Analysis of team success based on technical match-play performance in the Australian Football League Women's (AFLW) competition

Van der Vegt, Braedan; Gepp, Adrian; Keogh, Justin; Farley, Jessica B.

International Journal of Performance Analysis in Sport

DOI:
10.1080/24748668.2024.2386833

E-pub ahead of print: 04/08/2024

Publisher's PDF, also known as Version of record

International Journal of Performance Analysis in Sport

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rpan20

Analysis of team success based on technical match-play performance in the Australian Football League Women's (AFLW) competition

Braedan van der Vegt, Adrian Gepp, Justin Keogh & Jessica B. Farley

To cite this article: Braedan van der Vegt, Adrian Gepp, Justin Keogh & Jessica B. Farley (04 Aug 2024): Analysis of team success based on technical match-play performance in the Australian Football League Women's (AFLW) competition, International Journal of Performance Analysis in Sport, DOI: 10.1080/24748668.2024.2386833

To link to this article: https://doi.org/10.1080/24748668.2024.2386833

INTERNATIONAL JOURNAL OF PERFORMANCE ANALYSIS IN SPORT
https://doi.org/10.1080/24748668.2024.2386833

Routledge
Taylor & Francis Group

Analysis of team success based on technical match-play performance in the Australian Football League Women's (AFLW) competition

Braedan van der Vegt[a], Adrian Gepp[a,b], Justin Keogh[c,d,e] and Jessica B. Farley[c]

[a]Centre for Data Analytics, Bond Business School, Bond University, Gold Coast, Queensland, Australia; [b]Bangor Business School, Bangor University, Bangor, Wales; [c]Faculty of Heath Sciences & Medicine, Bond University, Gold Coast, Queensland, Australia; [d]Sports Performance Research Centre New Zealand, Auckland University of Technology, Auckland, New Zealand; [e]Kasturba Medical College, Manipal Academy of Higher Education, Mangalore, Karnataka, India

ABSTRACT

An understanding of the effect contextual data may have on key match-play technical performance indicators in the Australian Football League Women's (AFLW) competition is warranted due to its rapid evolution. To address this, predictive models were fit to determine which technical match-play data, including new contextual information, more accurately predict AFLW match outcomes (win/loss, margin), and what are the most important contexts and technical predictors of team performance? Thirteen random forest models were fit, each with greater data contextual interaction including relative to opposition and harder-to-attain match-play variables, field location, and individual player contributions. Models were assessed by prediction performance on match outcome in a holdout sample and variable importance through Mean Decrease in Gini Index. Effective kicks and entries into attacking locations were important in models. Territory gained, contexts of relative performance to the opposition, and locational information around actions improved prediction. This methodology represents the most in-depth analysis of women's Australian football technical match-play performance to date. Commentary presented surrounded issues of using aggregated datasets, prediction with match-play success as a dependent variable, and that detailed, process-oriented approaches are needed in future to avoid large assumptions.

ARTICLE HISTORY
Received 2 October 2023
Accepted 21 July 2024

KEYWORDS
Performance analysis; Australian rules football; machine learning; female sport; AFLW; technical skills

1. Introduction

The women's Australian football (AF) elite competition, the Australian Football League Women's (AFLW), was established in 2017. Performance analysis techniques play a key

CONTACT Braedan van der Vegt bvegt@bond.edu.au Centre for Data Analytics, Bond Business School, Bond University, 14 University Dr, Gold Coast, Queensland 4226, Australia

role in establishing industry practices to support the competition. Specifically, performance analysis involving technical and tactical variables capturing player and team skill actions during match-play and training can inform coaching, player recruitment and development, and gameday tactics (Lord et al., 2020) as it has in prior men's AF (Young, Luo, Gastin, Tran, et al., 2019) and soccer research (Moreira Praça et al., 2023).

Previous studies have analysed technical and tactical performance in women's AF (Black et al., 2019; Cust et al., 2019; Dwyer et al., 2022). Key variables in match-play success included Kicks, Uncontested Possessions, and Disposal Efficiency (Black et al., 2019), Relative Inside 50s between teams, and by key players (Cust et al., 2019), and Inside 50s, Disposals, Marks Inside 50, and Contested Possessions (Dwyer et al., 2022). These datasets included 13 variables (Black et al., 2019) and 23 variables (Dwyer et al., 2022) while Cust et al. (2019) had 12 base variables with feature distributions applied providing a representation of individual player contribution to each variable. Additionally, field location of technical match-play performance actions in the AFLW has been shown to impact key performance indicators characteristic of positional roles (van der Vegt et al., 2023a).

Previous studies (Black et al., 2019; Cust et al., 2019; Dwyer et al., 2022) utilised data from the first three seasons of the AFLW. As it has been demonstrated that the AFLW is a fast-developing competition (van der Vegt et al., 2023b), this may indicate that reassessment with more recent seasons has the potential to produce results more representative of the current competition. Analysis incorporating additional data that has become available since these studies have been published, both in number of variables and contextual information, may build on the results of the current AFLW literature.

Precedent of detailed analysis of wider technical indicators (91 variables after variable reduction) has been conducted in the men's elite competition, the Australian Football League (AFL), across the 2001–2016 seasons, separated into eras for each model fit (Young, Luo, Gastin, Tran, et al., 2019). Key performance indicators produced provided insight into the value of specific data and its effect on prediction accuracy of team match-play success. Contextual information including team technical variables relative to the opposition demonstrated greater prediction accuracy to match outcome than the standard version, with key variables of interest including relative kicks and metres gained, and inside 50-metre entries (Young, Luo, Gastin, Tran, et al., 2019).

These AFL results indicate that the inclusion of additional variables with greater detail and contextual information surrounding each variable (e.g. relative between teams, metres gained on the field through actions by a team, etc.) has the potential to provide greater information of the scenarios in which key indicators influence AFLW match-play success (van der Vegt et al., 2023b). This is further evidenced by the influence of individual player contributions in the AFLW (Cust et al., 2019), meaning the integration of this context may improve prediction accuracy and insights. As a result, investigating how a range of contextual aspects that may be influential (e.g. locational information of actions (van der Vegt et al., 2023a) or metres gained (Young, Luo, Gastin, Tran, et al., 2019)) may have the potential to show the value of current performance data captured.

Pertinent issues affecting the ability for performance analysis to be conducted and interpreted, have been identified in previous literature regarding both women's sport (Emmonds et al., 2019) and specifically women's AF (van der Vegt et al., 2023b). The

comparatively scarce resources in women's competitions relative to their men's equivalent, in terms of facilities, funding, less staff taking on more duties, and comparative contact hours for players under a semi-professional environment, may constrain conducting performance analysis. As such, understanding the value of data currently captured in the AFLW competition, better utilising available data, inclusive of a wider range of technical match-play performance variables and their contextual information, is warranted (van der Vegt et al., 2023b). Potential findings regarding prediction accuracy on match outcomes utilising various levels of detailed data may assist sport practitioners and club personnel regarding decision-making concerning priority of data usage given available resources, training practices, and strategy implementation.

