What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data

Charbel Merhej
University of Southampton
United Kingdom
cm2u19@soton.ac.uk

Ryan Beal
University of Southampton
United Kingdom
ryan.beal@soton.ac.uk

Tim Matthews
Sentient Sports
United Kingdom
tim.matthews@sentientsports.com

Sarvapali Ramchurn
University of Southampton
United Kingdom
sdr1@soton.ac.uk

ABSTRACT

Objectively quantifying the value of player actions in football (soccer) is a challenging problem. To date, studies in football analytics have mainly focused on the attacking side of the game, while there has been less work on event-driven metrics for valuing defensive actions (e.g., tackles and interceptions). Therefore in this paper, we use deep learning techniques to define a novel metric that values such defensive actions by studying the threat of passages of play that preceded them. By doing so, we are able to value defensive actions based on what they prevented from happening in the game. Our Defensive Action Expected Threat (DAxT) model has been validated using real-world event-data from the 2017/2018 and 2018/2019 English Premier League seasons, and we combine our model outputs with additional features to derive an overall rating of defensive ability for players. Overall, we find that our model is able to predict the impact of defensive actions allowing us to better value defenders using event-data.

CCS CONCEPTS

• **Applied computing**; • **Computing methodologies → Model development and analysis**; *Neural networks*;

KEYWORDS

Sports Analytics; Football; Deep Learning; Neural Networks; Applied Machine Learning; Defensive Actions

1 INTRODUCTION

Valuing the actions of humans and agents in the real-world is a problem in many industries. By assigning values to actions committed, we can help evaluate the performance of agents and aid the learning and improvement of future actions. To date, examples of work that explore methods for valuing actions are shown in industrial optimisation [18], agent negotiation [14] and sports analytics [6, 23]. Although all these papers aim to assign value to what a human or agent has helped happen, there are many domains in the real-world where some humans/agents are tasked with preventing actions from happening (e.g., in security games agents aim to protect bases or ports and prevent attacks [25]).

One domain where this is also key is in the sports world. In games such as football, American football and basketball, there are players in each team who are tasked with stopping actions (e.g., an American footballer defensive player aims to prevent touchdowns). Therefore in this paper, we focus our attentions on valuing the actions of defenders in Association Football (soccer).[1]

Due to the rapid advancements in technology in football, the past decade has seen a growing line of data science research and football analysis. Multiple companies[2] have implemented advanced methods for collecting detailed match actions in the form of event stream data, and more recently in the form of tracking data. The growing availability of these datasets have led to wider applications of artificial intelligence (AI) techniques in football analytics such as machine and deep learning [4].

An area where this has been particularly true is in the creation of new models for descriptive player performance metrics. This is due to the immense benefits it can provide for football clubs. As explained in [4], measuring player performance is an important factor in the decision making processes in team sport. However, due to the low scoring and dynamic nature of football, this presents a unique challenge to AI to value actions that often do not impact the scoreline directly but have important longer-term effects [6]. In recent years, multiple successful metrics have been adopted and used universally. One such example is the "Expected Goals (xG)" metric [22], which provides the probability of a shot resulting in a goal in comparison to other similar shots.

---

[1] Referred to as just "football" throughout this paper (as it is known outside the U.S.)
[2] Examples of these companies include but are not limited to: STATSPerform, WyScout and StatsBomb.

A key limitation of such event-based action models is the lack of application for defensive actions. This is an inherent problem with event-based data due to the fact that defensive actions mainly prevent events from completion and are therefore a challenge to value. This leads to a lot of players not getting truly acknowledged for their defensive duties. To combat this problem, we propose a novel data-driven model that focuses on defensive and out of possession actions by quantifying these actions in a unique and accurate way. What differentiates this model from other traditional football metrics [2, 3, 6, 22] is that, instead of studying actions that lead to events happening, our model values defensive actions that stop events from occurring. Thus, this paper advances the state of the art in the following ways:

(1) Introduces a novel model to value defensive actions in football named DAxT (Defensive Action Expected Threat). The model is targeted at the football analytics community and brings together new research in football with deep learning techniques to accurately value defensive actions based on what they have stopped from happening.

(2) Using real-world data from 760 real-world football games from the past two seasons of the English Premier League (EPL), we train and test our model and find that we can accurately predict the impact of future events to an MAE of 0.015.

(3) We use this model to identify the top football defenders in the EPL and validate this against leading providers of football performance statistics.

The rest of this paper is structured as follows. In Section 2 we review the literature around player performance metrics in football and introduce basic concepts about defending. Section 3 formally defines the goal of our model. In Sections 4 and 5 we discuss how we created and solved the model and illustrate the multiple experiments conducted to validate it. Section 6 shows the results of applying our model on real-world datasets. Finally, in Section 7 we discuss the outcomes obtained and we conclude in Section 8.

2 BACKGROUND

In this section, we present an overview of related work regarding player performance metrics. Moreover, we introduce a background on basic concepts in football and the importance of defensive actions.

2.1 Related Work

One of the key areas in football analytics is the study of new player performance metrics. Because of the fast-paced and continuous nature of the game, player performance analysis has been a difficult task in football. As mentioned in [11], "Baseball, for example, is a less fluid sports than soccer. It is easier to break down a baseball game into discrete events to be analyzed. The same cannot be said about soccer". However, with the increased interest in football analytics and advances in technology, more researchers have focused their studies on this specific area.

As previously mentioned, xG is a metric used to describe the quality of a certain shot. To accomplish this, the authors in [15] analysed the spatio-temporal patterns of the ten-second window of play before a shot for 10,000 shots. In addition to the historical

data, they used strategic features such as defender proximity, speed of play, shot location, and game phase to determine the likelihood of a goal being scored. A similar study was done in [22], where the authors used distance of the shot taken from goal, the angle in relation to goal and the type of assist as features to calculate xG. The model proved accurate for mid-table teams, whereas the top table teams and lower table teams over and under achieved respectively. Both papers mentioned above used logistic regression in their models. The different features that could be used for xG models are highlighted in Figure 1 [5].
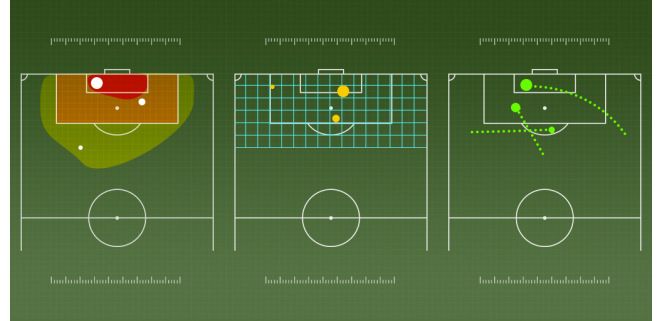

Figure 1: Distance to goal, angle, and type of assist as features used to calculate xG

Since xG only measures scoring actions, players on the pitch who are not tasked to score goals cannot be evaluated using this metric. This is why the authors in [7] defined a metric called "Expected Possession Value" (EPV), that quantifies the expected outcome at every moment in a possession. Their model studies passes and turnover probabilities, action likelihood, and pass/ball-drive value expectations that were estimated independently using logistic regression, convolutional neural network and deep neural networks respectively.

