

# **Integrating machine learning and decision support in tactical decision-making in rugby union**

Neil Watson<sup>a</sup>, Sharief Hendricks<sup>b,c</sup>, Theodor Stewart<sup>a</sup>, and Ian Durbach<sup>a,d</sup>,

<sup>a</sup>Department of Statistical Sciences, University of Cape Town, Rondebosch, Cape Town, South Africa; <sup>b</sup>Division of Exercise Science and Sports Medicine, Department of Human Biology, Faculty of Health Sciences, University of Cape Town, Rondebosch, Cape Town, South Africa; <sup>c</sup>Institute for Sport, Physical Activity and Health, Leeds Beckett University, Leeds, United Kingdom; <sup>d</sup>Research Fellow, Centre for Research into Ecological and Environmental Modelling, University of St Andrews

Correspondence regarding this article should be addressed to Neil Watson, Department of Statistical Sciences, University of Cape Town. Contact: nm.watson@uct.ac.za

## **ARTICLE HISTORY**

Compiled June 19, 2020

**ABSTRACT**

Rugby union, like many sports, is based around sequences of play, yet this sequential nature is often overlooked, for example in analyses that aggregate performance measures over a fixed time interval. We use recent developments in convolutional and recurrent neural networks to predict the outcomes of sequences of play, based on the ordered sequence of actions they contain and where on the field these actions occur. The outcomes considered are gaining territory, retaining possession, scoring a try, and being awarded or conceding a penalty. We consider several artificial neural network architectures and compare their performance against baseline models. Accounting for sequential data and using field location improved classification accuracy over the baseline for some outcomes. We then investigate how these prediction models can provide tactical decision support to coaches. We demonstrate that tactical insight can be gained by conducting scenario analyses with data visualisations to investigate which strategies yield the highest probability of achieving the desired outcome.

**KEYWORDS**

Decision support, classification, machine learning, neural networks, performance analysis, rugby union

## Introduction

In all sport, teams and players develop and employ ‘tactics’ – plans to guide how they should play during a performance to maximise their chance of success against their opposition. Tactical decision support – being able to provide feedback to coaches or players on their chosen actions, in light of the available options and game circumstances – is an important goal for performance analysis (Nevill, Atkinson, & Hughes, 2008; Wright, 2009). Historically this has been done mainly using univariate key performance indicators (KPIs), but the ability of these measures to accurately predict key outcomes is limited (Watson, Durbach, Hendricks, & Stewart, 2017). A key reason for this is that rugby union involves complex inter-player interactions that exhibit non-linear self-organising features (Dutt-Mazumder, Button, Robins, & Bartlett, 2011), where success is determined by sequences of actions that occur in space as well as time. Input data is thus fundamentally spatio-temporal.

As far as we are aware, this study is the first to consider the sequential nature of rugby union in predicting intra-game outcomes. We build on recent advances in artificial neural networks (ANNs) to model game outcomes as a function of sequences of actions as well as the field locations where these occur. We assess the degree to which doing so improves our ability to discriminate successful from unsuccessful passages of play.

Sequences of actions have been used to predict the outcome of entire sports games (Bosch & Bhulai, 2018; Teich, Lutz, & Kassarnig, 2016), short passages of play (Harmon, Lucey, & Klabjan, 2016; Shah & Romijnders, 2016; Wang & Zemel, 2016), and physical ball or player movements (Harmon et al., 2016; Mehrasa, Zhong, Tung, Bornn, & Mori, 2018; Zhao, Yang, Chevalier, Shah, & Romijnders, 2018). Most of these studies have employed convolutional or recurrent neural network and have focused on basketball or American football. The sequential, spatio-temporal nature of rugby union thus presents an excellent opportunity to apply these architectures in a novel setting to provide tactical decision support.

The objective of all ‘invasion’ sports is to ‘invade’ the opposition’s territory in advancing towards their goal line to score a goal. The majority of performance research

in invasion sports consists of studies that examine the relationship between KPIs and only the overall outcomes of contests, with few relating KPIs to intra-game outcomes (Clarke & Norman, 1998; dos Santos et al., 2017; Lim, Lay, Dawson, Wallman, & Anderson, 2011). The application of ANNs to invasion sports is expanding, in part due to their ability to model non-linear behaviour. ANNs have primarily been applied in predicting the outcome of sporting events (McCabe & Trevathan, 2008), where most of these studies used only multilayer perceptrons (MLP). In rugby union, Reed and O'Donoghue (2005) and O'Donoghue and Williams (2004) used MLP in predicting the outcome of games, while Passos, Araújo, Davids, Gouveia, and Serpa (2006) investigated the use of ANNs in developing models of interpersonal dynamics between players.

Recently, spatio-temporal research in invasion sports has grown (see Gudmundsson and Horton (2017)), with a considerable amount of work using location data in analysing tactical issues in football (Memmert, Lemmink, & Sampaio, 2017). Currently, only a small subset of this corpus has used machine learning techniques in analysing and classifying intra-game events in basketball (Franks, Miller, Bornn, & Goldsberry, 2015; Miller, Bornn, Adams, & Goldsberry, 2014; Yue, Lucey, Carr, Bialkowski, & Matthews, 2014), football (Horton, Gudmundsson, Chawla, & Estephan, 2015) and American football (Strange & Shamir, 2014). Few other studies have recognised the sequential, time-bounded, and spatial aspects of invasion sports (Dutt-Mazumder et al., 2011), a problem that has only recently been acknowledged in the literature (Woods, Sinclair, & Robertson, 2017).

Even if, as we find in the current study, advanced ANN models can improve predictive accuracy, these have limited value for decision support if the link between inputs and outcomes is not easily interpreted. Traditionally, decision-makers have struggled to trust the insights generated by these models (Namatēvs, Aleksejeva, & Połaka, 2016). We address this issue by demonstrating how to provide tactical insight through combining complex models with data visualisation in scenario-type analyses.

This study illustrates how data science and operational research (OR) can be combined to provide decision support. The rise in popularity of data science and the associated focus on generating insights from big datasets creates an opportunity for

OR. Traditionally, the availability of large amounts of data has not been a focal point of OR. In practice OR methods are often used in data-scarce environments where they develop mathematical models that elicit data from decision-makers or are not subject to the amount of data available. The increasing availability of more copious amounts of sports data can be leveraged in at least two ways to improve the decision support provided by an OR intervention – by incorporating some methods that would not typically be considered and by improving the parametrisation of existing OR models. Hence, investigating how data science and OR can be integrated to provide decision support represents a pertinent research avenue for OR.

We make the following contributions in this paper:

- We apply convolutional and recurrent ANNs to model sequences of player actions in rugby union, and compare the accuracy of several architectures of these models with other machine learning methods;
- We quantify the value of including the field locations and sequential nature of actions in classifying intra-game outcomes;
- We demonstrate how these models can provide tactical decision support via scenario analyses using data visualisations to contrast the probabilities of achieving a desired outcome across various plays.

## Methods

*Data.* Opta provided data for all games of five competitions across 2013, 2014 and 2015: the Heineken Cup, European Rugby Championship, Super Rugby, the Six Nations, and the Rugby Championship.

