. 8 In the first row, one distribution for every formation Liverpool's opponents used during Liverpool's organized buildups, showing the difference between the distribution of off-ball advantages and the mean distribution. The second row is analogous to the first one, presenting the on-ball EPV added distributions. The green circle represents the ball location (Color figure online)

Liverpool can take advantage of this and create spaces and make valuable passes towards those locations. If the coach has fast wing-backs that could press receptions on long balls to the sides, this could be an adequate formation; otherwise, 4-3-3 is still preferable. Finally, the 5-3-2 provides significant advantages to Liverpool that can create spaces both by the inside above the first pressure line and behind the defenders back, while also playing towards those locations effectively.

This kind of information can be highly useful to a coach to decide tactical approaches for solving specific game situations. If we add the knowledge that the coach has of his players' qualities, he can make a fine-tuned design of the pressing he wants his team to develop.

7.4 Growing around David Silva

Most teams in the best professional soccer leagues have at least one player who is the key playmaker. Often, coaches want to ensure that the team's strategy is aligned with maximizing the performance of these key players. In this section, we leverage tracking data and the passing components of the EPV model to analyze the relationship between the well known attacking midfielder David Silva and his teammates when playing at Manchester City in season 14/15. We calculated the playing minutes each player shared with Silva and aggregated both the on-ball EPV added and expected off-ball EPV added of passes between each player pair for each match in the season. We analyze two different situations: when Silva has the ball and when any other player has the ball and Silva is on the field. We also calculate the selection percentage, defined as the percentage of time Silva chooses to pass to that player when available (and vice versa). Figure 9 presents the sending and receiving maps involving David Silva and each of the two players with more minutes by position in the team. Every player is placed according to the most commonly used position in the league. Players represented by a circle with a solid
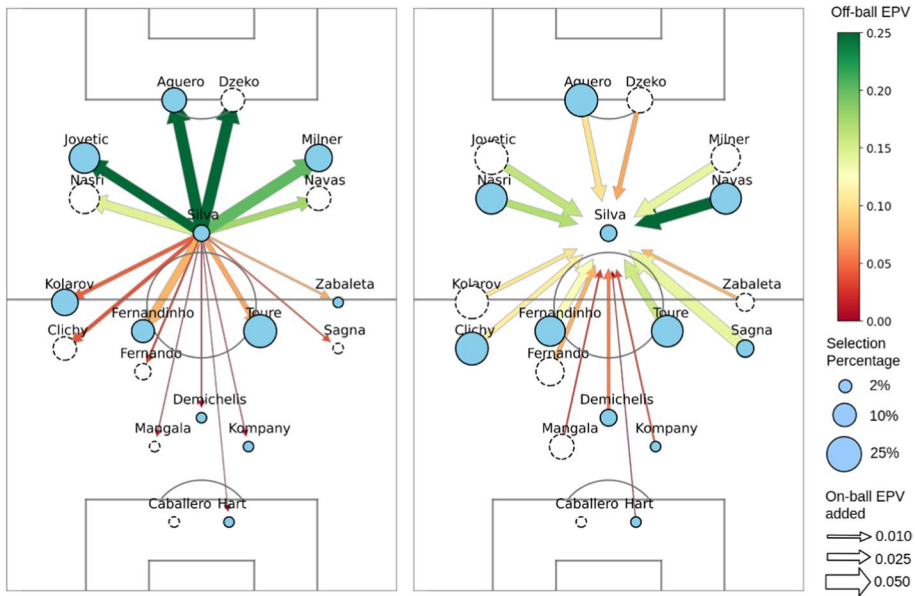
Fig. 9 Two passing maps representing the relationship between David Silva and each of the two players with more minutes by position in the Manchester City team during season 14/15. The figure on the left represents passes attempted by Silva, while the figure on the right represents passes received by Silva. The color of the arrow represents the average expected off-ball EPV added of the passes. The size of the circle represents the selection percentage of the destination player of the pass. Circles present a solid contour when that player is considered better for Silva than the teammate in the same position. The size of the arrow represents the mean on-ball EPV added of attempted passes. Players are placed according to their highest used position on the field. All metrics are normalized by minutes played together and multiplied by 90 minutes (Color figure online)

contour have the higher sum of off-ball and on-ball EPV in each situation than the teammate assigned for the same position, presented with a dashed circle. The size of the circle represents the selection percentage of the player in each situation. We represent off-ball EPV added by the arrows' color, and on-ball EPV added of attempted passes by the arrow's size.