Contrarily, instead of measuring periods of possession, the value of each action in a game was studied in [6]. Their VAEP (Valuing Actions by Estimating Probabilities) model calculates this value $V(a_i)$ by estimating the change in probability of scoring and conceding a goal after every action. This is shown in Equation 1.

$$V(a_i) = \Delta P_{scores(a_i,x)} + (-\Delta P_{concedes(a_i,x)}) \tag{1}$$

where $\Delta P_{scores(a_i,x)}$ is the offensive value of team x, denoting the change in probability of scoring after an action a_i is committed, and $(-\Delta P_{concedes(a_i,x)})$ is the defensive value of team x, denoting the change in probability of conceding after a_i is committed.

The VAEP metric performed better than traditional player performance metrics and successfully identified promising young players and minor league talents. To estimate the probabilities mentioned above, the authors used features such as the action type, the location of action start and end, time elapsed, etc. Moreover, since the probability of one team scoring is equal to the probability of the other conceding, only two probability estimates were needed for Equation 1. The task was then divided into two separate binary classification problems, where the authors trained both models using CatBoost algorithm.

What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data

KDD '21, August 14–18, 2021, Singapore

Another important player performance metric is the Expected Threat (xT) [26], which is explained in further details in Section 4.1, as it directly correlates to our model.

2.2 Defending in Football

Similar to some other team sports, football is an invasion game where the whole team attacks when in possession of the ball, and defends when out of possession. Even though scoring a goal directly contributes to the scoreline, not conceding one is equally as important. Although the main goal of defending is for teams not to concede goals, the way it happens varies according to teams' different tactics and styles of play. Some teams sit deep in their half of the pitch (field) and stack up defenders, while others press high on the field to gain possession back and attack.

There are two main defensive actions that happen in a football game: Interceptions and Tackles. An interception is when the defending player intentionally intercepts a pass by moving into the line of the intended ball direction[3]. According to [19], intercepting a pass takes the ability to anticipate, read the play and keep the receiving player guessing. Getting the ball after a misplaced pass is also counted as an interception.

A tackle can be of 2 different types [31]. A blocking/standing tackle is one where the defender remains on his feet. It is generally used when the player with the ball is coming directly towards the defender. The other type is the sliding tackle, where the defender is off his feet and slides to get the ball [28].

Due to a strong focus in football analysis on attacking metrics that lead to direct contributions, there is a lack of research on the defensive actions mentioned above. In this paper, we aim to quantify different interceptions and tackles happening on the pitch by using our DAxT model - how we model these actions is explained in the following section.

3 MODELLING DEFENSIVE ACTIONS

In this paper, our goal is to model defensive actions in football to be able to better value the impact they have on the overall game. When modelling attacking actions in games (such as in [6, 22]), the effect of a specific action can be seen according to what happened next and the impact it had on the game. However, the same cannot be said about defensive actions as they usually prevent further actions from happening. Therefore, a logical way to value defensive actions is by assessing what they have prevented from happening and predicting what would have happened if not for that defensive action. By doing so, we present a model for valuing actions that prevent other actions from happening. Although in this paper we focus on football as an application domain, this model could be applied elsewhere where it is important to understand the value of players or agents who prevent actions from happening in games e.g., security games [9, 10] and emergency response [20, 21].

In football, events (such as passes, shots and tackles) happen in a sequence or passage of play. We define a sequence of events as $S = \{e_1, e_2, \ldots, e_N\}$ where S is the sequence, e is an event and N is the number of events in the sequence. Each sequence of events can end in a number of ways, for example the ball may go out of play, there may be a foul leading to a free kick, or a defensive action DA

[3]https://www.mlssoccer.com/glossary

occurs for the opposition to win the ball back from the attacking team. Two key types of defensive actions that we focus on this paper are tackles and interceptions, defined in Section 2.2. These actions can be valued in different ways depending on the game context.

To value these actions, we aim to predict what was the 'threat' of the passage of play that the defensive action has stopped. To calculate this, we use a metric called 'Expected Threat (xT)' from [26]. This model assigns a value to each attacking event based on the probability they have added to the attacking team's likelihood to score a goal from that passage of play (defined in more details in Section 4.1).

We define the function to calculate xT as $f(e)$. We use this model to calculate the xT value of each event e in a sequence S, which then allows us to train a machine learning model that is able to predict the xT of e that should have happened after a sequence S. This can be defined as $f(e_n) = \Theta(S_a)$ where Θ is the trained machine learning model to predict xT, e_n is the event we aim to predict, S_a is the passage of play made of a events before a DA, and $f(e_n)$ is the xT of the event.

Using the model to predict the 'threat' of an event that did not happen means that we can then value defensive actions by predicting the xT of the event that the defensive action stopped from happening. In the following sections, we give more details regarding the techniques and experiments conducted to create the model described in this paper.

4 THE DAXT MODEL

In this section we present how we assign values to the defensive actions discussed in the previous section. We discuss how we applied deep learning to predict the xT of an event that a DA has stopped, which led to the valuation of all interceptions and tackles in games. The first challenge for us to be able to predict what was stopped from happening is assigning the threat of the actions that have happened. This is achieved through the Expected Threat (xT) model discussed in the next subsection.