A total of 313 games were available for analysis. Two teams of two Opta analysts collected the data live, after which another two analysts performed a post-match screening involving numerous accuracy checks. Analysts undergo training for 3 to 6 months before coding live games. Opta performs regular data accuracy checks and monitors each analyst’s accuracy throughout the season. Each match consists of approximately 1500 observations that capture each action performed by a player or a group of players. There are 29 unique categories of actions recorded in the data,

e.g. a ‘carry’, a ‘tackle’, or a ‘pass’. These actions are augmented by variables that describe the type of action. There are 475 unique action descriptors, e.g. for a carry it may be any one of a: ‘one out drive’, ‘pick and go’ or ‘support carry’, each of which has more detailed descriptions. For the current study, we considered only the actions and action description variables along with the  $x$  and  $y$  field locations for each action. These  $x$  and  $y$  coordinates are relative to the team in possession, with  $x = 0$  representing the team in possession’s try-line,  $x = 100$  the opposition’s try-line,  $y = 0$  the left-hand touchline and  $y = 68$  the right-hand touchline of the team in possession.

*Data preparation.* All data preparation and analyses were conducted in the R programming language (R Core Team, 2017). In rugby union, every possession is made up of one or more sequences of play called ‘phases’ that consist of the actions performed by players in-between ‘breakdowns’ in play. A breakdown typically occurs when a player is tackled to the ground by opposition player(s), and one or more players from each team contest for the ball on the ground. This situation is referred to as a ‘ruck’. If the team in possession protects the ball at the ruck, the next phase of play begins. We considered both individual phases of play and possessions of multiple phases of play with their outcomes. Each phase or possession of actions, action descriptions and  $(x; y)$ -coordinates were extracted from the data. Table 1 shows one sample of each of these four input variables for a possession that consists of three phases. The outcomes of the possessions or phases were chosen based on questions that we deemed of interest to coaches:

- Was territory gained or lost?
- Was possession maintained or lost?
- Was a try scored?
- Was a penalty awarded to the team in possession?
- Was a penalty conceded by the team in possession?

Each of these outcomes (ten in total across phases and possessions) is binary, with a ‘positive’ or ‘negative’ outcome. Thus the models were used to classify which sequences of actions, action descriptions and their associated  $(x; y)$  field positions would yield a

positive label for each of the five outcomes.

We truncated the earliest parts of the phase input data to 11 actions and the possession input data to 39, ensuring that over 90% of each of the sequences remained intact. This truncation reduced the sparsity of the data, as a feature of sequential models is that shorter sequences must be “padded” with additional zero values to ensure equal length of all input sequences.

Most of the outcomes consisted of imbalanced data, with the prevalence of the majority label ranging between 59.88% and 98.25%. Since some classifiers perform poorly on heavily imbalanced data (Mazurowski et al., 2008), we decided to either up-sample or down-sample the training data to counteract the effect of the class imbalance on model performance. We conducted experiments on the impact of up-sampling versus down-sampling on test set error across the four outcomes that had the most potential for improvement over a null model that only predicts the majority label. Up-sampling resulted in an average decrease in test set error of 3.54% across these outcomes when compared to down-sampling. Hence the training data were up-sampled for all future experimentation.

We decided to retain the field locations for the ‘gain in territory’ outcome for the following reasons: (1) we removed the actions ‘end of possession’ and ‘end of sequence’ that contain information on the final  $x$ -coordinate value, (2) the territory gained from the last action in a sequence is not included, and (3) the variation in  $x$ -coordinates during a sequence is seldom linear or monotonic.

As most possessions ended with the overt actions ‘end of possession’, ‘restart of play’ or ‘end of sequence’, these actions were removed. Similarly, for the ‘try’ and ‘penalty’ outcomes, the actions ‘try scored’, ‘penalty’ and ‘goal kick’ were removed to force the models to learn patterns in the sequences instead of recognising only one patent action.

*Modelling strategy.* We considered the following ANNs capable of modelling sequences:

- A one-dimensional convolution neural network (CNN) (LeCun, Bengio, et al., 1995)
- Three types of recurrent neural networks (RNN) (Rumelhart, Hinton, Williams, et al., 1988), each of which incorporates different mechanisms to ‘remember’ the

entire sequence of data:

- Gated recurrent units (GRU) (Cho, Van Merriënboer, Bahdanau, & Bengio, 2014)
- Long short-term memory units (LSTM) (Hochreiter & Schmidhuber, 1997)
- Bi-directional recurrent units (BI-RNN) (Schuster & Paliwal, 1997)
- A combined CNN-RNN model consisting of both convolutional layers and recurrent layers with gated recurrent units

We compared the classification performance of these networks with the performance of a null model that predicts the most commonly observed label, a random forest (RF) and a MLP as baseline models. The RF was fitted using the *RandomForest* package (Liaw & Wiener, 2002) that specifies default parameters of 500 trees and number of variables available at each split equal to the square root of the length of the input sequence. These parameters were not optimised as the goal was only to evaluate the classification performance of a non-sequential model. The MLP baseline model was tuned similarly to the other sequential models.

For each network, including the MLP, we fitted three varieties: (i) a single-input model using only the actions, (ii) a multi-input model of the actions and their field locations, and (iii) a multi-input model of the actions, action descriptions and their field locations. The RF model can only take a single input and was thus fed only the sequences of actions. We used a cut-off threshold probability of 0.5 for all models, whereby all samples with an output probability greater than 0.5 were classified as ‘1’.

We represented each categorical action as a numerical vector whose entries are parameters to be estimated by the model, a standard practice in machine learning known as “embedding” (Bengio, Ducharme, Vincent, & Jauvin, 2003). The choice of the embedding layer dimension is subjective, although typically it is much smaller than the total number of unique actions in the data. We experimented with embedding dimensions of 3 to 15 for the actions, choosing 5, and 12 to 50 for descriptors, choosing 25.

Models were fitted using a 60/20/20 training, validation and test split of the data. The training data were then up-sampled to ensure an even separation between the classes for each outcome, while the validation and test datasets were left with the



original class balance. Tuning was undertaken for each ANN architecture separately, and early stopping was employed in all experiments to prevent over-fitting, monitoring both the validation loss and validation accuracy. The maximum number of complete passes through the training data (epochs) any model could perform was set at 200. However, with early stopping employed, most models took 15-35 epochs (mean = 27.19, sd = 9.33) to train.

*Hyperparameter experimentation.* We conducted a hyperparameter experiment, varying the dropout rate (0.05 or 0.5), embedding dimension (3, 9 or 15), kernel size (3, 9 or 15), minibatch size (32, 128 or 256) and the number of layers (1, 2 or 3). We evaluated the performance of the models for all hyperparameter combinations, resulting in a total of 162 ‘runs’ for each type of ANN. The only hyperparameter that yielded any noticeable difference was the number of layers, with a 1.087% test set performance improvement between 2 and 3 hidden layers. Thus, only the effect of the number of hidden layers was investigated when fitting the models.