Therefore, a gap exists regarding usage of existing wider range of variables and contexts to inform technical performance, which can in turn also produce a current representative analysis of match-play actions that contribute most to match outcomes within women's AF, allowing for comparison to and potential reinforcement of previous literature. Utilising more seasons of data with additional detail also enables the use of more intensive predictive machine learning models that have the potential to capture further relationships between technical and tactical variables, which may inform key performance indicators that can be used to assist coaching and match analysis (van der Vegt et al., 2023b).

Considering the motivations of building on the current understanding of key AFLW match-play technical variables with the additional contexts available and potential shifts in more recent seasons in a fast-developing league, while accounting for the needs of women's AF of quick, interpretable, and reproducible results, the following study design has been created. This study aims to utilise existing technical match-play datasets while introducing new contextual information to understand how well they predict AFLW match outcomes (win/loss and margin), and to identify which variables and contexts are most important to prediction with consideration of reproducibility and interpretability.

2. Methods

2.1. Data

Data used in this analysis were derived from Champion Data across all AFLW matches from 2017 to the first of two seasons of 2022 (referred to as 2022.1). The data has been validated within men's AF research (Robertson et al., 2016), with similar collection processes in the AFLW, although no validation has been conducted. Quarterly data was chosen to be consistent with past analyses (Black et al., 2019; Cust et al., 2019), for easier comparison and greater sample size at the expense of increasingly sparse data. The study was approved by the Bond University Human Research Ethics Committee (BV00011).

Table 1 explains the detail of each data type used and whether a precedent exists for its use in current women's AF literature. Table 2 presents the different levels of data used in the analysis. Each model utilises incrementally more detailed data or interactions with different detail levels. Models are produced in model sets (e.g. Model 1.1, 1.2 is a set) with different iterations representing different detail combinations.

Table 1. Description of data detail available and use in previous women's Australian football literature.

Data	Rationale	Used in AFLW literature	Examples
Basic Statistics	Basic statistics – Simplest baseline	Yes (2017-18 seasons only) Black et al. (2019); Cust et al. (2019)	**Kicks, Handballs, Marks**
Advanced Statistics – Metres Gained Statistics	Additional variables representing information of team performance in terms of the total distance gained on the field by a team in a match.	No.	**Metres Gained, Metres Gained Retained, Chain Metres**
Relative between Teams	Relative performance in a variable to the other team that is often key to understanding match-play success.	Yes (2017-18 seasons only) Cust et al. (2019).	**Kicks_Relative** (Team Total Kicks – Opposition Total Kicks)
Locational Component	Includes where statistics occur which can give information of why these statistics are important. Five zones on the field as determined and assessed by Champion Data (Attacking Midfield, Defensive Midfield, True Midfield, Forward 50 & Defensive 50). Forward 50 results were also removed in certain dataset iterations as it was anticipated that having the location where almost all scoring actions occur in the data may bias prediction.	No.	**Kicks_AM** – Attacking Midfield Kicks. **Kicks_F50** – Forward 50 Kicks **Kicks_D50** – Defensive 50 Kicks. **Kicks_DM** – Defensive Midfield Kicks.
Feature Distributions of Player Contributions	Represent the contributions of player thresholds (e.g. the sum of how many goals the top 5/20/50% of players [highest goal player, top 4, top 11 players] scored in a game/quarter) and use that as a predictor. Means comment on player contributions to team can be made as in Cust et al. (2019).	Yes (2017-18 seasons only). Only done on 12 variable datasets Cust et al. (2019).	**Kicks_5per** – Kicks attributable to the top 5% of players (approx. 1 player) on the team by kicks. **Kicks_25per** - Kicks attributable to the top 25% of players (approx. 5 players) on the team by kicks. **Kicks_50per** - Kicks attributable to the top 50% of players (approx. 10 players) on the team by kicks.

Table 2. Combinations of data used to fit models to predict Australian Football League Women's (AFLW) match outcomes.

Model set + iteration	Dataset combination used	Number of variables used
Model 1.1	Baseline Data	27
Model 1.2	Model 1.1 + Relative between Teams	54
Model 2.1	All Statistics Data	320
Model 2.2	Model 2.1 + Relative between Teams	640
Model 3.1	Zone of Field Data with Baseline variables	130
Model 3.2	Model 3.1 + Relative between Teams	260
Model 3.3	Model 3.2 - Forward 50 Statistics removed	208
Model 4.1	Model 3.1 + Metres Gained variables	390
Model 4.2	Model 4.1 + Relative between Teams	780
Model 4.3	Model 4.2 - Forward 50 Statistics removed	624
Model 5.1	Model 3.1 + Feature Distributions of Players on Team	520
Model 5.2	Model 5.1 + Relative between Teams	1040
Model 5.3	Model 5.2 - Forward 50 Statistics removed	832

Previous literature removed the less predictive variable of any pair with correlation above 0.95 to avoid multicollinearity issues (Young, Luo, Gastin, Tran, et al., 2019). However, we take a different approach for two reasons. First, the random forest process applied is largely immune to multicollinearity issues in addition to it being performed with greater rigour surrounding sampling, number of iterations, and cross-validation usage (Hastie et al., 2009). Secondly, the technique considers complex non-linear interactive relationships, so exclusion based on linear correlation could remove important variables, reducing model performance (Hastie et al., 2009).

A standard 80:20 split stratified by each season (2017–2022.1–6 seasons) for train:test data was used for modelling and final verification of model performance (Chollet, 2017). This stratification ensured an equal representation of all seasons' data were included in the training and test sets. Splitting by year rather than stratifying can lead to lower accuracy on the test set as it becomes a future prediction of match result problem based on past data rather than in line with the aim of this analysis to understand important variables through quality models varying in data detail (Chollet, 2017). This stratification choice is further substantiated as due to previous temporal differences in variables noted in the AFLW (Dwyer et al., 2022), a break in the dataset was tested temporally through bootstrapped decision trees (Aminikhanghahi & Cook, 2017) where no large enough differences were found to justify an era split.

Quarters that ended in a draw were removed from the dataset due to the small sample size for a classification problem and specificity needed for a regression prediction. Variables that are functions of scoring or match result were removed in all datasets, due to the bias introduced by utilising these variables in prediction (Cust et al., 2019).