4.1 Expected Threat (xT)

The model defined in [26] aims to measure the threat of any single game state at any moment in a game. In order to do so, the model calculates the probability that the player with the ball will shoot and score, with the probability he will move the ball to another location. This is shown by the following equation:

$$\mathbf{xT}_{x,y} = (s_{x,y} \times g_{x,y}) + (m_{x,y} \times \sum_{z=1}^{16} \sum_{w=1}^{12} T_{(x,y) \to (z,w)} \mathbf{xT}_{z,w}) \quad (2)$$

The left side of the equation ($s_{x,y} \times g_{x,y}$) represents the probability of shooting from position x,y (how often players opt to shoot from this position), and the probability that this shot is scored (xG [22] of the shot). For the right side (after the + sign), the author is calculating the probability the player will move the ball to another zone, and the value of this movement. It is worth noting that since players can only shoot or move, then $s_{x,y} + m_{x,y} = 100\%$.

As seen in Equation 2, in order to calculate the xT at zone (x,y), we need to know beforehand the xT at all other zones. To solve

this problem, the author suggests starting with $xT_{x,y} = 0$ at first for all zones, and then evaluating this formula iteratively until convergence. For each evaluation, we use the xT value of the previous iteration. When implementing this model, we ran xT for 43 iterations until full convergence.

To illustrate an example, Figure 2 shows a passage of play by Manchester United in a game with the xT of each action shown. The first pass (1 to 2) and dribble (2 to 3) have negative xT value since they are moving away from goal, decreasing the probability of the attacking team to score. However, the successful pass of Alexis Sanchez into the penalty box resulted in a large positive xT value, indicating the increased threat of scoring a goal.
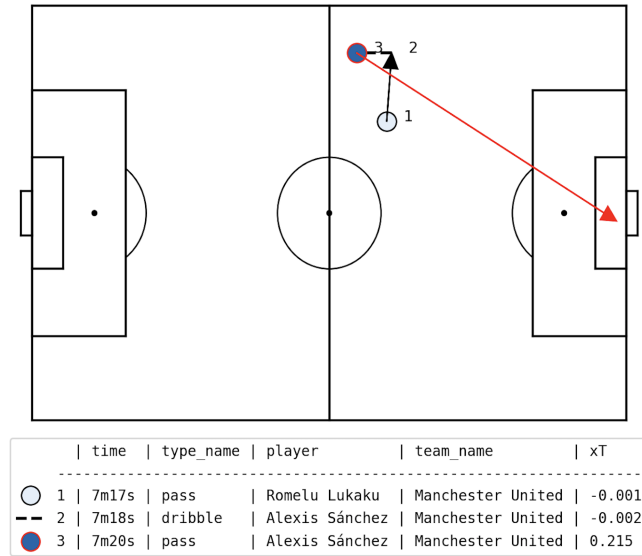

	time	type_name	player	team_name	xT
◯	1	7m17s	pass	Romelu Lukaku	Manchester United	-0.001
--	2	7m18s	dribble	Alexis Sánchez	Manchester United	-0.002
●	3	7m20s	pass	Alexis Sánchez	Manchester United	0.215

Figure 2: Passage of play showing xT of each action

In the next subsection we look at how we can use xT to train a model that can predict what would happen after a sequence of events.

4.2 Predicting What Was Stopped

Due to its generalization mechanism [13], and application to large sets of data (such as game events), we built a simple neural network model to predict the future xT of the events. Specifically, we have built a Multi Layer Perceptron [30] model using the Keras[4] library. As explained in Section 3, Θ is fed as input a fixed number of actions S_a in a passage of play (Section 5.1 shows the experiment done to determine the number of actions a), and as output the xT of the action after this passage of play. Table 1 shows an example of one input instance S_a with $a = 3$. Each event e in S_a is represented by its xT ($f(e)$) and x,y coordinates. The output in this case would be "xT4", signaling the xT of the action after this passage of play.

Due to the distribution of the target values being mainly Laplacian with few outliers (explained further in Section 5.3), the network was trained using a Mean Absolute Error (MAE) loss function, shown in Equation 3.

---
[4]https://keras.io

Table 1: Input example for our Neural Network model

xT1	x1	y1	xT2	x2	y2	xT3	x3	y3

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |\hat{y}_t - y_t| \tag{3}$$

where n is the number of data points used, $\hat{y}_t$ is the predicted xT for a specific instance, and y_t is the actual xT for this instance.

We generate and use three main datasets. The first was formed by collecting all sequences S_a (of a consecutive actions) where each sequence precedes a successful event e such that $e \neq DA$. These instances (example in Table 1) are used as input for Θ, with the xT of the following event as output. All training and testing (random 80-20 split) of Θ will be done on this dataset. The other two were made up of valid passages of play S_a that came before failed events that were interrupted by a DA (one dataset for interceptions and one for tackles). After validating Θ, it will be fed as input these datasets to predict $f(e_n)$ (equation in Section 3) i.e to predict the xT of event e_n that would have happened without the DA. This xT output would thus be the valuation of each interception and tackle.

In the next subsection we discuss how we assign value to each event as well as to the ability of an individual player in a given game for interceptions, tackles, and overall score.

4.3 Assigning Value to Events

As previously mentioned, we apply our DAxT model on the 2 defensive action datasets to predict the xT of the event that was interrupted by a DA. After having each interception and tackle assigned its corresponding Interception value I_V and Tackle value T_V, we grouped these defensive actions according to the player committing them. We then calculated the total I_V and T_V for each player by summing all values together and also computed the average for each per interception and tackle respectively. The results are shown in Section 6.

Moreover, we used these metrics to deduce an overall defender score. In total, four features were used to obtain the score: I_V, T_V, Clearance xT, and Pass xT. We combined the two metrics we got from the DAxT model with the latter two to derive a final score. Clearance xT is the expected threat of a clearance[5] committed by a defender. The advantage of clearances is that they are moving actions (unlike interceptions and tackles), thus have an xT value which correlates to the DAxT calculated. The final feature is Pass xT, which considers the expected threat of each pass committed by a defender. This feature was used to value the importance of defenders from an attack perspective in build up play, as this is becoming increasingly important at the top levels.

We used the cumulative values of each feature for each player, and normalized them to a score between 0-100. When simply calculating the average score of the four features, the rankings were topped by defensive-minded players. This is why we introduced weights to the equation. We applied the weights in such a way where the impact of defensive values will be equal to the impact of offensive values. This can be seen in the following equation:

---
[5]When a player kicks the ball away from their own goal [1]

What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data

KDD '21, August 14–18, 2021, Singapore

$$Sc = ((I_V + T_V + CxT) \div 3 + PxT) \div 4 \qquad (4)$$

where Sc is the final defender score, CxT is the Clearance xT, and PxT is the Pass xT. The division by 3 was done in order to get the mean of the defensive values and the whole equation was divided by 4 to get the average of all values. Section 7.1 illustrates and analyzes the real-world results.