*Model fitting.* The models were fitted on an Amazon Web Server instance with an Intel Xeon Platinum processor with 8 virtual CPUs, 16GB of RAM and a clock speed of 3.5 GHz. Figure 1 illustrates the architecture of the multi-input CNN-RNN model. A single convolutional layer followed by a single recurrent layer was used for each input ‘stream’ – the embedded actions, embedded action descriptors,  $x$  and  $y$ -coordinates. Outputs from these hidden layers were concatenated and passed to two fully-connected layers, separated by a single dropout layer with 30% dropout rate, before the final output layer. All the other multi-input models have similar model structures.

5-fold cross-validation was used with a minibatch size of 32, as smaller minibatch sizes have recently been shown to result in better training stability and generalizability of models (Masters & Lusch, 2018). We used binary cross-entropy as the loss function and the Adam (adaptive moment estimation) optimiser with its default parameters. The models were compared based on their test error and their percentage improvement over a null model that predicts the most commonly observed label for every prediction. The performance metrics calculated for each fold included the test, validation and

training error and loss, the specificity, sensitivity, precision, F1 score (the harmonic mean of sensitivity and precision) and area under the receiver-operator curve (AUC).

The final values in Tables 2 and 3 are the averages of these metrics across the 5 folds, while Table 4 shows summary results across all outcomes.

## Results

The performance of the models across each outcome for phase and possession sequences are displayed in Table 2 and Table 3 respectively. The single-input models that involve some form of recurrency outperform a feed-forward MLP in eight out of the ten outcomes, and the RF in three outcomes (Tables 2 and 3) on classification accuracy. The ability of these networks to ‘remember’ information further back in the sequences improves their performance. The single-input CNN-RNN and LSTM networks outperform the BI-RNN, CNN and GRU networks across all outcomes and the MLP in nine out of ten outcomes. The CNN-RNN was the best-performing single-input model in 6 outcomes and the LSTM in 4 outcomes.

The inclusion of field locations improves performance, with the networks including these inputs outperforming all others in all five of the phase outcomes (Table 2). The best-performing sequential ANN with field locations (*xy* models) outperforms the MLP<sub>xy</sub> models across all outcomes (Tables 2 and 3). The best-performing network with field locations and action descriptors (*xyd* models) outperforms all other models in nine out of ten outcomes, with the CNN-RNN<sub>xyd</sub> model being the best in seven outcomes and the LSTM<sub>xyd</sub> model in two. A similar trend to the single-input models is observed in both the *xy* and *xyd* models, with the CNN-RNN and LSTM variants most often the best-performing models. The amount of improvement over the best-performing single-input model, i.e. one of the RF, MLP or best single-input network, by including the field locations varies considerably across the network types and outcomes:

- MLP<sub>xy</sub> model (min = -11.691%, max = 98.642%, mean = 5.894%)
- *xy* model (min = 0.432%, max = 90.369%, mean = 25.416%)
- *xyd* model (min = 3.019%, max = 98.642%, mean = 30.753%)

The mean relative improvement over the best single-input model is 4.3 times greater for the best-performing  $xy$  model and 5.2 times greater for the best-performing  $xyd$  model than the MLP\_xy model, and more than 5% greater for the best-performing  $xyd$  models than the best-performing  $xy$  models. All models generally perform better on the phase outcomes than the possession outcomes. The mean relative improvement drops from 18.168% to -6.380% from phases to possessions for the MLP\_xy model, and from 43.661% to 7.172% and 48.477% to 12.029% for the  $xy$  and  $xyd$  models respectively.

The mean test set error across all outcomes is lower for the best-performing MLP\_xy,  $xy$  and  $xyd$  models than the RF (Tables 2 and 3). The RF has the second-lowest sensitivity and the lowest specificity and AUC across all outcomes. The best-performing  $xy$  and  $xyd$  models have much higher mean percentage improvement across all outcomes than any of the single-input models or the MLP\_xy model.

The AUC for the best performing single-input ANNs is higher than the single-input MLP across all outcomes and the RF and in nine out of ten outcomes, even when their test error is larger (Tables 2 and 3). This result indicates that these ANNs are more robust to the threshold classification probability. In general, the specificity of all the models was higher than the sensitivity, implying that they were better at correctly classifying negative than positive labels. The sensitivity of the models should be evaluated in light of the often heavily imbalanced class balances. Half of the outcomes have fewer than 10% positive labels, and nine out of ten outcomes have fewer than 30% positive labels.

If compared to random chance predictions, the outcomes that are most difficult to predict are ‘gain in territory’ and ‘maintaining possession’ for longer sequences of play (possessions). However, when compared to a null model that predicts the majority label, the ‘penalty’ outcomes were the hardest to predict. This result is likely due to there being a wide range of patterns of play that lead to penalties, making them particularly challenging to model. Predicting penalties in shorter sequences of play (phases) is more challenging than longer sequences (possessions). Tries are well-predicted by the models – even though they are two of the three most imbalanced datasets – implying that there are distinct patterns of play or actions that lead to

tries. For example, a sequence of play that involves a linebreak in the opposition's 22m area would have a high probability of leading to a try.

### **Integrating models into decision support**

After fitting the models, we investigated how to use the best overall model to identify tactics that are successful in achieving their desired outcome. We used the CNN-RNN<sub>xyd</sub> model to predict the outcome probabilities of hypothetical sequences of actions, action descriptions and their field locations. These predictions were then visualised as 'heat map' scenario plots – a demonstration of how the models could be used in a decision support system to provide tactical insight. The approach detailed here could assess any sequence of play and would ideally be performed in consultation with stakeholders in a performance analysis system.

To demonstrate the type of decision support that the models can provide, we extracted possession sequences from the data representing different hypothetical scenarios. Each scenario starts with a 'setup' event (one of a lineout, a scrum or from broken play) and ends with a 'strategy' (e.g. carry or pass or tackle). For this study, the scenario analysis was conducted only on possessions consisting of multiple phases of play. The scenario of actions, action descriptions,  $x$  and  $y$ -coordinates were input into the model to predict the probability of the desired outcome. Predictions were made for many different scenarios, as detailed in the paragraphs that follow. This results in vectors of probabilities across almost all possible starting field locations, with each vector corresponding to a different scenario of 'setup event', 'strategy' and 'number of phases'. These probabilities are what are plotted as heat maps in Figures 2 and 3.

To investigate the effect of sequence length, we calculated the average number of phases per possession (and double this average), based on the mean number of actions in a possession and a phase, as  $14.71/5.37 = 2.74$  and  $2 \times 2.74 = 5.48$ , and rounded up to the nearest integer. We thus considered possessions consisting of either three or six phases, and evaluated each combination of setup event, strategy and number of phases.

To assess the sensitivity of predicted probabilities to field locations, we simulated

the same sequence of actions across multiple field positions. This simulation allowed us to compare many different scenarios across the entire field of play. For example, comparing the probability of scoring a try when starting from a set-piece versus broken play in the opposition's half, or the probability of conceding a penalty when in possession for only three phases versus six phases in the 22m area. The illustrative examples in this paper consider sequences that advanced only 5m up the field and 15m to either the left or right of the starting  $y$ -coordinate, i.e.  $(x+5; y\pm 15)$ , but this is not restrictive.