We can see that both the wingers and forwards generate space for Silva and receive high added value from his passes. However, the most frequently selected player is the central midfielder Yaya Touré, who also looks for Silva often and is the midfielder providing the highest value to him. Regarding the other central midfielder, Fernandinho has a better relationship with Silva in terms of received and added value than Fernando. Silva shows a high tendency to play with the wingers; however, while Milner and Jovetic can create space and receive value from Silva, Navas and Nasri find Silva more often, with higher added value. Based on this, the coach can decide whether he prefers to lineup wingers that can benefit from Silva's passes or wingers, increasing Silva's participation in the game. A similar situation is presented with the right and left-backs. Additionally, we can observe that Silva tends to be a highly preferable passing option for most players. This information allows the coach to gain a deeper understanding of the effective off-ball and on-ball value relationship that is expected from every pair of players and can be useful for designing playing strategies before a match.

8 Discussion

This paper presents a comprehensive approach for estimating the instantaneous expected value of possessions in soccer. One of the main contributions of this work is showing that by deconstructing a single expectation into a series of lower-level statistical components and then estimating each of these components separately, we can gain greater interpretation insight into how these different elements impact the final joint estimation. Also, instead of depending on a single-model approach, we can make a more specialized selection of the models, learning approach, and input information that is better suited for learning the specific problem represented by each sub-component of the EPV decomposition. The deep learning architectures presented for the different passing components produce full probability surfaces, providing rich visual information for coaches that can be used to perform fine-grained analysis of player and team performances. We show that we can obtain calibrated estimations for all the decomposed model components, including the single-value estimation of the expected possession value of soccer possessions. We develop a broad set of novel spatial and contextual features for the different models presented, allowing rich state representations. Finally, we present a series of practical applications showing how this framework could be used as a support tool for coaches, allowing them to solve new upcoming questions and accelerating the problem-solving necessities that arise daily in professional soccer.

We consider that this work provides a relevant contribution to improving the practitioners' interpretation of the complex dynamics of professional soccer. With this approach, soccer coaches gain more convenient access to detailed statistical estimations that are unusual in their practice and find a visual approach to analyze game situations and communicate tactics to players. Additionally, on top of this framework, there is a large set of novel research that can be derived, including on-ball and off-ball player performance analysis, team performance and tactical analysis for pre-match and post-match evaluation, player profile identification for scouting, young players evolution analysis, match highlights detection, and enriched visual interpretation of game situations, among many others.

9 Limitations and future work

In this section we present the main limitations of the EPV framework and provide guidelines for improving the model in future work.

Player and team-specific features These kind of features, such as player skills or team-level playing tendencies could provide enriched information for producing more accurate estimation in specific game situations. Some examples of these features are: a player-passing skill feature (e.g. average pass completion) for the pass probability model, an action-selection feature indicating the either the player's or the team's tendency to pass, keep the ball or shoot, or a player-level shooting skill. Alternatively, a series of components could be added to the different neural network architectures proposed in this work to learn team-specific features, within the same learning process. For example, a group of neurons encoding team passing-style could be added to produce pass-likelihood surfaces adjusted to each team in the dataset. However, it is important to notice that considering player-level or team-level features could make the attribution of value more challenging.

Meta-information of the game and team state The overall estimation of EPV could benefit from including information such as the time when actions occur, the current score, an estimation of the match importance, known rivalry between the two teams, or an estimation of a team's mental pressure at any given time (Bransen et al. 2019).

Training with a more recent and broader dataset The current model could be improved by incorporating tracking data from different leagues other than the English Premier League, in order to capture soccer spatiotemporal dynamics more broadly. Also, training the model with more recent tracking data might help capturing up-to-date characteristics of soccer such as a higher physical conditions of players or a tendency to prefer short passes over long passes.