5 EMPIRICAL EVALUATION

To evaluate and optimise our models, we used a dataset collected from two seasons (2017/18 and 2018/19) from the English Premier League (EPL).[6] The dataset breaks down each of the games from the tournament into an event-by-event analysis where each event gives different metrics including: event type (e.g., pass, shot, tackle etc.), the pitch coordinates of the event and the event outcome. This is a rich real-world dataset that allows us to rigorously assess the value of our model. The experiments[7] performed are as follows:

5.1 Experiment 1: Setting the Model Parameters

First, to test the model generalisation, we separated our data into training and validation sets (random split of 80-20). The experiments were then ran on both sets and the MAE (Equation 3) was the metric used.

We ran an experiment to determine the number of actions a that should be taken into consideration in S_a. There could be an argument about using $a = 1$, since having the xT and location of an event could be enough "threat" evidence. Another argument would be that using more actions would be beneficial to our model since it is learning more details about the passages of play. Table 2 shows the training and validation losses, amount of training data, and number of each DA with different number of previous actions.

Table 2: MAE and instances for different number of actions

	$a = 1$	$a = 2$	$a = 3$
Training Loss	0.0174	**0.0154**	0.0123
Validation Loss	0.0169	**0.0153**	0.012
Training Data	950,139	**802,046**	686,529
Interceptions	98,235	**75,691**	59,593
Tackles	22,124	**17,423**	13,676

The number of training data and defensive actions available decreases as a increases. This is logical, since finding 3 consecutive, successful, and moving actions in the dataset is harder than finding 2 or even 1. Thus, this leads to having less available defensive actions to value as the passages of play become longer.

As highlighted in the table, we chose $a = 2$ as the optimal number of actions for passages of play. This result was selected to balance between minimizing the loss of the model and maximizing the number of defensive actions being valued. Although their number was decreasing, not valuing these DA was worth the increase in accuracy. This is due to the fact that by choosing 2 actions instead

of 1, we minimized the randomness of having only $a = 1$ in S_a (which could be considered as outliers that confuse the model), thus allowing better and more accurate sequences to be considered. This decrease in DA eliminated the repeated actions that occur in some games, with actions happening quickly after one another that would not interpret a defender's true ability.

5.2 Experiment 2: Selecting the Features

Another experiment was conducted in order to test different feature sets for our model. According to [2], most existing models that analyze football event data only use location and action type. Since our action type is constant, we tested different combinations of features that include body part, time of game and team ID, other than the already mentioned xT and location.

After testing different combinations, the results showed that xT and x,y coordinates were truly the most important features. The omission of either one of them (or both) resulted in an increase in MAE > 0.002, whereas trying both features with others resulted in differences of ≈ 0.001. The best model according to the lowest MAE value and best learning curve was the model that uses both of these features alone. Adding other variables was either overfitting the model (team ID), which was expected since it disturbs the initial tactical interpretation, or was too general to make a difference (body part, where the large majority of actions was with foot).

5.3 Experiment 3: Model Prediction

The first step towards validating our model is by checking whether its predictive performance deteriorates when applied to unseen data. When exercised on the test set, the model returned an MAE loss of 0.016, which is an accepted result according to the training and validation loss functions in Section 5.1.[8]

The first statistical test we performed was comparing the residuals of the training and testing datasets. The residuals (errors) are the differences between the actual and predicted values of the model. To make sure our model does not over fit to our training data, we used the Levene test [24] and Kolmogorov-Smirnov (KS) [16] test on both residuals. These tests are conducted to compare the variance (Levene) and probability distribution (KS) of our training and testing residuals, which are expected to be similar for our model to be considered a good fit. The Levene test returned a statistic value $v = 1.209$ and a p-value $p = 0.272$. Since the p-value is greater than our 5% threshold, we can then conclude that there is evidence that both residuals have the same variance. Similarly, the KS test's return of $v = 0.0174$ and $p = 0.8333 > 0.05$ suggests that the two probability distributions are the same.

When sketching the fitted line of the probability plot (also known as Q-Q plot [12]) in Figure 3, we could see that the residual's distribution is somewhat normal with long tails on both sides. By visualizing the predictions yielded by our model and comparing them to the actual values, we observed that 96.1% of the data was between 0.05 and -0.05, explaining the tails in the plot. This reasoning was logical since actions with very high xT (or very low) rarely happen in games compared to regular, less significant actions.

---

[6]All data provided by StatsBomb - www.statsbomb.com.
[7]Experiments have been run using Keras and TensorFlow.

[8]For comparative purposes a baseline model predicting 0's would have a MAE of 0.0267.
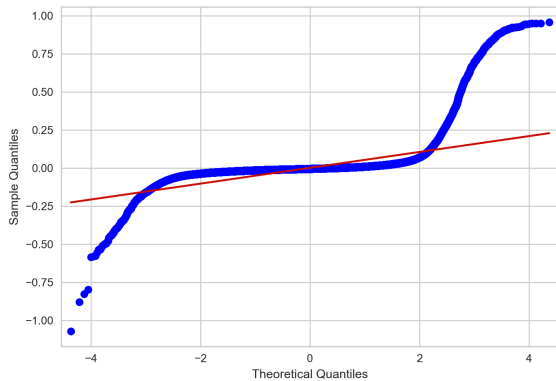
Figure 3: Q-Q plot of normalized residuals

Finally, to show that there is a correlation between the model's predictions and the actual values, we ran a Pearson correlation test which resulted in an r-value $r = 0.0985$ and a p-value less than 0.05. These results are statistically significant and show that, using the data available, we have been able to train a model that can predict the xT of the next event in games of football.