The resulting scenario probabilities were overlaid onto a plot of a rugby union field, along with probability contour lines. These plots provide a graphical display of the probability of achieving the desired outcome subject to a scenario, for any starting field position on the field where the top of the plot is the opposition's try-line. The aim is to stimulate discussion around what underlying tactics are most likely to yield the outcome of interest, and thus provide decision support to decision-makers. For example, such plots could be used to answer a question like: 'If we have the option of a scrum or a lineout in our half, which decision will incur less risk of conceding a penalty?'.

Figure 2 displays the probability of conceding a penalty from a turnover in the sixth phase of play, between possessions that start with broken-play versus a lineout versus a scrum where the scrum-half passes from the base of the scrum. There is little difference in the chance of being penalised due to conceding a turnover from broken-play or a lineout. However, conceding a turnover from a scrum from 30 metres onwards yields a much higher chance of being penalised. What is not readily apparent is the reason for this difference, but that discussion is beyond the scope of the present paper. Even from this 'static' visualisation, a coach gleans valuable decision support. If they are concerned about conceding a penalty, and they have the option of choosing between a lineout and a scrum to start their possession, then for any starting position more than 30 metres from their try-line they should opt for a lineout.

Figure 3 shows the probability of scoring a try in the third phase of play when the setup event is a lineout and the strategy is to carry the ball. This sequence is one that is typical in a rugby union game. The ball travels from left to right on the field, while

the defending team attempt to tackle the current ball carrier, resulting in rucks and multiple phases of play. The sequence of actions is randomly overlaid to demonstrate the progression of the team in possession. The different point markers indicate the three phases of play. The decision support provided here could be that if one wishes to score a try within three phases of play from a lineout, then there is a good chance of doing so from around 70 metres onwards.

The plots presented here use all the data across all teams and competitions. These plots are not intended as a panacea; they are hypothetical illustrations of the potential these models have to provide decision support. There is room for further work that would incorporate important factors that we have not considered, e.g. accounting for the relative strengths of teams in the model. This factor could be incorporated by including a ‘game-state’ variable – some function of the current score difference and time left in the game.

There is considerable variation in the plot patterns when any of the outcome, setup event or strategy are changed. Hence, these plots have the potential to provide more robust support should they be made interactive. For example, in Figure 3 a more dynamic visualisation may allow a coach to change any of the actions in the sequence and render a new heat map.

A decision support system developed around these models would likely be used offline in a post-match analysis setting. An online, game-specific, system could be developed, but that would necessitate first training the model offline and then using it to make in-game predictions. A key question here would be whether live data of the required detail would be available. How much of the significant amount of data pre-processing needed to create the input data could be automated would also need to be considered. Discussion around the ethics of such an online system is beyond the scope of the current study. Both an offline and online version of such a system could provide practical decision support.

## Discussion

Our findings highlight the importance of the sequential and spatial dimensions of rugby union when evaluating team performance. In invasion sports, no actions are performed without considering what previous actions have occurred (or anticipating what consequent actions may occur) and their locations. We posit that not incorporating this information is a key reason why few KPIs in rugby union reliably discriminate between successful and unsuccessful teams. To develop KPIs with more discriminatory power, one has to move on from univariate KPIs and simple regression-based methods, while still clearly communicating insights in a manner accessible to stakeholders. These aspects have received little attention in the literature, highlighting the need for the transparent application of methods to model the spatio-temporal nature of the game. The demonstrated predictive ability of ANNs in modelling the spatio-temporal nature of events here emphasises the value of applying such methods to sport. Our study serves as an example of how this can be achieved.

Dutt-Mazumder et al. (2011) recommended that ANNs are more appropriate for sports performance analysis than conventional statistical methods and should be used in a qualitative way to understand the dynamical attributes of football players. The current study has shown that ANNs can be used quantitatively to provide tactical decision support to coaches in rugby union by combining them with data visualisation. We recommend further research that applies ANNs and other appropriate machine learning techniques to better model the inherent complexity of rugby union and other invasion sports.

Across most invasion sports, the focus is almost exclusively placed on the player or team in possession (Franks et al., 2015). This emphasis reflects in the data, where often observations of only the player or team in possession are available. However, the advent of global positioning (GPS) and other camera-based motion tracking systems has translated into the availability of an increasingly large amount of spatial information on both attacking and defending players. Nascent analysis of this type of data has concentrated on American-based sports. While some studies have focused on modelling sports and the interactions between players as dynamic systems, there is a

need for more studies to include the spatial information of the defence when evaluating performance.

In any applied research, the question of how the intended audience can assimilate insight gleaned from experiments is crucial. To date, the vast majority of KPI analyses in rugby union seek to answer this question by offering only a few suggestions as to how stakeholders could apply research insights. It is not clear how likely it is to provide constructive decision support to stakeholders – both the utility of research insights and how this information is assimilated into their decision making. This question is one that needs attention.

This paper serves as a starting point in demonstrating how data science and OR methods can be combined to provide decision support. Here we have developed a tool that can visualise tactical insight in a way that is easy to understand and apply. However, there is room for more research. For this study, this is partly due to some outcomes not being well predicted. Further research is needed to answer questions around why that is and how to better predict those outcomes. There are other related questions that further investigation could answer. For example, the sheer range of possible intra-game outcomes – how might one guide the choice of which to model? Machine learning methods need larger datasets than most OR or traditional statistical methods. What does a small team with limited data do in light of this? The nature of the easily accessible data, particularly the lack of important spatial information of the defence, is relevant here too.

Many kinds of OR models are designed to operate in relatively data-scarce environments. The increasing availability of big datasets means that aggregating this data can now allow for the parametrisation of some OR models that would previously have been achieved via consulting a small number of experts. For example, decision support for sports may have used interviews with coaches, as in expert systems. The two processes are neither the same nor mutually exclusive and work on integrating data into traditional OR models is ongoing, as evidenced by recent calls for studies that blend data science and OR across a variety of topics and methods (“Call for Papers”, 2018; Giesecke, Liberali, Nazerzadeh, Shanthikumar, & Teo, 2018). Research of this type is growing, with healthcare (Galetsi & Katsaliaki, 2019) and financial services



(Andriosopoulos, Doumpos, Pardalos, & Zopounidis, 2019) receiving attention in recent years. From an OR perspective, research that combines softer OR methods and data science when engaging with decision-makers has great potential to yield valuable contributions to the literature.

The goal of OR in sport should be to provide transparent, actionable decision support to stakeholders (Hurley, 2006). There are relatively few instances detailing how data science is combined with OR, with the current study serving as an example of how this could be achieved. This lack is especially evident when considering the question of how to successfully engage with stakeholders to ensure the uptake of decision support.

Many factors may come into play to determine whether a decision support intervention is effective, all of which revolve around generating ‘buy-in’ from stakeholders. These include navigating the firmly held beliefs of coaches and the knowledge gap between the OR practitioner and stakeholders around the scientific method (Hurley, 2006), having a short time-frame to convince stakeholders of the value of the proffered decision support, and the perceived utility and ease of use of the decision support system (Shibl, Lawley, & Debuse, 2013). Having a competent model (or at least a functional prototype) that is transparent, robust and easy to explain to stakeholders (Levasseur, 2015) is critical. In a high-pressured sports performance environment, there is little room for coaches to engage in discussions around what a decision support system may be able to achieve. There is a heightened need for discussions to revolve around the implementation and improvement of a decision support system that can already be demonstrated. Hence, careful planning must go into how to manage interactions with stakeholders (Hurley, 2006) tactfully. The application of problem structuring methods to this crucial area could provide helpful insight.