Training set-pieces separately The components related to passes and shots could be improved by differentiating between open-play and set-pieces, at an implementation level.

Appendix 1: List of spatial and contextual features features

Tables 5 and 6 describe the complete set of features used as input for each presented model. The concept type column refers to the general feature grouping described in Sect. 4, including a prefix indicating whether the feature is a spatial feature (SP), a contextual feature (CX), or other types (OT). Model names are presented with acronyms, including: pass success probability (PP), pass selection probability (PS), pass success and missed EPV (PE), ball drive probability (DP), ball drive success and missed EPV (DE), action selection probability (AS), and shot EPV (SE). For PP, PS, and PE models, the input features are either sparse or full matrix of 104×68. When the feature description indicates the value is set of every location, this input will correspond to a full matrix; otherwise, it corresponds to a sparse matrix. For the rest of the models, each feature is provided as a single variable. We refer to the team in control of the ball as the *attacking team*, and its players as the *attacking players*. We refer to the other team as the *defending team*, and its players as the *defending players*. All the features are normalized, assuming a left to right attacking direction of the team in control of the ball (attacking team).

Appendix 2: Pitch control and influence model

The concepts of pitch influence and pitch control are adapted from a recent statistical approach based on modeling players' reachability surface through normal distributions (Fernandez and Bornn 2018). The pitch influence I, shown below at expression 10 is a normally distributed random variable whose mean vector and covariance matrix are adjusted to account for players' velocity and ball location. Let p_i be player's i location in 1 second and let $f_i(p, t)$ be the value of the probability density function of I related to player i at location p and time t, we obtain the player's influence value at location $I_i(p, t)$ following expression 11.

$$I \sim \mathcal{N}(\mu, \Sigma) \tag{10}$$

$$I_i(p, t) = \frac{f_i(p, t)}{f_i(p_i, t))}. \tag{11}$$

This influence value is normalized in the [0,1] range and provides a degree of influence for a given player. Having a quantification of individual players influence, we can calculate

Table 5 Set of spatial features used as input for each presented model

Concept type	Feature	PP	PS	PE	DP	DE	AS	SE
SP - (x,y) location	1 on attacking players' location (x,y)	✓	✓	✓				✓
SP - (x,y) location	Ball location (x)	✓	✓		✓	✓	✓	
SP - (x,y) location	1 on defending players' location (x,y)	✓	✓	✓				
SP - (x,y) location	1 if the ball is closer to the goal than the opponent's goalkeeper	✓	✓					✓
SP - Velocity	Attacking team players' speed (m/s) (x)	✓	✓	✓				
SP - Velocity	Attacking team players' speed (m/s) (y)	✓	✓	✓				
SP - Velocity	Defending team players' speed (m/s) (y)	✓	✓	✓				
SP - Velocity	Defending team players' speed (m/s) (y)	✓	✓	✓				
SP - Angle	Angle between every location and the goal	✓	✓	✓				
SP - Angle	Angle between the ball and the goal	✓	✓		✓	✓	✓	✓
SP - Angle	Sine of the angle between every location and the ball location	✓	✓					
SP - Angle	Cosine of the angle between every location and the ball location	✓	✓					
SP - Angle	Sine of the angle between the ball carrier velocity vector and every other location	✓	✓					
SP - Angle	Cosine of the angle between the ball carrier velocity vector and every other location	✓	✓					
SP - Distance	Distance between every location and the goal	✓	✓	✓				
SP - Distance	Distance between every location and the ball	✓	✓	✓				
SP - Distance	Distance between the ball and the goal				✓	✓	✓	✓
SP - Distance	Distance between the ball and the goalkeeper in y-axis				✓	✓	✓	✓
SP - Distance	Distance between the ball and the goalkeeper				✓	✓	✓	✓
SP - Pitch control	Pitch control of the attacking team at the ball location				✓	✓	✓	
SP - Pitch influence	Pitch influence of the defending team at the ball location				✓	✓	✓	