5.4 Experiment 4: Valuing Defenders

After applying our model on all defensive actions, we used Equation 4 to calculate the overall defender score for all players. In order to validate the results, we compared the score of the top 25 center backs, full backs, and defensive midfielders respectively, with their market value[9] at the beginning of 2019. Figure 4 shows the scatter plot that illustrates the correlation between the overall defender score and their market value (in millions of €).
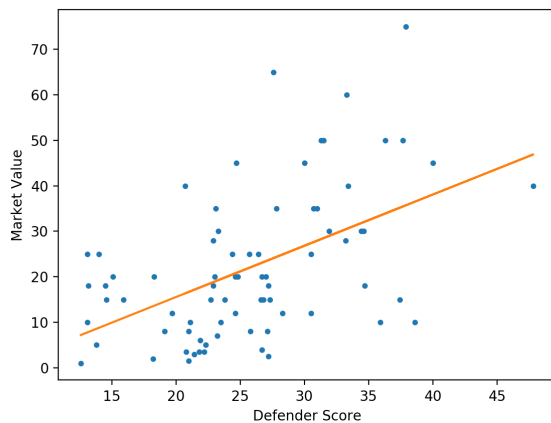

Figure 4: Correlation between Defender Score and Market Value (in millions of €)

---

[9]according to transfermarkt.com

After plotting the regression line, we calculated the Pearson correlation coefficient r and the p-value. This resulted in $r = 0.533$ and $p = 3.37 \times 10^{-6}$. These values show that there is a positive correlation between our defender score and their market value, thus validating the results generated by our model. The reason for not having a stronger correlation coefficient is due to the many factors that add noise to this data away from the actual ability and performance of the player (intangibles that affect market value). For example, other factors include but are not limited to age, nationality, political factors, current club and even popularity [17]. One example in the plot would be Fernandinho who, although considered a world class player and has a high defensive score, was 34 years old in 2019 and arguably past his prime meaning he would have a lower market value.

6 MODEL APPLICATION TO ENGLISH PREMIER LEAGUE

In this section, we show the results of our model when applied on the English Premier League data for the 2017/18 and 2018/19 seasons. As mentioned in Section 4.3, we separated interceptions from tackles and used our DAxT metric to calculate the Interception Value I_V and Tackle Value T_V. Tables 3 and 4 show the results after combining all interceptions and tackles respectively for individual players over the 2 Premier League seasons.

The results shown in both tables are logical since almost all players that appear are defenders that play with teams that finished in the bottom half of the table, meaning they tend to commit more defensive actions as their teams on average have a smaller share of possession in each game. This could be normalized based on the possession of the team to assess how much defending a player does. This is also proved in the number of goals conceded column (GC). Although not directly related, committing more defensive actions increases the overall DAxT of defenders as the sum shown in Tables 3 and 4 are cumulative sums for 2 Premier League seasons.

Table 3: Top 10 players with most cumulative I_V

Player Name	I_V	Count	GC	Team
James Tarkowski	1.605	641	93	Burnley
Ben Mee	1.576	660	95	Burnley
Nathan Aké	1.537	616	129	AFC Bournemouth
Lewis Dunk	1.508	566	109	Brighton & Hove Albion
Shane Duffy	1.473	604	106	Brighton & Hove Albion
Harry Maguire	1.457	593	100	Leicester City
Virgil van Dijk	1.447	657	50	Southampton/Liverpool
Christopher Schindler	1.411	620	124	Huddersfield Town
Jamal Lascelles	1.386	565	67	Newcastle United
Steve Cook	1.357	612	109	AFC Bournemouth

To get a clearer idea about which players carried out the most important *DA*, we normalized the values by *DA* committed i.e. showing the average I_V and T_V for all interceptions and tackles committed respectively by each player. We set thresholds to only include players that committed more than 100 interceptions and more than 50 tackles (this is so we remove the players who had not played a high number of minutes and skew the results). Tables 5 and 6 show the top 10 players with highest average values for each category.

What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data

KDD '21, August 14–18, 2021, Singapore

Table 4: Top 10 players with most cumulative T_V

Player Name	T_V	Count	GC	Team
Wilfred Ndidi	0.435	169	95	Leicester City
Christopher Schindler	0.4	118	124	Huddersfield Town
Idrissa Gana Gueye	0.384	142	86	Everton
James Tarkowski	0.38	85	93	Burnley
César Azpilicueta	0.353	120	77	Chelsea
Nathan Aké	0.335	90	129	AFC Bournemouth
Luka Milivojević	0.318	110	91	Crystal Palace
Aaron Wan-Bissaka	0.315	108	56	Crystal Palace
Ben Mee	0.292	68	129	Burnley
Harry Maguire	0.286	67	100	Leicester City

Table 5: Top 10 players with highest average I_V per interception

Player Name	I_V Average	Count	GC	Team
Fabián Balbuena	0.0031	232	32	West Ham United
James Ward-Prowse	0.00306	125	92	Southampton
Paul Pogba	0.00305	138	70	Manchester United
Cheikhou Kouyaté	0.00294	176	97	West Ham United/Crystal Palace
Geoff Cameron	0.00294	122	30	Stoke City
Isaac Hayden	0.00291	162	65	Newcastle United
Jefferson Lerma	0.0029	140	61	Bournemouth
Solomon March	0.00288	125	106	Brighton & Hove Albion
Patrick van Aanholt	0.00288	289	84	Crystal Palace
Joe Bennett	0.00287	168	50	Cardiff City

Table 6: Top 10 players with highest average T_V per tackle

Player Name	T_V Average	Count	GC	Team
James Tarkowski	0.00448	85	93	Burnley
Shane Duffy	0.00444	54	106	Brighton & Hove Albion
Jamal Lascelles	0.00443	53	67	Newcastle United
Ben Mee	0.00429	68	95	Burnley
Harry Maguire	0.00427	67	100	Leicester City
James Tomkins	0.00426	62	69	Crystal Palace
Issa Diop	0.00416	51	43	West Ham United
Wesley Hoedt	0.00414	69	66	Southampton
Adrian Mariappa	0.00413	53	74	Watford
David Luiz	0.00404	51	49	Chelsea/Arsenal

The inputs were first standardized when training the model, but an inverse transform function was then applied to get back the original corresponding output values. This way, the outputs can be more interpretable and understandable i.e. Fabian Balbuena's I_V average directly correlates to the fact that on average, his interceptions prevented actions of xT = 0.0031 of happening.

7 DISCUSSION

One advantage this model offers is a framework to compare defensive players against each other. To take 2 examples, we used N'Golo Kanté and Aaron Wan-Bissaka as benchmarks [32]. Although both players are world class in many ways, their most well-known important trait (affects their market value) is their ability to intercept (Kanté) and tackle (Wan-Bissaka). We searched for players having similar number of defensive actions as both and compared their defensive and market values.[10]

---

[10]according to transfermarkt.com

Table 7: Comparing players by I_V and market value

Player Name	I_V Average	Interceptions	Market Value
N'Golo Kanté	0.00255	247	€100M
Pierre-Emile Højbjerg	0.00216	241	€12M
Fernando	0.00205	264	€5M

Table 8: Comparing players by T_V and market value

Player Name	T_V Average	Tackles	Market Value
Aaron Wan-Bissaka	0.00291	108	€30M
DeAndre Yedlin	0.00139	89	€8M
Pablo Zabaleta	0.00128	97	€4M

In Tables 7 and 8, our benchmark players are compared with other players having similar number of defensive actions and playing in the same position. Their I_V and T_V average per interception and tackle respectively are observed, as well as their market value at the beginning of 2019. As we can see in both tables, the I_V and T_V are directly proportional to the market value, with our benchmark players having much higher market value than the others. Without the results yielded by our model, we would only be able to compare the raw numbers of interceptions and tackles, without truly understanding why our benchmarks players are top of the class in their defensive ability.