## **Conclusion**

This study is the first known of its kind in rugby union where machine learning methods have been combined with data visualisation to provide coaches with decision support for their tactical decision-making. We have demonstrated that the sequential nature

of rugby union is an important factor in the assessment of team performance. This paper is one of only a handful to have applied multi-input RNN and CNN to sport and shown the utility of these networks in classifying sequences of intra-game events. We have established that incorporating field locations improve the accuracy of these models for some outcomes. The best ‘sequential’ network outperformed both a RF and a MLP in every outcome.

The model developed in this study was limited by the nature of the data available. For example, the data did not include the positional information of any player other than the one currently in possession. This information, along with other variables like individual player identifiers, current score difference and time left in the game would facilitate richer scenario analyses and be easily input into the model. The outcomes chosen for this study were also limited to actual measurable events, but there may be other, more subjective measures of success. Here, engaging with coaches and other stakeholders in a real-world performance analysis system would provide valuable insight and feedback on the decision support tool. Each of these limitations offers opportunities for future research.

We envision that coaches and other stakeholders would use the tools developed here to provide tactical decision support, as part of their post-match analysis. This decision support would be achieved by investigating similar scenarios to what happened in the game to inform them of which tactics would be optimal to achieve their desired outcome for that scenario. The system could also be used to inform coaches on what tactical plays to focus on in training before an upcoming game. Finally, the system can serve as a tool via which important tactical information is communicated to and discussed with players via the visualisations.

The methodology employed here is an example of how data science and OR can be combined to provide decision support and is transferable to other invasion sports. With the increasing availability of more massive sports datasets, we recommend that more research be conducted across all sports to leverage this data in blending these two approaches. We have raised and discussed some pertinent issues around the nature and availability of data for this purpose in rugby union. We have highlighted the value of combining data science and OR techniques in providing decision support and the need

for more studies of this type. We have also discussed challenges to the effectiveness of OR interventions in terms of the uptake of decision support and identified the need for more soft OR studies to detail examples of successful engagement with stakeholders in this regard.

## **Funding**

The work was supported by the National Research Foundation of South Africa and the Department of Higher Education and Training via the Teaching and Development Grant (IRMA:29113)

## **Notes on contributors**

Opta, part of Perform Content, a division of Perform Group, is acknowledged for providing the data that made this study possible. All copyright rests with Opta; hence the data is not able to be made publicly available. Opta is the world's leading live, detailed sports data provider. See <http://www.optasports.com/contact/> for more details.

## **References**

- Andriosopoulos, D., Doumpos, M., Pardalos, P. M., & Zopounidis, C. (2019). Computational approaches and data analytics in financial services: A literature review. *Journal of the Operational Research Society*, 1–19.
- Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, 3(Feb), 1137–1155.
- Bosch, P., & Bhulai, S. (2018). Predicting the winner of NFL-games using machine and deep learning..
- Call for papers. (2018). *Journal of the Operational Research Society*, 69(3), 485-486. Retrieved from <https://doi.org/10.1080/01605682.2018.1429707>
- Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

- Clarke, S. R., & Norman, J. M. (1998). When to rush a ‘behind’ in Australian rules football: a dynamic programming approach. *Journal of the Operational Research Society*, 49(5), 530–536.
- dos Santos, F. J. L., Sarmiento, H., Mendes, B., Mauricio, N., Furtado, B., Sousa, P. M., & Pinheiro, V. (2017). Complementary analysis of goal in football through notational analysis, sequential analysis and t-patterns detection/analise complementar do gol no futebol atraves de analise notacional, analise sequencial e detecao de t-patterns. *Revista Brasileira de Futsal e Futebol*, 9(34), 238–250.
- Dutt-Mazumder, A., Button, C., Robins, A., & Bartlett, R. (2011). Neural network modelling and dynamical system theory: Are they relevant to study the governing dynamics of association football players? *Sports Medicine (Auckland, N.Z.)*, 41, 1003-17.
- Franks, A., Miller, A., Bornn, L., & Goldsberry, K. (2015). Characterizing the spatial structure of defensive skill in professional basketball. *The Annals of Applied Statistics*, 9(1), 94–121.
- Galetsis, P., & Katsaliaki, K. (2019). A review of the literature on big data analytics in healthcare. *Journal of the Operational Research Society*, 71, 1–19.
- Giesecke, K., Liberali, G., Nazerzadeh, H., Shanthikumar, J. G., & Teo, C. P. (2018). Call for papers—management science—special issue on data-driven prescriptive analytics. *Management Science*, 64(6), 2972-2972. Retrieved from <https://doi.org/10.1287/mnsc.2018.3120>
- Gudmundsson, J., & Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys*, 50.
- Harmon, M., Lucey, P., & Klabjan, D. (2016). Predicting shot making in basketball using convolutional neural networks learnt from adversarial multiagent trajectories. *Stat*, 1050, 15.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
- Horton, M., Gudmundsson, J., Chawla, S., & Estephan, J. (2015). Automated classification of passing in football. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages=319–330.
- Hurley, W. J. (2006). Foreword: special issue on operations research in sport. *Computers and Operations Research*, 33(7), 1871–1873.
- LeCun, Y., Bengio, Y., et al. (1995). Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, 3361(10), 1995.

- Levasseur, R. E. (2015). People skills: Building analytics decision models that managers use—a change management perspective. *Interfaces*, *45*(4), 363–364.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomforest. *R News*, *2*(3), 18–22. Retrieved from <https://CRAN.R-project.org/doc/Rnews/>
- Lim, E., Lay, B., Dawson, B., Wallman, K., & Anderson, S. (2011). Predicting try scoring in Super 14 rugby union – the development of a superior attacking team scoring system. *International Journal of Performance Analysis in Sport*, *11*(3), 464–475.
- Masters, D., & Luschi, C. (2018). Revisiting small batch training for deep neural networks. *arXiv preprint arXiv:1804.07612*.
- Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A., & Tourassi, G. D. (2008). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural networks*, *21*(2-3), 427–436.
- McCabe, A., & Trevathan, J. (2008). Artificial intelligence in sports prediction. In *Fifth International Conference on Information Technology: New Generations* (pp. 1194–1197).
- Mehrasa, N., Zhong, Y., Tung, F., Bornn, L., & Mori, G. (2018). Deep learning of player trajectory representations for team activity analysis. In *11th MIT Sloan Sports Analytics Conference*.
- Memmert, D., Lemmink, K. A., & Sampaio, J. (2017). Current approaches to tactical performance analyses in soccer using position data. *Sports Medicine*, *47*(1), 1–10.
- Miller, A., Bornn, L., Adams, R., & Goldsberry, K. (2014). Factorized point process intensities: A spatial analysis of professional basketball. In *International Conference on Machine Learning* (pp. 235–243).
- Namatēvs, I., Aleksejeva, L., & Poļaka, I. (2016). Neural network modelling for sports performance classification as a complex socio-technical system. *Information Technology and Management Science*, *19*(1), 45–52.
- Nevill, A., Atkinson, G., & Hughes, M. (2008). Twenty-five years of sport performance research in the journal of sports sciences. *Journal of Sports Sciences*, *26*(4), 413–426.
- O’Donoghue, P., & Williams, J. (2004). An evaluation of human and computer-based predictions of the 2003 rugby union world cup. *International Journal of Computer Science in Sport*, *3*(1), 5–22.
- Passos, P., Araújo, D., Davids, K., Gouveia, L., & Serpa, S. (2006). Interpersonal dynamics in sport: The role of artificial neural networks and 3-d analysis. *Behavior Research Methods*, *38*(4), 683–691.