2.2. Model fit and selection

Each dataset had a random forest model applied for testing following the precedent of Young, Luo, Gastin, Tran, et al. (2019) in men's AF literature utilising R computing software (R Core Team, 2018). A random forest method was selected for several reasons. An initial decision of employing random forests over advanced regression models and decision trees was made due to the ability to tune parameters, particularly through iterating the sampling size of the model, perform feature selection, and capture highly non-linear relationships for a more representative model enabling conditional two or more variable relationships (Chollet, 2017; James et al., 2013). More interpretable and ease of producing results were a key influence to the selection of the random forest rather than more intensive machine learning models like neural networks (Chollet, 2017). Using random forests are less computational and decision intensive relative to designing neural network architecture. Neural network variable interpretation is difficult considering the need for additional methods such as DeepSHAP (Lundberg & Lee, 2017), which requires interpretation of variable importance per each sample under the assumption of variable independence. In addition, given the specificity of relationships that would arise on a dataset such as ours that is not as large as optimal for neural networks, a greater danger of overfitting models would be present when fitting neural networks of multiple layers (Chollet, 2017). Random forests maintain a degree of the interacting variable structure to still allow for more conditional relationships to arise. Random forests ordinal

and nonparametric structure, means that outliers do not skew results and conformity to a specific distribution is not required (Chollet, 2017).

Given the large number of variables relative to the number of data points in most cases, five-fold cross-validation was performed when fitting models, stratified by season within the training set (Hastie et al., 2009). Each model was also fit multiple times sampling different variable amounts, with the results analysed to select the ideal sample size to avoid overfit and ensure dimensionality problems within the datasets were addressed (Chollet, 2017). The best sampling size iteration was utilised to fit the final model used for prediction on all available training data.

2.3. Model evaluation and variable importance

Comparison of modelling performance between each dataset was measured using one of two metrics: Mean Absolute Error (MAE) (Black et al., 2019; Cust et al., 2019) for numeric margin prediction and percentage accuracy for win-loss classification. Both are necessary as scenarios can arise where prediction may be close by MAE but wrong in classification, such as an incorrect prediction of a 1-point win for team A, when team B won by 1-point. The scale of the MAE is important as it can be interpreted as an exact error by the number of points scored, with the context of six points for a goal, and one point for a behind (missed shot that still scores) showing the scale of error relative to the number of scoring actions.

Variable importance was determined via the Mean Decrease in Gini Index (MDGI) for every final model; this is common when applying random forests while Mean Decrease in Accuracy assessment provided similar results (Breiman, 2001). Assessment of variable importance from random forests models has the benefit of being able to handle correlated variables without suffering from multicollinearity issues (Chollet, 2017).

As a result, themes can be extracted on key variable groups with similar traits that are important to models, which can then be utilised as key performance indicators by practitioners and coaches. MDGI does not indicate relationship direction, meaning that correlation directions to the dependent variable (quarter margin) were analysed to determine the direction of the relationship of important variables in each model's dataset. This comes with the caveat that caution should be taken in directional interpretation as the non-linear relationship is likely not simplistic given the interacting nature of variable importance in random forests (Chollet, 2017).

3. Results

A total of 2280 data points (285 matches × 2 teams × 4 quarters per match) were available for use with 52 (26 × 2) drawn quarters removed from the dataset. This resulted in a final dataset of 2228 data points. Figure 1 presents the MAE and accuracy results of each model from Table 2 above. Further detail in table form can be found in supplementary material Table 1.

Tables 3 and 4 display the 20 most important variables by MDGI for each model along with whether they are positively or negatively correlated with the dependent variable (quarter margin).

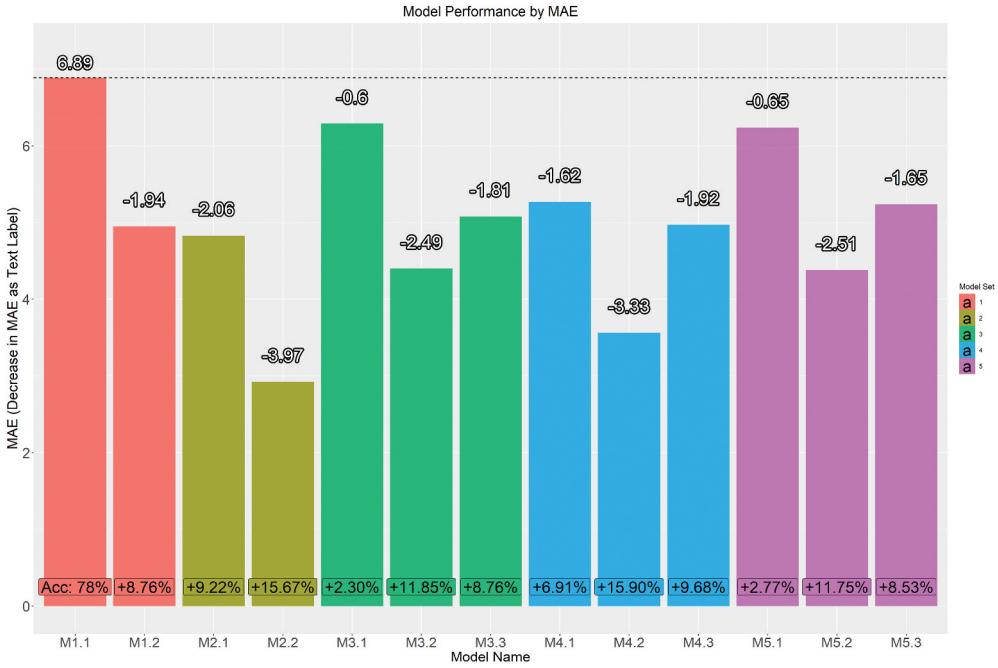

Figure 1. Model performance to determine which technical match-play data best predicts Australian Football League Women's competition match outcomes in terms of difference in mean absolute error and prediction accuracy relative to the baseline Model 1.1.

In Model 1.1 (M1.1), Inside 50 (I50) was the most important variable by MDGI. Effective and Total Kicks, and Clearances were also strongly positively associated with success, while Turnovers (TO) and Rebound 50s (R50s) were negatively associated. Model 1.2 (M1.2) revealed similar important variables except in relative form, producing better prediction than M1.1 by MAE and accuracy.

Model 2.1 (M2.1) indicated the Chain Metres variable was responsible for a large portion of predictive power, while Repeat I50s also was a strong indicator. The relative form of similar statistics were again stronger indicators in Model 2.2 (M2.2), with Chain Metres and Metres Gained (MG) constituting most of the predictive power. M2.1 and M2.2 both improve on M1.1 and M1.2 predictions with the relative version performing particularly well.

Model 3.1 (M3.1) was dominated by efficiency inside the F50 statistics and I50s, indicating the importance of field position with the strongest negative association coming through Free Kicks conceded in defence. Model 3.2 (M3.2) shows many of the same statistics as M3.1 in relative form, although with a large amount of predictive power attributable to Effective Kicks inside F50. When removing F50 variables, I50s and less R50s once again dominated prediction. All of Model set 3 improved on Model set 1 although the removal of F50 location in Model 3.3 (M3.3) reduced accuracy.