The concept type column includes a prefix indicating the feature belongs to the spatial feature type (SP). For the rest of the columns a checkmark indicates the models where the feature is used, including: pass success probability (PP), pass selection probability (PS), pass success and missed EPV (PE), ball drive probability (DP), ball drive success and missed EPV (DE), action selection probability (AS), and shot EPV (SE)

Table 6 Contextual features and other feature types used as input for each presented model

Concept type	Feature	PP	PS	PE	DP	DE	AS	SE
CX - Dynamic pressure lines	Index of the closest attacking team line to every location			✓				
CX - Dynamic pressure lines	Index of the closest attacking team line to the ball location				✓	✓	✓	
CX - Dynamic pressure lines	Index of the closest defending team line to every location			✓				
CX - Dynamic pressure lines	Index of the closest defending team line to the ball location				✓	✓	✓	
CX - Outplayed players	Number of attacking team's players between the ball and every other location			✓				
CX - Outplayed players	Number of defending players between the ball and every other location			✓				
CX - Outplayed players	Number of attacking players between the opponent's goal and every other location			✓				
CX - Outplayed players	Number of players of the defending team between the opponent's goal and every other location			✓				
CX - Interceptability	Number of defending players inside the triangle formed between the ball location and the posts of the opponent's goal							✓
CX - Interceptability	Number of defending players located less than 3 meters away from the ball location						✓	✓
CX - Event-based xG	Expected goals based on the action location and the angle to the goal					✓	✓	✓
OT - Type	1 of action is attempted with the head			✓				
OT - Probability	Pass probability surface					✓		
OT - Probability	Ball drive probability				✓			

The concept type column includes a prefix indicating whether the feature is a contextual feature (CX) or other types (OT). For the rest of the columns a checkmark indicates the models where the feature is used, including pass success probability (PP), pass success probability (PS), pass selection probability (DP), ball drive probability (DP), ball drive success and missed EPV (DE), action selection probability (AS), and shot EPV (SE)
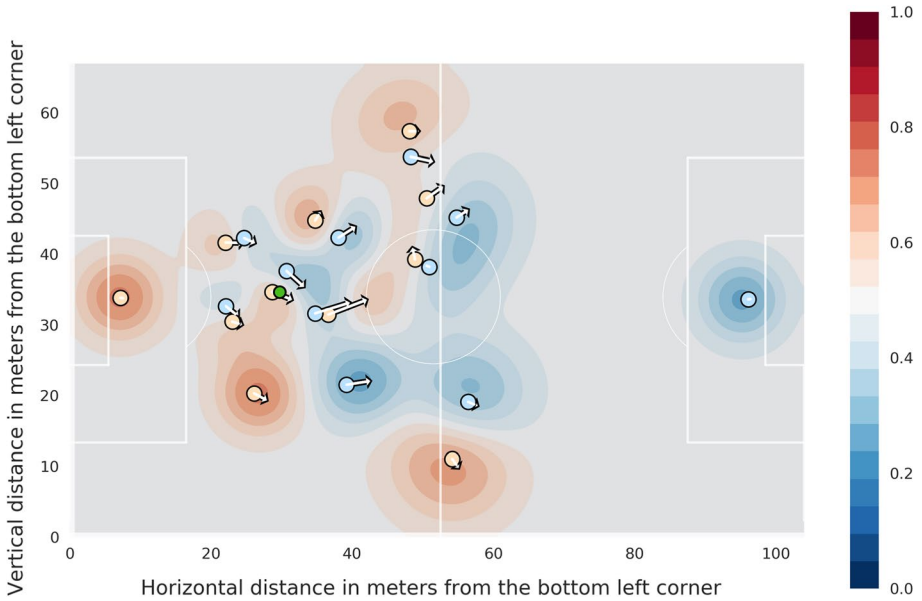
Fig. 10 Surface representing the attacking team's pitch control. The dots correspond to players' locations where the attacking team's players appear in yellow and the opponent team's players in blue. The ball location is represented as a green circle. White arrows show the direction of the player's velocity vector, ending at the expected location in one second (Color figure online)
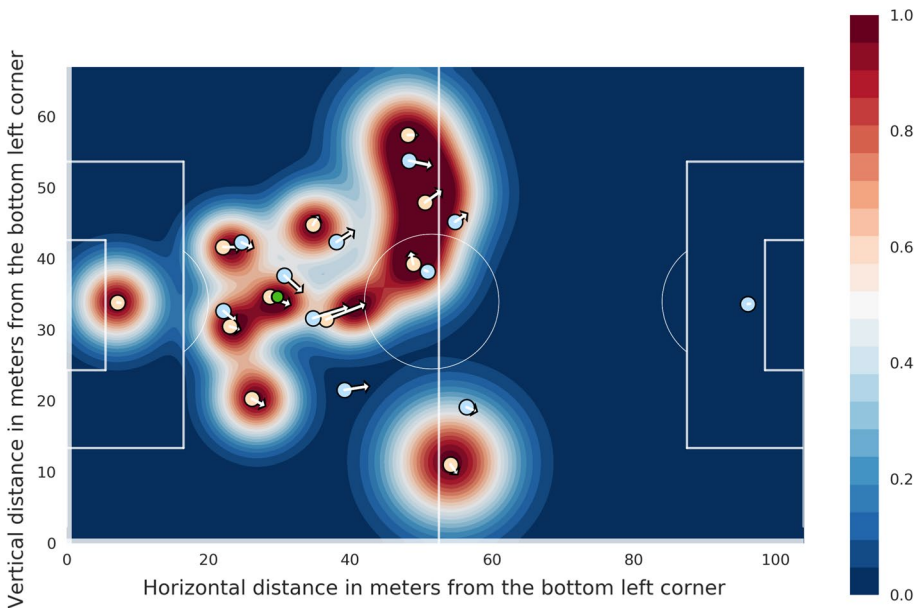