To make sure our model is outputting logical results, we plotted the defensive actions of players in each category and colored each action according to its DAxT. The blue color signals a very high DAxT (top 10 % of all DAxT in the dataset), with green (top 30%), yellow (top 50%) and red (the rest) coming after.
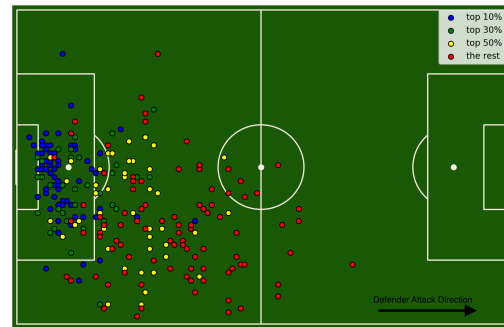

Figure 5: Interceptions committed by Fabian Balbuena

Figure 5 illustrates the interceptions committed by Fabian Balbuena. The ones closest to goal (defensively) are shown to be the most important defensive actions (top 10%). These interceptions illustrate that the preceding passages of play were of high threat, thus increasing the I_V of each interception. The values gradually decrease as actions occur further up the pitch, which is logical since passages of play in the middle of the pitch do not usually lead to
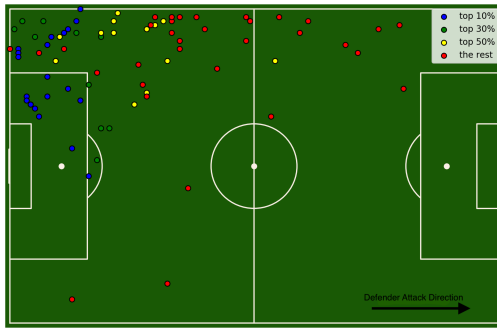
Figure 6: Tackles committed by Andrew Robertson

conceding goals. The same concept applies to Figure 6, that illustrates the tackles committed by Robertson. The ones which directly led to stopping actions of high threat are in blue, and most of the tackles are on the left side of the pitch due to Robertson's position as a left back.

7.1 Overall Defender Score

We applied the overall defender score (introduced in Section 4.3) for ranking and filtered out according to position (center back or full back). We then used CIES Football Observatory and InStat Performance Index [8] for the 2019 season as a means of comparison for our rankings (CIES). Moreover, we calculated the number of goals conceded (GC), appearances (A), goals conceded per appearance (GC/A), and assists (AST) for full backs. Table 9 shows our rankings for center backs while Table 10 shows the rankings for full backs.

Table 9: Ranking Center Backs according to defender score

Player Name	Score	GC	A	GC/A	Rank	CIES
Virgil Van Dijk	37.901	50	64	0.781	1	1
Harry Maguire	36.3	100	69	1.449	2	4
Ben Mee	35.942	95	67	1.418	3	37
James Tarkowski	34.733	93	66	1.409	4	13
Shkodran Mustafi	34.636	70	58	1.207	5	20
Jan Vertonghen	34.426	54	58	0.931	6	20
Toby Alderweireld	33.459	44	48	0.916	7	11
Aymeric Laporte	33.284	26	44	0.591	8	6
Nathan Aké	33.25	129	76	1.697	9	31
Michael Keane	31.913	85	63	1.349	10	37
Lewis Dunk	30.572	109	74	1.472	11	26
David Luiz	30.536	49	46	1.065	12	20
Antonio Rüdiger	29.975	57	60	0.95	13	17
Shane Duffy	28.295	106	72	1.472	14	20
Nicolas Otamendi	27.826	32	52	0.615	15	17

Although there is a season difference between the rankings, we can see many similarities in player standings. Most of the players mentioned are found in both studies, assuring that our results are logical. Moreover, our findings match the top defenders identified by

Table 10: Ranking Full Backs according to defender score

Player Name	Score	GC	A	GC/A	AST	Rank	CIES
César Azpilicueta	47.823	77	75	1.026	11	1	6
Andrew Robertson	40.047	38	58	0.655	16	2	2
Patrick van Aanholt	38.568	84	64	1.3125	3	3	24
Kyle Walker	31.464	39	65	0.6	7	4	12
Trent Alexander-Arnold	31.31	32	48	0.66	13	5	5
Kieran Trippier	31.082	44	51	0.86	8	6	N/A
Ryan Bertrand	27.206	91	59	1.54	4	7	27
Pablo Zabaleta	26.69	97	63	1.54	2	8	N/A
Ben Davies	26.44	46	56	0.82	6	9	24
Marcos Alonso	24.691	58	64	0.906	6	10	1
Ben Chilwell	24.442	73	60	1.216	6	11	8
Aaron Wan-Bissaka	23.331	56	42	1.33	3	12	9
Ricardo Pereira	23.046	45	35	1.285	6	13	4
Luke Shaw	22.994	41	40	1.025	4	14	13
Cédric Soares	22.73	75	50	1.5	5	15	19

experts in the press for that season[11]. This indicates the importance of applying statistical models such as our metrics and xT, rather than using raw data.

Furthermore, the tables illustrate consistency between our defender score and the other variables. For example, we can see that 5 of our top 6 full backs have a low average of goals conceded/appearance with a high number of assists relative to the rest. Also, although Harry Maguire and Nathan Aké, for instance, have a high number of goals conceded/appearance due to the teams they were playing for, their transfer fees (transfermarkt.com) of €87M to Manchester United and €45M to Manchester City respectively show why they are top of our ranking.

As different teams have different playing styles, valuing defenders as such has many advantages. Other than using it directly to compare between players, scouts and teams could use the same metrics but change the weights according to their needs. Instead of equally weighing the defensive and attacking side of players, some teams (such as relegation-threatened ones) could want to focus more on specific defensive metrics, weighing each one accordingly. Moreover, the same defender score could also be used for defensive midfielders. Table 11 shows the top 10 defensive midfielders according to score.