The re-addition of MG statistics in Model 4.1 (M4.1), as seen in Table 4, found many of these variables included, with MG Retained and Effective Kicks in the F50 the two most important variables. Model 4.2 (M4.2) shows much the same results in relative form. Removal of F50 statistics shows MG in Attacking Midfield, both in total and when

Table 3. Twenty most important variables for each model fit to predict quarter margins in the Australian Football League Women's competition and accompanying variable importance score.

Highest importance variables for each model

Model 1.1	MDGI	Model 1.2	MDGI	Model 2.1	MDGI	Model 2.2	MDGI
INSIDE_50	49272.47	INSIDE_50_relative	81934.22	CHAIN_METRES_NET	135352.69	CHAIN_METRES_relative	144338.62
EFFECTIVE_KICK	19686.92	EFFECTIVE_KICK_relative	33416.41	CHAIN_METRES_NET_PG	10473.92	TOTAL_GAINED_METRES_relative	48234.55
KICK	16271.30	KICK_relative	22169.26	IN50_RPT_RETURN	4698.35	KICK_GAIN_METRES_relative	8965.94
TURNOVER	**11302.89**	**REBOUND_50_relative**	**17609.66**	IN50_MID_STOP_RETURN	3488.84	CHAIN_METRES_NET_relative	2778.08
CLEARANCE	9111.40	DISPOSAL_EFFICIENCY_relative	4102.15	CLEARANCE_WIN_PCT	3345.25	METRES_GAINED_RETAINED_relative	2241.13
REBOUND_50	**9104.72**	MARK_relative	3984.27	KICK_GAIN_METRES	3212.32	IN50_RPT_RETURN_relative	1553.17
KICK_EFFICIENCY	8968.47	**TURNOVER_relative**	**3764.70**	METRES_GAINED_RETAINED	2952.41	METRES_GAINED_NET_relative	1100.19
EFFECTIVE_DISPOSAL	8643.47	KICK_EFFICIENCY_relative	3319.92	KICK	2190.50	IN50_TRANS_RETURN_relative	1050.90
TACKLE	8071.75	CLEARANCE_relative	2943.41	METRES_GAINED_NET	1541.23	IN50_MID_STOP_RETURN_relative	771.69
DISPOSAL	7748.23	INTERCEPT_relative	2738.92	METRES_GAINED_EFF	1497.84	**REBOUND_50_relative**	**355.28**
DISPOSAL_EFFICIENCY	7337.57	EFFECTIVE_DISPOSAL_relative	2668.35	TOTAL_GAINED_METRES	1280.18	IN50_INIT_RETURN_relative	344.07
KICK_TO_HANDBALL_RATIO	6827.40	TACKLE_relative	2654.14	**SQUAD_PRESSURE_CHANCE**	**1208.62**	IN50_KICK_RETAIN_PCT_relative	282.33
HANDBALL_EFFICIENCY	6479.93	INSIDE_50	2563.05	IN50_INIT_RETURN	1125.00	METRES_GAINED_RETAINED	280.06
FREE_FOR	6162.29	HANDBALL_EFFICIENCY_relative	2387.28	IN50_TRANS_RETURN	1046.00	METRES_GAINED_EFF_relative	273.15
HARD_BALL_GET	5985.83	**REBOUND_50**	**2368.96**	SQUAD_TACKLE_EFFICIENCY	945.76	SHRT_KICK_FROM_POS_EFF_PCT_relative	272.78
BEHIND	5757.23	HANDBALL_EFFICIENCY	2088.20	CHAIN_METRES	906.50	IN50_RPT_RETURN	233.11
MARK	5685.83	KICK_EFFICIENCY	1958.74	CHAIN_LAUNCH	877.91	IN50_RPT_DIR_RETURN_relative	223.67
INTERCEPT	5672.25	DISPOSAL_EFFICIENCY	1950.00	CHAIN_METRES_NET_STOP	868.58	CHAIN_METRES_NET	217.97
FREE_AGAINST	**5095.78**	CLEARANCE	1877.97	PLY_TACKLE_EFFICIENCY	789.82	IN50_LAUNCH_PCT_relative	211.61
HITOUT	4815.72	KICK_TO_HANDBALL_RATIO_relative	1814.06	SQUAD_PRESSURE_FACTOR	779.17	DISPOSAL_EFFICIENCY_relative	209.78

Model 3.1	MDGI	Model 3.2	MDGI	Model 3.3	MDGI
EFFECTIVE_KICK_F50	97738.94	EFFECTIVE_KICK_F50_relative	178054.78	INSIDE_50_AM_relative	96523.12
INSIDE_50_AM	9502.39	KICK_EFFICIENCY_F50_relative	1183.11	**REBOUND_50_D50_relative**	**34659.13**
KICK_EFFICIENCY_F50	7726.63	DISPOSAL_EFFICIENCY_AM_relative	1151.37	INSIDE_50_DM_relative	7854.96
KICK_F50	6680.79	KICK_EFFICIENCY_AM_relative	1035.39	EFFECTIVE_KICK_AM_relative	4902.23
FREE_AGAINST_D50	**4704.60**	KICK_TO_HANDBALL_RATIO_AM_relative	1014.21	DISPOSAL_EFFICIENCY_AM_relative	3800.87
DISPOSAL_F50	2880.59	INSIDE_50_AM_relative	942.60	**FREE_AGAINST_D50_relative**	**2821.65**
EFFECTIVE_DISPOSAL_F50	2862.27	DISPOSAL_EFFICIENCY_F50_relative	872.95	KICK_EFFICIENCY_AM_relative	2558.10
KICK_TO_HANDBALL_RATIO_DM	2724.42	KICK_EFFICIENCY_AM	765.46	EFFECTIVE_DISPOSAL_AM_relative	2517.28
TACKLE_AM	2501.62	DISPOSAL_EFFICIENCY_AM	755.58	**REBOUND_50_D50**	**2451.83**
DISPOSAL_EFFICIENCY_D50	2342.14	CLEARANCE_DM	754.60	INSIDE_50_AM	2259.00
TURNOVER_D50	**2313.29**	KICK_TO_HANDBALL_RATIO_DM	722.07	DISPOSAL_EFFICIENCY_D50_relative	1882.37
KICK_EFFICIENCY_DM	2296.03	KICK_TO_HANDBALL_RATIO_F50_relative	686.52	**INTERCEPT_D50_relative**	**1523.72**
KICK_DM	2269.08	EFFECTIVE_KICK_F50	685.46	CLEARANCE_DM	1494.74

(*Continued*)

Table 3. (Continued).