Fig. 11 The sum of player pitch influence in every location for every player in the attacking team. The dots correspond to players' locations where the attacking team's players appear in yellow and the opponent team's players in blue. The ball location is represented as a green circle. White arrows show the direction of the player's velocity vector, ending at the expected location in one second (Color figure online)

pitch control of a team *PC* as the difference between the added influence of the possession
team's players and the influence of the opponent team's players, at any given location, as
shown in equation 12

$$PC(p, t) = \sigma(\gamma(\lambda_1 \sum_i I_i(p, t) - \lambda_2 \sum_j I_j(p, t)))$$

(12)

where σ is the logistic function, λ_1 and λ_2 are weight parameters to allow balancing each
team's overall influence, and γ is a shrinking factor for the input logistic function. Fig-
ure 10 presents this probabilistic pitch control surface on a given soccer situation, while
Figure 11 presents the influence surface of the attacking team players. Pitch influence and
pitch control provide a rich summary of players' spatial distribution and impact along the
playing surface and can be used to enrich the information on locations where players are
not directly present but are having a certain influence from soccer's tactical perspective. In
this work we set these parameters to the following values $\lambda_1 = 1$, $\lambda_2 = 1$, and $\gamma = 1$.

Appendix 3: Dynamic pressure lines model

As described in Sect. 4.2, we consider a contextual factor identifying the team player's
alignment in a given instance. The alignment of players is often observed in soccer
through concepts such as team formation or the identification of forwards, midfielders,
and defenders. However, this organization of players is manifested dynamically during
the game. Instead of following a strict and predefined positioning, players tend to adapt
their location to the specific situation of a given time instance in the game. Specifically,
while defending, players tend to align within groups of pressure across the field. We call
this alignment group dynamic pressure lines.

Extending from this idea, we first define dynamic pressure lines with higher generality
as the centroids of a number k of clusters representing hard partitions for players of the
same team, where the intra-cluster distance is minimized, and the inter-cluster distance is
maximized. If the clustering is based on the breadth-wise location of players (x-axis), we
call them vertical dynamic pressure lines, and if it is based on the depth-wise location of
players (y-axis), we call them horizontal dynamic pressure lines.