Table 11: Ranking Defensive Midfielders according to defender score

Rank	Player Name	Score
1	Granit Xhaka	37.648
2	Fernandinho	37.437
3	Nemanja Matić	30.737
4	Jorginho	27.652
5	Luka Milivojević	27.299
6	Abdoulaye Doucouré	25.734
7	Idrissa Gueye	24.633
8	Oriol Romeu	24.629
9	Mark Noble	23.197
10	Wilfred Ndidi	23.107

[11]https://www.mirror.co.uk/sport/football/news/top-10-central-defenders-premier-13772574

What Happened Next? Using Deep Learning to Value Defensive Actions in Football Event-Data

KDD '21, August 14–18, 2021, Singapore

7.2 Remaining Challenges

One limitation of our DAxT model is that some defensive actions will be valued less than they should be due to unique passages of plays. For instance, the opposition could counterattack and be in a 3vs1 situation (3 players attacking while only 1 is defending) in the middle of the pitch. If the defender intercepts the ball, our model would not highly reward this action because similar passages of play are not usually of high threat (without taking into consideration the number of defenders behind the ball). This is why future work for valuing defenders' contributions should be centered around the growing collection of tracking data. This will allow analysts to assess the off-ball contributions and movement of defenders with models such as pitch control [27], evaluations of opposition marking and stopping events before they have happened. However, due to the low availability of tracking data at this time, especially across different leagues, it is key to have methods that allow defensive actions to be valued from event-data in smarter ways.

A further challenge our model faces is the dependency on another player metric. This becomes a problem in terms of interpretation, where people need to understand Expected Threat first in order to comprehend our DAxT model. Moreover, metrics such as Possession Adjusted Interceptions and True Tackle Win % are slightly more intuitive than our ratings, which complicates the task for analytically less inclined scouts to fully grasp our model.

8 CONCLUSION

Even though the results have illustrated that our model works well, comparing it to other baseline models remains difficult due to the lack of similar work. One analogous paper is [29], where the authors make use of tracking data to understand the defensive impact of players.

In conclusion, we have presented a novel model for valuing defensive actions by using deep learning to predict a future event that a defensive action has stopped. We have introduced a new metric DAxT by focusing on the values of tackles and interceptions made by defenders. This can help clubs to better understand the contribution of defenders and identify new talent for recruitment in the over 100 leagues that event-data is collected for.

REFERENCES

[1] Alvin Almazov. [n.d.]. *Clearances.* http://alvin-almazov.com/soccer-eng/clearance/
[2] Daniel Altman. 2015. Beyond Shots: A New Approach to Quantifying Scoring Opportunities. In *OptaPro Analytics Forum.*
[3] Ryan Beal, Georgios Chalkiadakis, Timothy J Norman, and Sarvapali D Ramchurn. 2020. Optimising Game Tactics for Football. *International Conference on Autonomous Agents and Multiagent Systems* (2020).
[4] Ryan Beal, Timothy J Norman, and Sarvapali D Ramchurn. 2019. Artificial intelligence for team sports: a survey. *The Knowledge Engineering Review* 34 (2019).
[5] Benjamin Cronin. [n.d.]. *An analysis of different expected goals models.* https://www.pinnacle.com/en/betting-articles/Soccer/expected-goals-model-analysis/MEP2N9VMG5CTW99D
[6] Tom Decroos, Lotte Bransen, Jan Van Haaren, and Jesse Davis. 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.* 1851–1861.
[7] Javier Fernández, Luke Bornn, and Dan Cervone. 2019. Decomposing the Immeasurable Sport: A deep learning expected possession value framework for soccer. In *13 th Annual MIT Sloan Sports Analytics Conference.*
[8] CIES Football and InStat. [n.d.]. *CIES Football Observatory and InStat Performance Index.* https://football-observatory.com/IMG/sites/instatindex/
[9] Jens Grossklags, Nicolas Christin, and John Chuang. 2008. Secure or insure? A game-theoretic analysis of information security games. In *Proceedings of the 17th international conference on World Wide Web.* 209–218.
[10] Christopher Kiekintveld, Manish Jain, Jason Tsai, James Pita, Fernando Ordóñez, and Milind Tambe. 2009. Computing optimal randomized resource allocations for massive security games. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1.* 689–696.
[11] Gunjan Kumar. 2013. Machine learning for soccer analytics. *University of Leuven* (2013).
[12] Usha A Kumar. 2005. Comparison of neural networks and regression analysis: A new insight. *Expert Systems with Applications* 29, 2 (2005), 424–430.
[13] Steve Lawrence, C Lee Giles, and Ah Chung Tsoi. 1998. *What size neural network gives optimal generalization? Convergence properties of backpropagation.* Technical Report.
[14] Wen-Yau Liang, Chun-Che Huang, Tzu-Liang Bill Tseng, Yin-Chen Lin, and Juotzu Tseng. 2012. The evaluation of intelligent agent performance—an example of B2C e-commerce negotiation. *Computer Standards & Interfaces* 34, 5 (2012), 439–446.
[15] Patrick Lucey, Alina Bialkowski, Mathew Monfort, Peter Carr, and Iain Matthews. 2014. quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. In *Proc. 8th annual mit sloan sports analytics conference.* 1–9.
[16] Frank J Massey Jr. 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association* 46, 253 (1951), 68–78.
[17] Oliver Müller, Alexander Simons, and Markus Weinmann. 2017. Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research* 263, 2 (2017), 611–624.
[18] Thomy Phan, Lenz Belzner, Thomas Gabor, and Kyrill Schmid. 2018. Leveraging statistical multi-agent online planning with emergent value function approximation. *arXiv preprint arXiv:1804.06311* (2018).
[19] Dick's pro tips. [n.d.]. *Defensive Soccer Skills: Anticipating Intercepting the Pass with Becky Sauerbrunn.* https://protips.dickssportinggoods.com/sports-and-activities/soccer/defensive-soccer-skills-anticipating-intercepting-pass
[20] Sarvapali D Ramchurn, Trung Dong Huynh, Yuki Ikuno, Jack Flann, Feng Wu, Luc Moreau, Nicholas R Jennings, Joel E Fischer, Wenchao Jiang, Tom Rodden, et al. 2015. HAC-ER: A disaster response system based on human-agent collectives. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS).* 533–541.
[21] Sarvapali D Ramchurn, Feng Wu, Wenchao Jiang, Joel E Fischer, Steve Reece, Stephen Roberts, Tom Rodden, Chris Greenhalgh, and Nicholas R Jennings. 2016. Human–agent collaboration for disaster response. *Autonomous Agents and Multi-Agent Systems* 30, 1 (2016), 82–111.
[22] Alex Rathke. 2017. An examination of expected goals and shot efficiency in soccer. *Journal of Human Sport and Exercise* 12, 2 (2017), 514–529.
[23] Oliver Schulte, Mahmoud Khademi, Sajjad Gholami, Zeyu Zhao, Mehrsan Javan, and Philippe Desaulniers. 2017. A Markov Game model for valuing actions, locations, and team performance in ice hockey. *Data Mining and Knowledge Discovery* 31, 6 (2017), 1735–1757.
[24] Brian B Schultz. 1985. Levene's test for relative variation. *Systematic Zoology* 34, 4 (1985), 449–456.
[25] Eric Shieh, Bo An, Rong Yang, Milind Tambe, Craig Baldwin, Joseph DiRenzo, Ben Maule, and Garrett Meyer. 2012. PROTECT: An application of computational game theory for the security of the ports of the United States. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 26.
[26] Karun Singh. [n.d.]. *Introducing Expected Threat (xT).* https://karun.in/blog/expected-threat.html
[27] William Spearman. 2018. Beyond expected goals. In *Proceedings of the 12th MIT sloan sports analytics conference.* 1–17.
[28] SPIFFYD. [n.d.]. *Soccer Tackling Techniques: Types of Tackles.* https://howtheyplay.com/team-sports/Soccer-tackling-techniques-Types-of-tackles
[29] Michael Stöckl, Thomas Seidl, Daniel Marley, and Paul Power. [n.d.]. Making Offensive Play Predictable-Using a Graph Convolutional Network to Understand Defensive Performance in Soccer. ([n. d.]).
[30] Jiexiong Tang, Chenwei Deng, and Guang-Bin Huang. 2015. Extreme learning machine for multilayer perceptron. *IEEE transactions on neural networks and learning systems* 27, 4 (2015), 809–821.
[31] Lim Weiyang. [n.d.]. *How do you tackle in football?* https://www.myactivesg.com/Sports/Football/Training-Methods/Football-for-Beginners/How-to-tackle-in-football#:~:text=There%20are%202%20types%20of,known%20as%20the%20sliding%20tackle.
[32] Tom Worville. [n.d.]. *Underlying stats stars of 2020: Giroud, Lloris, Buendia, Rice – and Leeds.* https://theathletic.com/2280965/2020/12/31/analytics-review-giroud-lloris-rice-leeds/