- R Core Team, R. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.
- Reed, D., & O'Donoghue, P. (2005). Development and application of computer-based prediction methods. *International Journal of Performance Analysis in Sport*, 5(3), 12–28.
- Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al. (1988). Learning representations by back-propagating errors. *Cognitive modeling*, 5(3), 1.
- Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673–2681.
- Shah, R., & Romijnders, R. (2016). Applying deep learning to basketball trajectories. *arXiv preprint arXiv:1608.03793*.
- Shibl, R., Lawley, M., & Debuse, J. (2013). Factors influencing decision support system acceptance. *Decision Support Systems*, 54(2), 953–961.
- Strange, R., & Shamir, L. (2014). Prediction of American Football Plays Using Pattern Recognition. *International Journal of Computer Science in Sport (International Association of Computer Science in Sport)*, 13(2), 6.
- Teich, B., Lutz, R., & Kassarnig, V. (2016). NFL Play Prediction. *arXiv preprint arXiv:1601.00574*.
- Wang, K.-C., & Zemel, R. (2016). Classifying NBA offensive plays using neural networks. In *Proceedings of MIT Sloan Sports Analytics Conference* (Vol. 4).
- Watson, N., Durbach, I., Hendricks, S., & Stewart, T. (2017). On the validity of team performance indicators in rugby union. *International Journal of Performance Analysis in Sport*, 17(4), 609–621. Retrieved from <https://doi.org/10.1080/24748668.2017.1376998>
- Woods, C. T., Sinclair, W., & Robertson, S. (2017). Explaining match outcome and ladder position in the national rugby league using team performance indicators. *Journal of Science and Medicine in Sport*, 20(12), 1107–1111.
- Wright, M. (2009). 50 years of or in sport. *Journal of the Operational Research Society*, 60(sup1), S161–S168.
- Yue, Y., Lucey, P., Carr, P., Bialkowski, A., & Matthews, I. (2014). Learning fine-grained spatial models for dynamic sports play prediction. In *2014 IEEE International Conference on Data Mining* (pp. 670–679).
- Zhao, Y., Yang, R., Chevalier, G., Shah, R. C., & Romijnders, R. (2018). Applying deep bidirectional lstm and mixture density network for basketball trajectory prediction. *Optik*, 158, 266–272.

Possession example	Phase 1					Phase 2				Phase 3			
Actions	CL	CR	TK	TK	RK	PS	CR	TK	RK	PS	CR	TK	RK
Action descriptions	LB	OT	LN	LN	-	CP	OT	CT	-	CP	PG	LN	-
x-coordinate	54	54	59	60	61	60	59	61	64	64	63	68	69
y-coordinate	41	41	38	35	33	31	31	31	20	19	17	16	15

**Table 1.** Example of the four input ‘streams’ for a possession that consists of multiple phases, prior to being embedded (Actions & Action descriptions) and normalised (x-coordinates & y-coordinates). Each vector here represents one sample of each variable. This possession started with a player collecting (CL) a loose ball (LB) and carrying it (CR) before two opposition players tackled (TK) him, resulting in a ruck (RK) to end the first phase. The next two phases both involve a pass (PS) from the base of the ruck followed by a carry, a tackle and another ruck.