Definition 2 Given a set of n player locations $P = \{p_1, \ldots, p_n\}$, and let $d(p, q)$ be the
Euclidean distance between p and q, and $D(L_1, L_2)$ the distance between clusters L_1 and L_2,
the set L of dynamic pressure lines is conformed by the average locations of the player's
belonging to the complete-linkage clustering of P in k partitions, such that for $L_1, L_2 \in L$
and $D(L_1, L_2) = \max_{p^{L_1}, p^{L_2}} d(p^{L_1}, p^{L_2})$. When $p_i = (x_i, y_i) = (x_i, 0)$ we call L the set of verti-
cal dynamic pressure lines, and when $p_i = (x_i, y_i) = (0, y_i)$ we call L the set of horizontal
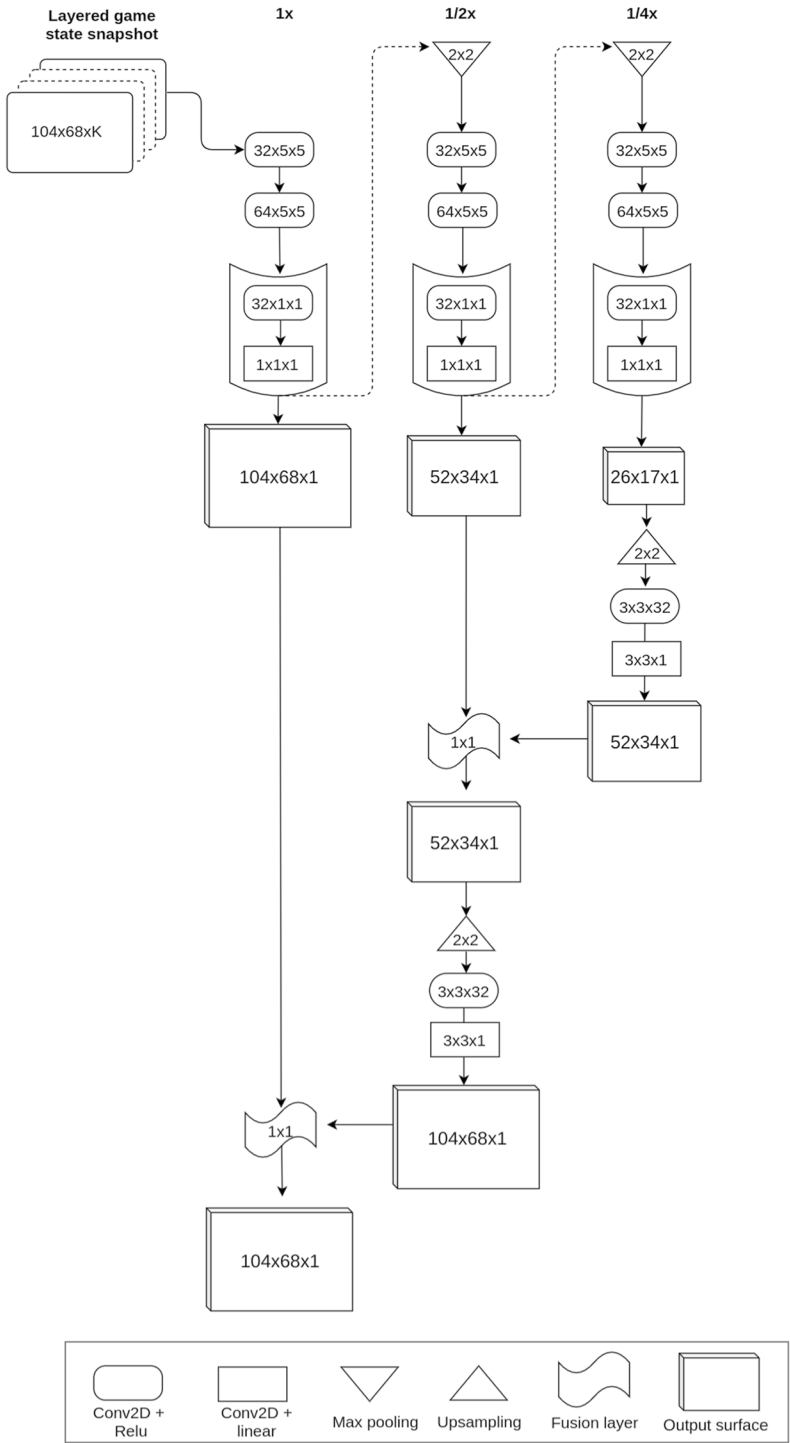dynamic pressure lines.