KICK_TO_HANDBALL_RATIO_D50	**2236.74**	**HITOUT_DM**	**668.36**	DISPOSAL_EFFICIENCY_DM_relative	1421.82
DISPOSAL_EFFICIENCY_AM	2220.45	HANDBALL_EFFICIENCY_AM_relative	622.75	INSIDE_50_DM	1333.55
KICK_TO_HANDBALL_RATIO_AM	2203.54	HANDBALL_EFFICIENCY_DM_relative	612.28	HANDBALL_EFFICIENCY_DM_relative	1208.78
FREE_AGAINST_DM	**2095.21**	KICK_TO_HANDBALL_RATIO_F50	608.31	DISPOSAL_EFFICIENCY_AM	1194.86
DISPOSAL_EFFICIENCY_DM	2005.34	TACKLE_AM_relative	608.13	CLEARANCE_DM_relative	1184.63
KICK_TO_HANDBALL_RATIO_F50	1989.77	**KICK_TO_HANDBALL_RATIO_D50_relative**	**606.38**	**HANDBALL_EFFICIENCY_D50_relative**	**1172.27**
HANDBALL_EFFICIENCY_DM	1973.35	KICK_TO_HANDBALL_RATIO_DM_relative	595.96	KICK_EFFICIENCY_DM_relative	1123.79

MDGI = Mean Decrease in Gini Index. AM = Attacking Midfield, DM = Defensive Midfield, F50 = Forward 50, D50 = Defensive 50, PG = Possession Gain,PCT = Percent, RPT = Repeat, STOP = Stoppage, EFF = Effective, TRANS = Transition, PLY = Player, DIR = Direct. Further definitions available in supplementary material. Variables in bold are those with a negative correlation to the dependent variables (quarter margin).

Table 4. Twenty most important variables for each model fit to predict quarter margins in the Australian Football League Women's competition and accompanying variable importance score.

Model 4.1	MDGI	Model 4.2	MDGI	Model 4.3	MDGI
EFFECTIVE_KICK_F50	62270.47	EFFECTIVE_KICK_F50_relative	80392.66	CHAIN_METRES_NET_AM_relative	45049.71
METRES_GAINED_RETAINED_F50	38798.35	METRES_GAINED_RETAINED_F50_relative	45228.34	METRES_GAINED_RETAINED_AM_relative	21404.54
CHAIN_METRES_NET_D50	6282.06	METRES_GAINED_EFF_F50_relative	27420.53	TOTAL_GAINED_METRES_AM_relative	19010.99
CHAIN_METRES_NET_DM	6124.34	KICK_GAIN_METRES_F50_relative	8837.95	CHAIN_METRES_NET_D50_relative	16737.91
TOTAL_GAINED_METRES_AM	6050.35	EFFECTIVE_DISPOSAL_F50_relative	6555.74	CHAIN_METRES_NET_DM_relative	11128.90
KICK_EFFICIENCY_F50	4885.64	TOTAL_GAINED_METRES_F50_relative	5541.99	KICK_GAIN_METRES_AM_relative	8006.19
KICK_GAIN_METRES_F50	4000.54	KICK_EFFICIENCY_F50_relative	3442.13	METRES_GAINED_NET_DM_relative	5916.67
METRES_GAINED_EFF_F50	3935.75	METRES_GAINED_NET_F50_relative	3357.98	METRES_GAINED_EFF_AM_relative	5422.25
KICK_GAIN_METRES_AM	3778.67	METRES_GAINED_RETAINED_F50	2383.32	TOTAL_GAINED_METRES_DM_relative	4274.89
CHAIN_METRES_NET_AM	3777.39	METRES_GAINED_RETAINED_AM_relative	2061.43	CHAIN_METRES_NET_PG_D50_relative	3053.62
TOTAL_GAINED_METRES_F50	3552.69	EFFECTIVE_KICK_F50	2032.86	INSIDE_50_AM_relative	2929.15
CHAIN_METRES_NET_PG_DM	3414.19	KICK_F50_relative	1349.01	CHAIN_METRES_NET_D50	2419.94
CHAIN_METRES_AM	3199.57	METRES_GAINED_EFF_F50	1330.08	CHAIN_METRES_NET_DM	2325.74
CHAIN_METRES_NET_PG_D50	3145.54	CHAIN_METRES_NET_D50_relative	1294.34	KICK_GAIN_METRES_DM_relative	2096.88
METRES_GAINED_NET_AM	3053.11	CHAIN_METRES_NET_AM_relative	1212.14	CHAIN_METRES_NET_M_relative	2091.75
KICK_F50	3040.17	METRES_GAINED_EFF_AM_relative	873.87	METRES_GAINED_NET_AM_relative	2025.58
METRES_GAINED_RETAINED_AM	1989.78	DISPOSAL_EFFICIENCY_F50_relative	868.35	CHAIN_METRES_NET_CB_M_relative	1830.08
EFFECTIVE_DISPOSAL_F50	1708.01	TOTAL_GAINED_METRES_DM_relative	630.18	CHAIN_METRES_NET_ST_M_relative	1664.43
CHAIN_METRES_NET_PG_AM	1627.93	CHAIN_METRES_NET_DM_relative	629.92	CHAIN_METRES_NET_PG_DM_relative	1407.21
METRES_GAINED_EFF_AM	1316.09	CHAIN_METRES_NET_PG_AM_relative	561.19	METRES_GAINED_RETAINED_AM	1194.83
Model 5.1	**MDGI**	**Model 5.2**	**MDGI**	**Model 5.3**	**MDGI**
EFFECTIVE_KICK_F50_50per	22347.56	EFFECTIVE_KICK_F50_25per_relative	48233.24	INSIDE_50_AM_relative	51769.97
EFFECTIVE_KICK_F50_25per	21987.79	EFFECTIVE_KICK_F50_relative	38396.56	INSIDE_50_AM_50per_relative	41228.64
EFFECTIVE_KICK_F50	20587.25	EFFECTIVE_KICK_F50_50per_relative	37280.68	**REBOUND_50_D50_relative**	**19365.48**
KICK_EFFICIENCY_F50_50per	8200.96	KICK_EFFICIENCY_F50_25per_relative	14945.41	**REBOUND_50_D50_50per_relative**	**11590.94**
KICK_EFFICIENCY_F50_25per	7732.34	KICK_EFFICIENCY_F50_50per_relative	14361.53	DISPOSAL_EFFICIENCY_AM_50per_relative	3392.33
KICK_TO_HANDBALL_RATIO_F50_50per	5326.61	EFFECTIVE_DISPOSAL_F50_25per_relative	5064.39	**FREE_AGAINST_D50_relative**	**2207.45**
KICK_F50_50per	4609.21	EFFECTIVE_DISPOSAL_F50_50per_relative	4138.14	DISPOSAL_EFFICIENCY_AM_relative	2198.40
KICK_F50	4208.14	EFFECTIVE_DISPOSAL_F50_relative	3543.89	INSIDE_50_AM_25per_relative	1930.43
KICK_TO_HANDBALL_RATIO_F50_25per	3754.12	EFFECTIVE_KICK_F50_5per_relative	1896.44	**REBOUND_50_D50_25per_relative**	**1765.27**
EFFECTIVE_DISPOSAL_F50	3287.54	DISPOSAL_EFFICIENCY_F50_50per_relative	1324.40	KICK_EFFICIENCY_AM_relative	1685.96
INSIDE_50_AM	3050.04	KICK_TO_HANDBALL_RATIO_F50_50per_relative	1318.24	INSIDE_50_DM_5per_relative	1640.59
KICK_EFFICIENCY_F50	2793.19	DISPOSAL_EFFICIENCY_F50_25per_relative	1218.09	INSIDE_50_DM_25per_relative	1611.20
INSIDE_50_AM_50per	2767.50	KICK_F50_25per_relative	1085.90	INSIDE_50_DM_50per_relative	1577.59
KICK_F50_25per	2452.15	KICK_F50_50per_relative	1063.39	KICK_EFFICIENCY_AM_50per_relative	1567.61

(*Continued*)

Table 4. (Continued).