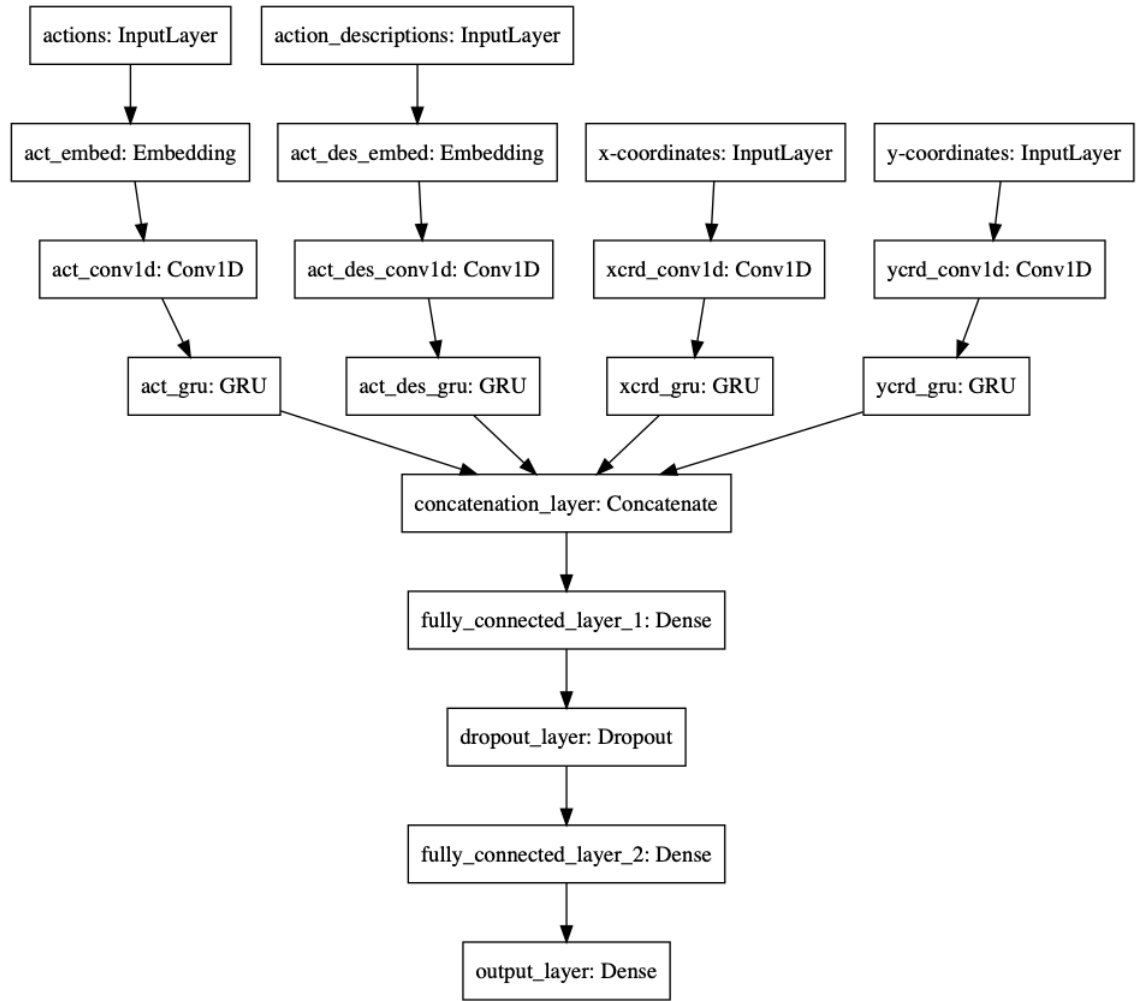


Figure 1. Model architecture of the combined CNN-RNN



Outcome	Model	Test error %	% Improvement	Sensitivity	Specificity	AUC
<i>Phase-territory</i>	RF	29.704	25.962	0.674	0.746	0.710
	MLP	31.064	22.572	0.602	0.820	0.782
Null model error % 40.120	LSTM	29.733	25.890	0.642	0.792	0.790
	MLP <sub>xy</sub>	12.066	69.925	0.847	0.928	0.956
	LSTM <sub>xy</sub>	8.609	78.542	0.904	0.929	0.975
	CNN-RNN <sub>xyd</sub>	8.683	78.357	0.906	0.923	0.972
<i>Phase-possession</i>	RF	6.172	75.984	0.933	0.952	0.943
	MLP	6.430	74.980	0.926	0.963	0.979
Null model error % 25.699	LSTM	6.179	75.956	0.928	0.968	0.982
	MLP <sub>xy</sub>	6.163	76.019	0.928	0.967	0.984
	LSTM <sub>xy</sub>	6.014	76.598	0.931	0.964	0.984
	CNN-RNN <sub>xyd</sub>	5.396	79.003	0.946	0.946	0.984
<i>Phase-try</i>	RF	1.375	26.667	0.899	0.988	0.943
	MLP	1.336	28.747	0.877	0.989	0.971
Null model error % 1.875	CNN-RNN	1.187	36.693	0.906	0.990	0.995
	MLP <sub>xy</sub>	0.721	61.547	0.949	0.994	0.993
	CNN-RNN <sub>xy</sub>	0.582	68.960	0.945	0.995	0.993
	LSTM <sub>xyd</sub>	0.364	80.587	0.931	0.998	0.987
<i>Phase-penalty-awarded</i>	RF	7.177	-47.766	0.731	0.938	0.835
	MLP	7.342	-51.163	0.707	0.938	0.822
Null model error % 4.857	CNN-RNN	7.299	-50.278	0.722	0.938	0.837
	MLP <sub>xy</sub>	6.853	-41.095	0.741	0.941	0.879
	CNN-RNN <sub>xy</sub>	5.601	-15.318	0.658	0.959	0.876
	CNN-RNN <sub>xyd</sub>	4.776	1.668	0.659	0.967	0.887
<i>Phase-penalty-conceded</i>	RF	5.485	-132.711	0.637	0.952	0.795
	MLP	5.796	-145.906	0.619	0.950	0.771
Null model error % 2.357	CNN-RNN	5.844	-147.942	0.620	0.949	0.790
	MLP <sub>xy</sub>	5.124	-117.395	0.595	0.957	0.836
	CNN-RNN <sub>xy</sub>	3.355	-42.342	0.496	0.978	0.838
	CNN-RNN <sub>xyd</sub>	3.160	-34.069	0.511	0.979	0.847

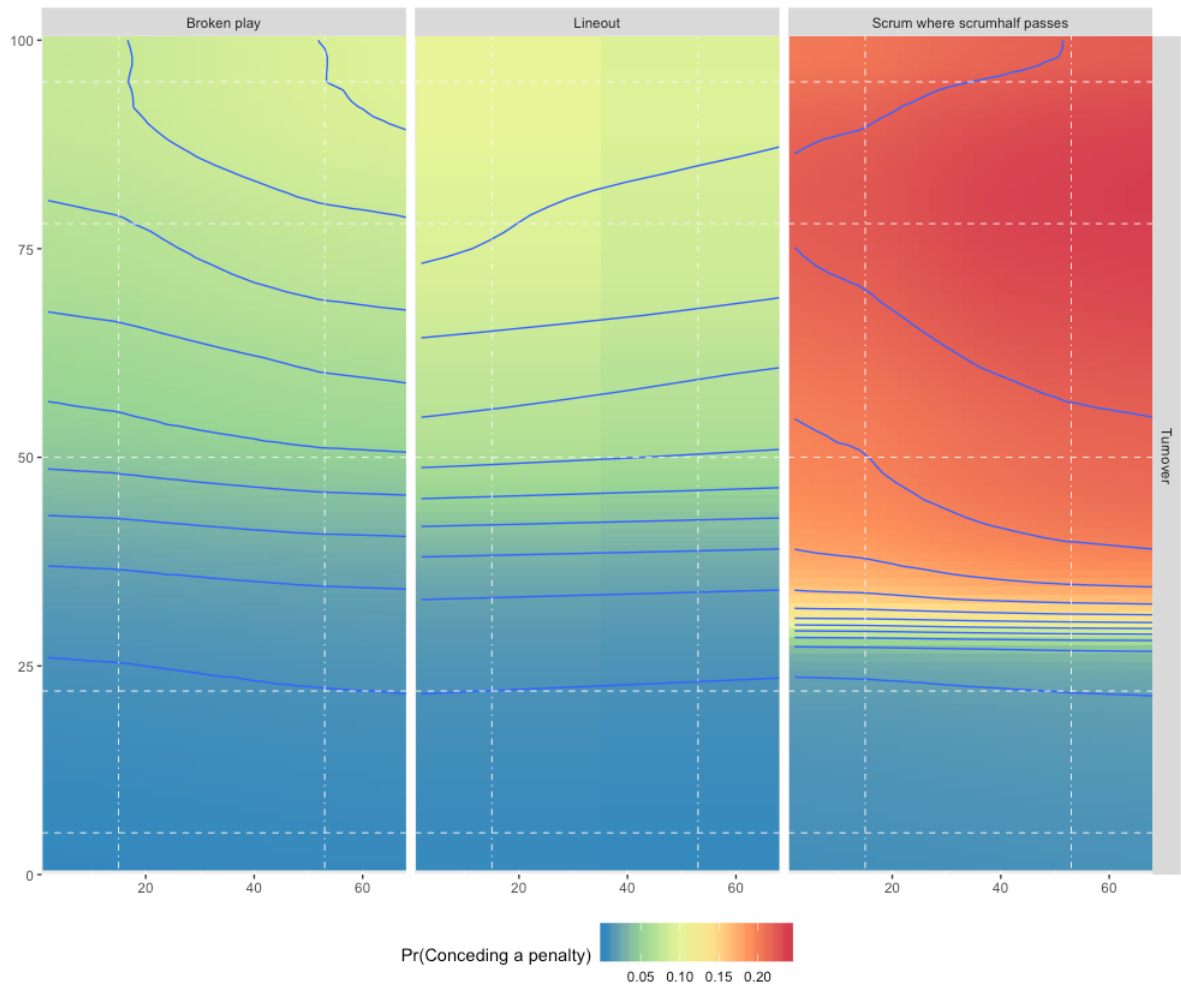
**Table 2.** Model fitting results for phase sequences. The top three models in each outcome are the best-performing single-input RF, MLP and sequential ANN models. Models with the suffixes ‘*xy*’ and ‘*xyd*’ are the best-performing models that include the field locations for actions and both the field locations and action descriptors, respectively.