Table 7 Total count of matches and shot events included within the event data dataset

Data type	Source	# Total	# Training	# Test	% Goals
Match	Event	4679	3509	1170	–
Shot	Event	117,948	87,980	30,645	10.4

In this work, we set $k = 3$ to identify vertical pressure lines, which conceptually represent forwards, midfielders, and defenders. For horizontal pressure lines, we set $k = 3$, which will tend to define the breadth-wise borderlines of the team formation block and split the inside of the block into two parts.

Appendix 4: SoccerMap architecture

Fully convolutional networks focus on estimating a full prediction surface from the input data, contrasting with the typical application of convolutional neural networks for classification, where outcomes tend to be either binomially or multinomially distributed. Such is the case of the image segmentation problem where, given an image, we intend to estimate a pixel-level correspondence to multiple objects present in the input image (Long et al. 2015; Yu and Koltun 2015; Pathak et al. 2015). SoccerMap is modeled as a fully convolutional network-based architecture. In its design, SoccerMap uses several components of successful architectures in other application fields such as convolutional filters, pooling and upsampling, fusion layers, and activation layers. Figure 12 presents the standard architecture for the feature extractor block, for a soccer field representation of sizes 104×68. First, the input data constituted by the layered data snapshot Υ_t goes through two layers of 32 and 64 convolutional filters, respectively, with 5×5 activation fields and stride of 1, and then is downsampled to 1/2x using max-pooling. This process is repeated twice, so three outputs of convolutional filters are produced at 1x, 1/2x, and 1/4x sampling scales. Each of the three outputs at each scale is fed to convolutional prediction layers that produce a prediction matrix at each sampling scale. The predictions at each scale are merged (previous upsampling to match dimensions) through convolutional layers with linear activation, called fusion layers. The output is fed to a final prediction layer constituted by 1×1 convolutional filters of stride 1, producing a 104×68×1 surface. The combination of layers at different resolutions allows capturing relevant information at both local and global levels, with the expectation of producing location-wise predictions that are spatial-aware. This approach is inspired by a nonlinear feature hierarchy called deep jet (Long et al. 2015).

Appendix 5: Baseline expected goals model

In both the action selection and goal expectation models we include a feature representing a general estimation of the goal expectation given a shot is taken, based on event data. In order to produce a calibrated baseline estimation of expected goals, we use a wide dataset of event data provided by *OPTA*, which contains 117, 948 shot events and 12, 266 goals as detailed in Table 7. Despite only providing the location and time of observed shots, this dataset is considerably larger than the 13, 735 shots available in the tracking data dataset

(see Sect. 6.1). Event data has been used successfully in previous work to obtain a calibrated estimation of expected goals (Eggels 2016).

We use a set of spatial features by the event location and the distance and angle between the ball location and the goal. Contextual features are composed of a one-hot encoded vector indicating the attacking type at the moment of the event (open-play, set-piece, free-kick, corner, penalty), and a boolean variable indicating whether the action is taken with the head or not. The matches are split into a training and test set. For every shot in the dataset, we label the outcome as 1 for the shots resulting in a goal, and 0 otherwise. We build the model using the extreme gradient boosting algorithm XGBoost (Chen and Guestrin 2016), and we perform an exhaustive grid-search on the following hyper-parameters of the model: number of trees ($\{50, 100, 250\}$), learning rate ($\{1e-3, 1e-2, 1e-1\}$), and maximum depth ($\{3, 5, 10\}$). Model selection is performed through a K-fold cross-validation procedure on the training set, with $K = 10$. All the features are standardized, obtaining a scaled feature set where each variable has a mean of 0 and a unitary standard deviation.