EFFECTIVE_DISPOSAL_F50_25per	2296.20	EFFECTIVE_KICK_F50_50per	1046.36	INSIDE_50_DM_relative	1463.54
EFFECTIVE_DISPOSAL_F50_50per	2037.69	KICK_EFFICIENCY_F50_relative	1043.27	**REBOUND_50_D50**	**1341.09**
EFFECTIVE_KICK_F50_5per	2017.55	KICK_TO_HANDBALL_RATIO_F50_25per_relative	988.90	DISPOSAL_EFFICIENCY_D50_relative	1289.05
FREE_AGAINST_D50	**1994.44**	EFFECTIVE_KICK_F50_25per	902.67	EFFECTIVE_KICK_AM_50per_relative	1288.12
KICK_EFFICIENCY_F50_5per	1746.38	EFFECTIVE_KICK_F50	833.19	EFFECTIVE_KICK_AM_relative	1219.87
INSIDE_50_AM_25per	1155.58	KICK_F50_relative	718.81	EFFECTIVE_KICK_AM_25per_relative	1166.78

MDGI = Mean Decrease in Gini Index. AM = Attacking Midfield, DM = Defensive Midfield, F50 = Forward 50, D50 = Defensive 50, *M* = True Midfield, EFF = Effective, PG = Possession Gain, 5per = Top 5% of players on team cumulative performance in variable, 25per = Top 25% of players on team cumulative performance in variable, 50per = Top 50% of players on team cumulative performance in variable. Further definitions available in supplementary material. Variables in bold are those with a negative correlation to the dependent variables (quarter margin).

possession is retained by a team to be key. All models improve on their Model set 3 baseline equivalent, attributable to these MG statistics, with prediction approaching levels seen with many more variables in M2.1 and M2.2.

Model 5.1 (M5.1) introduces feature distributions of player contributions, with Effective Kicks in total and attributable to the top 50% and 25% of players key. Model 5.2 (M5.2) shows similar results in relative form. Model 5.3 (M5.3) shows the importance of I50 in total and by the top 50% of a team, while the negative association of R50s is again present at 50% and whole team levels. Model set 5 iterations all produce results very similar to the Model set 3 equivalents, with little additional predictive power attributable to this additional data.

Residual plots of each model are available in Supplement 3, which suggest that better predictions are made when quarter margins are smaller, where a greater proportion of results are. Prediction improved largely uniformly regardless of margin albeit with slightly better prediction of larger margins (~20+ points) in the best performing models.

4. Discussion

This study sought to determine which incrementally more detailed technical and tactical match-play data better predicts AFLW match outcomes (win/loss and margin), and to identify the most important predictors of team performance. By using the most comprehensively detailed and longitudinal datasets within the AFLW to date, it enabled comparisons of the value associated with each level of detail, as well as when in combination with each other. Key variables for predictive team success have been identified in addition to general commentary on the usage of data going forward to assist with the previously outlined issue of the better utilisation and future allocation of scarce resources with a summary available in the conclusion in Table 5. Each model presented has practical value, meaning that interpreting each is beneficial, albeit some with greater predictive power and accuracy relative to others, although some of these improvements were obtained at the cost of a substantially greater number of variables utilised.

4.1. Value of detail levels in data

The addition of relative statistics between teams improves prediction substantially in all MAE and accuracy results. Important variable lists also primarily feature these relative statistics with high MDGI when applied, giving further weight to its importance. This is consistent with past observations in the AFL (Young, Luo, Gastin, Tran, et al., 2019) and AFLW (Cust et al., 2019). As a result, relative data should be considered an essential detail level for all team analyses of technical match-play performance.

The best prediction came through the application of all available variables including opposition statistics. While valuable, it is difficult to both access and manage all this data effectively when available. Notably, a lot of the value in the All Statistics dataset appeared to be derived from the MG variables also utilised in M4.1-M4.3. Evidence of this comes in both observation of the key variables of M2.1 and 2.2, as well as the MAE of Model set 4, which is the closest in performance to the All Statistics data. In the interests of data dimensionality, the usage of MG variables in addition to baseline variables is likely the

Table 5. Summary table of key practical insights surrounding usage of data details and key variables for coaches and sport practitioners.

Data detail	Conclusions surrounding usage	Key variables
Basic Statistics	Baseline variables commonly publicly available still produce a fair representation although utilisation of relative to other team statistics should always be introduced.	Inside 50s, kicks, effective Kicks, and clearances positive relationship. Turnovers and rebound 50s, negative relationship. All have more predictive power when expressed relative to opposition.
All Statistics	Produced best prediction by mean absolute error. Difficult to acquire all available data given usual restricted access.	Repeat Inside 50s, percentage of clearances won, and Metres Gained statistics covered below.
Advanced Statistics – Metres Gained Statistics	Found to account for a lot of the predictive value in the full dataset especially when interacting with field location. These variables represent the territory on the field gained by a team and should certainly be included or collected for future datasets.	Chain + Total Metres gained; Metres gained retained (i.e. ball remained in possession of the team from disposal). Relative form and in Forward 50 zone add more predictive power.
Relative between Teams	Found to make considerable improvement to prediction. Easy to implement and should be a level of data used for all analysis of team performance.	Covered through other data detail descriptions and seems to produce better prediction.
Locational Component	Should be sought wherever possible and potentially in more detail. Improved prediction substantially. Shows importance of context of where on the field an action occurs which affects variable importance. Caveat that Forward 50 zone seems to dominate prediction, which may be considered data snooping due to the requirement to get inside this zone to score.	Forward 50 locations highly positively associated with team success. Defensive 50 locations highly negatively associated with team success.
Feature Distributions of Player Contributions	Did not improve prediction when introduced. Can be difficult to create so may not be worth effort of producing. Individual player contributions likely still important but may need methodology that can represent whole dataset or alternate means like network analysis to be able to fully capture.	Top 25/50% player contributions to inside 50s.

best number of initial variables to start with, as it gives a fair representation of all available data in a more manageable way.