Outcome	Model	Test error %	% Improvement	Sensitivity	Specificity	AUC
<i>Possession-territory</i> Null model error % 15.417	RF	16.770	-8.776	0.903	0.450	0.676
	MLP	18.543	-20.276	0.872	0.504	0.721
	CNN-RNN	18.887	-22.508	0.863	0.529	0.776
	MLP <sub>xy</sub>	18.105	-17.435	0.832	0.746	0.856
	CNN-RNN <sub>xy</sub>	13.012	15.600	0.896	0.727	0.906
	CNN-RNN <sub>xyd</sub>	12.406	19.530	0.923	0.622	0.890
<i>Possession-possession</i> Null model error % 26.040	RF	18.191	30.142	0.700	0.859	0.780
	MLP	20.461	21.425	0.643	0.848	0.759
	LSTM	17.559	32.569	0.726	0.859	0.852
	MLP <sub>xy</sub>	17.785	31.701	0.707	0.862	0.850
	CNN-RNN <sub>xy</sub>	17.731	31.909	0.729	0.855	0.870
	CNN-RNN <sub>xyd</sub>	16.633	36.125	0.736	0.868	0.872
<i>Possession-try</i> Null model error % 5.294	RF	1.553	70.665	0.946	0.987	0.966
	MLP	1.595	69.872	0.883	0.990	0.989
	CNN-RNN	1.174	77.824	0.942	0.991	0.997
	MLP <sub>xy</sub>	1.383	73.876	0.931	0.989	0.993
	CNN-RNN <sub>xy</sub>	0.735	86.116	0.966	0.994	0.998
	LSTM <sub>xyd</sub>	0.576	89.120	0.969	0.996	0.998
<i>Possession-penalty-awarded</i> Null model error % 15.745	RF	10.409	33.890	0.804	0.913	0.859
	MLP	11.269	28.428	0.698	0.923	0.865
	LSTM	9.955	36.774	0.891	0.902	0.949
	MLP <sub>xy</sub>	11.015	30.041	0.771	0.912	0.896
	CNN-RNN <sub>xy</sub>	10.341	34.322	0.848	0.906	0.935
	CNN-RNN <sub>xyd</sub>	8.742	44.478	0.786	0.936	0.94
<i>Possession-penalty-conceded</i> Null model error % 7.553	RF	8.818	-16.748	0.547	0.941	0.745
	MLP	9.538	-26.281	0.472	0.940	0.755
	CNN-RNN	8.989	-19.012	0.537	0.941	0.817
	MLP <sub>xy</sub>	9.701	-28.439	0.510	0.935	0.788
	CNN-RNN <sub>xy</sub>	8.743	-15.755	0.485	0.947	0.832
	CNN <sub>xyd</sub>	8.117	-7.467	0.462	0.956	0.798

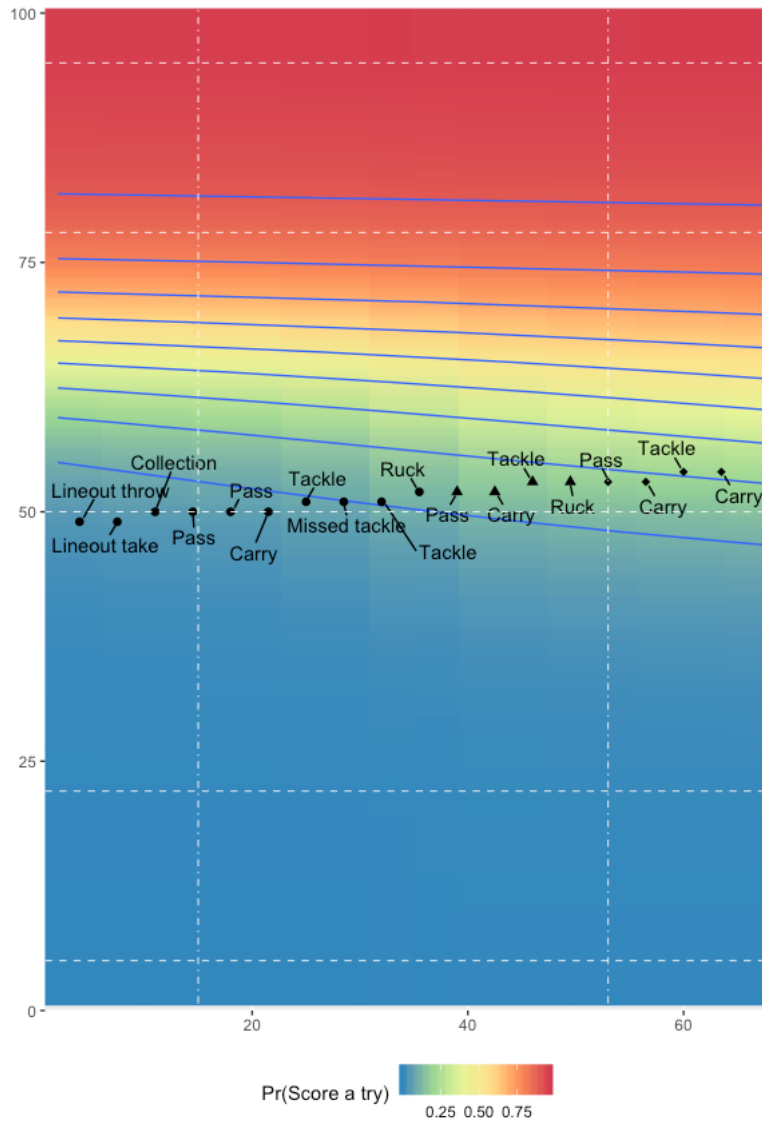
**Table 3.** Model fitting results for possession sequences. The top three models in each outcome are the best-performing single-input RF, MLP and sequential ANN models. Models with the suffixes ‘*xy*’ and ‘*xyd*’ are the best-performing models that include the field locations for actions and both the field locations and action descriptors, respectively.

	Test error %	% Improvement	Sensitivity	Specificity	AUC
RF	10.565	5.731	0.777	0.873	0.825
MLP	11.337	0.240	0.730	0.887	0.841
Best single input network	10.681	4.597	0.778	0.886	0.879
MLP <sub>xy</sub>	8.892	13.875	0.781	0.923	0.903
Best <i>xy</i> network	7.472	31.863	0.786	0.925	0.921
Best <i>xyd</i> network	6.885	38.733	0.783	0.919	0.918

**Table 4.** Model fitting results summary: mean values across all outcomes



**Figure 2.** The probability of conceding a penalty from a turnover is much higher for scrumhalf-pass-based sequences than if play emerges from broken play or a lineout, but only in the top half of the field. Close to a team's own try line there are little differences between the actions.



**Figure 3.** The probability of scoring a try when carrying the ball in the third phase of play from a lineout is uniformly low when starting in your half. It rises steadily when the sequence begins in the opposition's half and remains equally high within the opposition's 22m area. The shape of the point markers indicate the different phases in the possession.