The best model presented a log loss value of 0.2540 and a calibration ECE value of 0.00594, in the test set. The parameters of the best model where: 100 trees, a maximum depth of 3, and a learning rate of $1e-1$.

Availability of data and materials Proprietary spatiotemporal tracking data from *STATS, LLC* and *Stats Perform*, has been used in this project.

Code availability Source code is not provided in this submission.

Declarations

Conflict of interest FC Barcelona, Polytechnic University of Catalonia, Sacramento Kings, Simon Fraser University, Los Angeles Dodgers, Zelus Analytics.

References

Bransen, L., Robberechts, P., Van Haaren, J., & Davis, J. (2019). Choke or shine? Quantifying soccer players' abilities to perform under mental pressure. In: *Proceedings of the 13th MIT sloan sports analytics conference* (pp. 1–25).

Bransen, L., & Van Haaren, J. (2018). Measuring football players on-the-ball contributions from passes during games. In: *International Workshop on machine learning and data mining for sports analytics* (pp. 3–15) Springer.

Cervone, D., D'Amour, A., Bornn, L., & Goldsberry, K. (2016). A multiresolution stochastic process model for predicting basketball possession outcomes. *Journal of the American Statistical Association, 111*(514), 585–599.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). ACM.

Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). Actions speak louder than goals: Valuing player actions in soccer. In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 1851–1861).

Eggels, H. (2016). Expected goals in soccer: Explaining match results using predictive analytics. In: *The Machine Learning and Data Mining for Sports Analytics workshop*, p 16

Fernandez, J., & Bornn, L. (2018). Wide open spaces: A statistical technique for measuring space creation in professional soccer. In: *Sloan Sports Analytics Conference*

Fernández, J., & Bornn, L. (2020). Soccermap: A deep learning architecture for visually-interpretable analysis in soccer. arXiv preprint arXiv:201010202.

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In: *Proceedings of the 34th international conference on machine learning-Volume 70*, JMLR. org (pp. 1321–1330).

Gyarmati, L., & Stanojevic, R. (2016). Qpass: A merit-based evaluation of soccer passes. arXiv preprint arXiv:160803532.

Hubáček, O., Šourek, G., & Železný, F. (2018). Deep learning from spatial relations for soccer pass prediction. In *International workshop on machine learning and data mining for sports analytics* (pp. 159–166). Springer.

Kingma, D.P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:14126 980.

Link, D., Lang, S., & Seidenschwarz, P. (2016). Real time quantification of dangerousity in football using spatiotemporal tracking data. *PLoS ONE, 11*(12), e0168768.

Liu, G., & Schulte, O. (2018). Deep reinforcement learning in ice hockey for context-aware player evaluation. arXiv preprint arXiv:180511088.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431–3440).

Lucey, P., Bialkowski, A., Monfort, M., Carr, P., & Matthews, I. .(2014). Quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. In *Proceedings of 8th annual MIT sloan sports analytics conference* (pp. 1–9).

Pathak, D., Krahenbuhl, P., & Darrell, T. (2015). Constrained convolutional neural networks for weakly supervised segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 1796–1804).

Power, P., Ruiz, H., Wei, X., & Lucey, P. (2017). Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1605–1613). ACM.

Rudd, S. (2011). A framework for tactical analysis and individual offensive production assessment in soccer using markov chains. In *New England symposium on statistics in sports*. http://nessis.org/nessis11/rudd.pdf.

Singh, K. (2019). Introducing expected threat (XT). https://karun.in/blog/expected-threat.html. Accessed: 2020-10-16.

Spearman, W. (2018). Beyond expected goals. In *Proceedings of the 12th MIT sloan sports analytics conference*.

Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv: 151107122.

Yurko, R., Matano, F., Richardson, L. F., Granered, N., Pospisil, T., Pelechrinis, K., & Ventura, S.L. (2020). Going deep: models for continuous-time within-play valuation of game outcomes in American football with tracking data. *Journal of Quantitative Analysis in Sports 1(ahead-of-print)*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.