Zone of field data has previously been found to be influential in positional classification (Barake et al., 2021), with this study being the first time it has been applied in a match-play prediction context in women's AF. Zone of Field data was found to substantially increase prediction in all cases, as it is interacted with other variables. This shows the importance of the contextual information behind variables, with the location of Kicks, especially Effective Kicks, being a superior predictive variable relative to the non-locational version important in Model set 1. Further analysis of key variables in models using locational information finds that actions in F50, are the most predictive, which does suggest that the number of actions in that zone may be more important than the specific action undertaken in the zone. At the same time, the presence of Effective Kicks in F50, which represents effective ball usage in this area, is the most important aspect of match-play.

Removal of F50 data in the third iteration of Model sets 3–5 (M3.3, M4.3, M5.3) supports the view that getting the ball into this zone is of high importance, with a reduction in predictive power in all cases of removal, although quality predictions are still found. The presence and relative importance of I50 variables in these iterations may

still suggest that getting the ball into F50 is the biggest indicator of success (as it is more difficult to score a goal from outside the F50). This provides insight to utilise in terms of tactical practical application; however, it is also a factor to account for in future modelling interpretation due to it representing a variable that is a prerequisite to scoring. Overall, the capture and employment of this locational data should be prioritised, although care must be taken in its usage and interpretation of results derived from it.

While Model set 5 included feature distributions and locational components, the addition of these features did not improve prediction by MAE in the current study. This differs to findings reported by Cust et al. (2019), which suggested the importance of individual player contributions, particularly in the top 25% of the team (although locational components were not utilised in the model). This pulls into question the past assertion that individual key player performances contribute more to match-play success than team performance in the AFLW. To truly test this, alternative methodology should be sought, with the use of all individual player data through a method that can handle this dimensionality. One such method is neural networks (e.g. Convolution Neural Networks with multiple channels), albeit with issues surrounding large sample requirements and variable interpretation associated with its usage (Chollet, 2017). For now, this dataset requires additional feature generation that requires care in its use, which is likely not worth the comparatively marginal benefits that may be derived compared to results presented in this study.

4.2. Important variable interpretation

Model set 1 can be compared to previous literature with the caveat of a different time-frame and holdout test set limiting a truly fair comparison. In terms of MAE, Cust et al. (2019) reported a best result of 7.63 points on a holdout test set prediction of quarter margin using decision trees, while producing a best MAE of 5.12 points using Generalised Estimating Equations, although this was using training data. Models produced in this study improve on previous literature with the application of random forests, even on the comparable datasets of Model set 1. This may be in part due to the greater data availability, but our analysis outperformed previous literature on a test set, even when compared to performance on the training data in Cust et al. (2019), suggesting a greater robustness of our results.

Variables that were important in M1.1 and 1.2 shared some similarity with previous results (Black et al., 2019; Cust et al., 2019; Dwyer et al., 2022). Kicks, Disposal Efficiency, and Relative I50s were identified as important variables in previous models and remain so in our analyses. Other previously identified key variables, such as Uncontested Possessions (Black et al., 2019) and Contested Possessions (Dwyer et al., 2022), were not present in variable lists of either comparable model. Alternatively, Turnovers, R50s and Clearances were not previously deemed important in past analyses. This suggests that with a greater sample size of additional seasons, Turnovers, R50s and Clearances have become more prominent in influencing match outcome. Interestingly, Black et al. (2019) previously suggested that I50s were not associated with match-play success in the AFLW, a result directly contradictory to what is seen in not just M1.1, but the majority of our models. Such differences between the results of our study and that of previous literature may reflect the ongoing evolution of the game and how gaining possession of

the football from the opposition is becoming increasingly important regarding score margin and match outcome.

Dwyer et al. (2022) performed correlation significance tests, which do not allow for non-linear relationships like the random forests used in this research methodology. This is reflected in many important variables discovered in our analysis not being the highest in correlation to the dependent variable in initial exploratory data analysis or similar to what Dwyer et al. (2022) presented, showing the value of the non-linear analysis.

4.3. Usage of predictive modelling on aggregate match-play data

While this research presents results of key performance variables in available data, issues arise that largely stem from the application of aggregated data and predictive models. Many variables in these datasets are relatively obvious, high likelihood precursors to actions that result in increased win probability (e.g. an I50 is required to score a goal which naturally improves win probability). While a key variable, the contextual information of the process or chain-of-play that precedes an I50 may hold more value than these aggregate measures.

Approaching match-play performance from a technical and tactical perspective in this process-oriented manner is congruent with literature expressing match-play performance as dynamic systems in wider team invasion sport performance analysis (Travassos et al., 2013; Vilar et al., 2012). This may be achieved with the usage of analyses focused on the chain-of-play or states of play, for example through Markov modelling (Barkell et al., 2017; Meyer et al., 2006) or network analysis as has been done in the men's AFL (Young, Luo, Gastin, Lai, et al., 2019).

Despite these challenges associated with interpretation of results, there is still usefulness that comes from applying predictive models on aggregate team statistics. In this instance, variables like that of I50s, kicks and effective kicks in F50, and MG in F50 or MG Retained can be used as a state or objective to aim for which leads to positive results in match-play when employing process-oriented example methods like network analysis or Markov models. This is in addition to the presented aggregate metrics which can still be used by teams and practitioners as key performance indicators to assist in match analysis and pre-match preparation. It is evident though that great care should be taken in interpreting the insights aggregate measures produce, with value existing in results, but elaboration needed in following studies to verify or improve upon initial insights created.

Further contexts and combinations with other data sources like that of GPS performance and locational data, represent a way forward to produce more holistic representations, leading to more targeted, actionable insights for practitioners. Examples of this approach may include mapping of team structure, positioning, and its decision-making impacts, as is starting to be explored in the AFL (Alexander, Spencer, Mara, et al., 2019; Alexander, Spencer, Sweeting, et al., 2019; Spencer et al., 2019). With that in mind, challenges surrounding the utilisation of GPS data in the AFLW have been previously documented due to the requirement of team collaboration in a directly competitive league, otherwise results will be heavily biased to the available team's data (van der Vegt et al., 2023b). The next step to unlocking the ability to perform these analyses is greater care and depth of data collection within the AFLW; a difficult task in a league with limited resources compared to the AFL.

Despite the identified limitations and assumptions of this application, it produces the most robust representation of technical match-play to date while also presenting commentary on the detail captured in available data, assisting in future research and industry practices. Results should also be taken with the caveat for future use that this is a snapshot in time, meaning that relationships may need to be reassessed with similar methodology in the future as the game evolves. Individual team game styles may also influence the relevance of results; a factor to consider in future analyses.

5. Conclusion

This study provides the most in-depth analysis of women's AF technical and tactical performance to predict match outcome to date, in terms of detail and timeframe of data used. Findings bring clarity to the value of current data available, while suggesting areas to pursue for future research. Thirteen models were analysed, each representing different levels of detail available in the dataset. From each model, key technical variables contributing to match-play success were determined with commentary surrounding the value of individual levels of detail. This is summarised in Table 5.