A APPENDIX ON REPRODUCIBILITY

In this appendix we give further details on our models and techniques used for reproducibility purposes.

A.1 Data Conversion

For this project, we used StatsBomb's event-based data for the 2017/2018 and 2018/2019 English Premier League seasons (total of 760 games). Due to the inconsistency of events in terms of number of features, we used a data abstraction package called SPADL (Soccer Player Action Description Language). As mentioned in [6], SPADL transforms the event stream data of JSON format into "a common vocabulary that enables subsequent data analysis." This library was created to convert different data providers' event data into a consistent format. Therefore, our model is reproducible with any data. Table 12 shows some features of the first 5 instances in the event-stream data.

We then calculated the xT of every moving successful action in the dataset. Figure 7 illustrates the frequency of each action type.
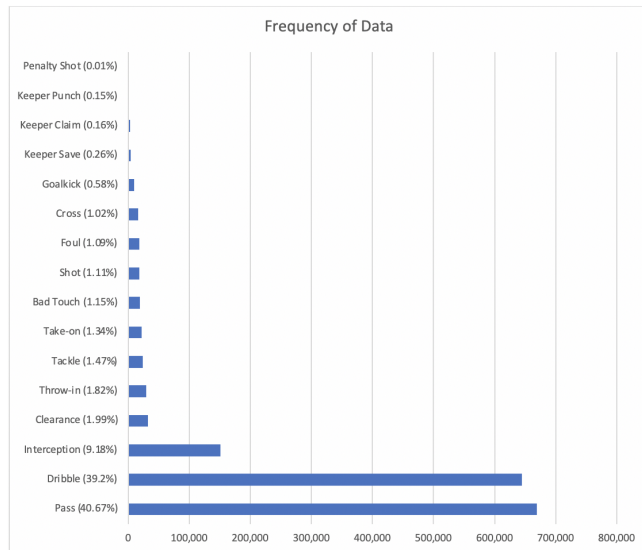

Figure 7: Frequency of each action type in StatBomb's data

After preparing the data, we extracted all 3 consecutive, moving, and successful actions for our first database. Each extraction was 1 instance where 2 actions were considered as input (xT and coordinates) and the third one as output (xT only). This amounted to 802,046 instances. The next step was extracting similar sequences (2 actions) that happened before a failed event due to a defensive action by the opposition. We separated interceptions from tackles which resulted in 75,691 interceptions to value and 17,423 tackles.

Before training, we applied Scikit-Learn's MinMaxScaler [12] on our dataset for normalization.

A.2 Model Setup

For our model, we used the implementation of a Multi Layer Perceptron from Keras' library. The network was trained using a Mean Absolute Error loss function and an Exponential Linear Unit activation function. The deep network was composed of 3 dense hidden layers with 10 neurons each, and ran using Adadelta optimizer for 50 epochs.

For the experiments in the paper, we used the scipy.stats [13] library for the statistical outcomes and the matplotlib [14] library for all graphs.

---

[12]https://scikit-learn.org/
[13]https://docs.scipy.org/doc/scipy/reference/stats.html
[14]https://matplotlib.org/

Table 12: StatBomb's Premier League data converted to SPADL

Game ID	Time in seconds	Team ID	Player ID	Start x	Start y	...	Result ID	Type name	Player name	Country name
14562	1	58	9923	52.0588	34.4304	...	1	pass	Bobby Reid	England
14562	2	58	9917	41.4706	34.4304	...	1	dribble	Joe Ralls	England
14562	2	58	9917	41.4706	34.4304	...	1	pass	Joe Ralls	England
14562	4	58	9924	25.5882	19.7975	...	1	dribble	Sean Morrison	England
14562	6	58	9924	26.4706	18.9367	...	0	pass	Sean Morrison	England