Future research should expand upon these findings to explore more granular, chain-of-play data and consider methodologies like Markov models or network analysis. Important technical performance indicators presented can be applied by practitioners as a method of determining subsequent dependent variables that represent an advantageous state of play in future process-oriented models.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Australian Government Research Training Program Scholarship.

ORCID

Braedan van der Vegt http://orcid.org/0000-0001-6028-1797
Adrian Gepp http://orcid.org/0000-0003-1666-5501
Justin Keogh http://orcid.org/0000-0001-9851-1068
Jessica B. Farley http://orcid.org/0000-0003-1272-0833

Data availability statement

Data not available due to commercial restrictions. Due to the nature of the research, commercial supporting data is not available.

References

Alexander, J. P., Spencer, B., Mara, J. K., & Robertson, S. (2019). Collective team behaviour of Australian rules football during phases of match play [article]. *Journal of Sports Sciences*, *37*(3), 237–243. https://doi.org/10.1080/02640414.2018.1491113

Alexander, J. P., Spencer, B., Sweeting, A. J., Mara, J. K., & Robertson, S. (2019). The influence of match phase and field position on collective team behaviour in Australian rules football [article]. *Journal of Sports Sciences*, *37*(15), 1699–1707. https://doi.org/10.1080/02640414.2019.1586077

Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection [article]. *Knowledge and Information Systems*, *51*(2), 339–367. https://doi.org/10.1007/s10115-016-0987-z

Barake, A. J., Mitchell, H., Stavros, C., Stewart, M. F., & Srivastava, P. (2021). Classifying player positions in second-tier Australian football competitions using technical skill indicators [article]. *International Journal of Sports Science & Coaching*, *17*(1), 73–82. https://doi.org/10.1177/17479541211010281

Barkell, J. F., Pope, A., O'Connor, D., & Cotton, W. G. (2017). Predictive game patterns in world rugby sevens series games using Markov chain analysis [article]. *International Journal of Performance Analysis in Sport*, *17*(4), 630–641. https://doi.org/10.1080/24748668.2017.1381459

Black, G. M., Gabbett, T. J., Johnston, R. D., Cole, M. H., Naughton, G., & Dawson, B. (2019). A skill profile of the national women's Australian football league (AFLW) [article]. *Science and Medicine in Football*, *3*(2), 138–142. https://doi.org/10.1080/24733938.2018.1489140

Breiman, L. (2001). Random forests [article]. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chollet, F. (2017). *Deep learning with Python*. Manning.

Cust, E. E., Sweeting, A. J., Ball, K., Anderson, H., & Robertson, S. (2019). The relationship of team and individual athlete performances on match quarter outcome in elite women's Australian rules football [article]. *Journal of Science & Medicine in Sport*, *22*(10), 1157–1162. https://doi.org/10.1016/j.jsams.2019.05.004

Dwyer, D. B., DiDomenico, I., & Young, C. M. (2022). Technical performance in elite women's Australian football–comparisons with men's football, identifying important performance characteristics and apparent trends [article]. *International Journal of Performance Analysis in Sport*, *22*(1), 29–37. https://doi.org/10.1080/24748668.2021.1993029

Emmonds, S., Heyward, O., & Jones, B. (2019). The challenge of applying and undertaking research in female sport [article]. *Sports Medicine - Open*, *5*(1), 51. https://doi.org/10.1186/s40798-019-0224-x

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer. https://doi.org/10.1007/978-0-387-84858-7

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (1 ed.). Springer. https://doi.org/10.1007/978-1-4614-7138-7

Lord, F., Pyne, D. B., Welvaert, M., & Mara, J. K. (2020). Methods of performance analysis in team invasion sports: A systematic review [article]. *Journal of Sports Sciences*, *38*(20), 2338–2349. https://doi.org/10.1080/02640414.2020.1785185

Lundberg, S., & Lee, S. (2017). *DeepSHAP. Advances in neural information processing systems*, 30. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

Meyer, D., Forbes, D., & Clarke, S. R. (2006). Statistical analysis of notational AFL data using continuous time Markov chains. *Journal of Sports Science & Medicine*, *5*(4), 525–532.

Moreira Praça, G., Braga Jacinto, A. L., de Sousa Pinheiro, G., de Oliveira Abreu, C., & Teoldo da Costa, V. (2023). What are the key performance indicators related to winning matches in the German Bundesliga? *International Journal of Performance Analysis in Sport*, *23*(4), 284–295. https://doi.org/10.1080/24748668.2023.2227923

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Robertson, S., Gupta, R., & McIntosh, S. (2016). A method to assess the influence of individual player performance distribution on match outcome in team sports [article]. *Journal of Sports Sciences*, *34*(19), 1893–1900. https://doi.org/10.1080/02640414.2016.1142106

Spencer, B., Jackson, K., Bedin, T., & Robertson, S. (2019). Modelling the quality of player passing decisions in Australian rules football relative to risk, reward and commitment [article]. *Frontiers in Psychology*, *10*(JULY), 1777. https://doi.org/10.3389/fpsyg.2019.01777

Travassos, B., Davids, K., Araújo, D., & Esteves, P. T. (2013). Performance analysis in team sports: Advances from an ecological dynamics approach [article]. *International Journal of Performance Analysis in Sport*, *13*(1), 83–95. https://doi.org/10.1080/24748668.2013.11868633

van der Vegt, B., Gepp, A., Keogh, J., & Farley, J. B. (2023a). An investigation of data-driven player positional roles within the Australian football league women's competition using technical skill match-play data [article]. *International Journal of Sports Science & Coaching*, *19*(3), 1130–1142. https://doi.org/10.1177/17479541231203895

van der Vegt, B., Gepp, A., Keogh, J., & Farley, J. B. (2023b). Methods of performance analysis in women's Australian football: A scoping review. *PeerJ*, *11*, e14946. https://doi.org/10.7717/peerj.14946

Vilar, L., Araújo, D., Davids, K., & Button, C. (2012). The role of ecological dynamics in analysing performance in team sports [article]. *Sports Medicine*, *42*(1), 1–10. https://doi.org/10.2165/11596520-000000000-00000

Young, C. M., Luo, W., Gastin, P., Lai, J., & Dwyer, D. B. (2019). Understanding effective tactics in Australian football using network analysis [article]. *International Journal of Performance Analysis in Sport*, *19*(3), 331–341. https://doi.org/10.1080/24748668.2019.1605562

Young, C. M., Luo, W., Gastin, P., Tran, J., & Dwyer, D. (2019). Modelling match outcome in Australian football: Improved accuracy with large databases [article]. *International Journal of Computer Science in Sport*, *18*(1), 80–92. https://doi.org/10.2478/ijcss-2019-0005