

The influence of playing styles and team proficiencies on play call predictability in the NFL

A thesis submitted to Middlesex University in partial fulfilment of
the requirements for the degree of Doctor of Philosophy (PhD)

Ben Adam Carter

M00552804

School of Science and Technology
Middlesex University London

APRIL 2024

Abstract

National football League (NFL) teams are unusual in team invasion sports as they are an aggregation of sub-teams (offence and defence), hence research conducted on American football should reflect this fundamental difference. However, research has typically quantified technical behaviours related to offensive play. Current understanding is insufficient with respect to offensive and defensive patterns in play given that defensive performance has not been satisfactorily considered. Accordingly, this thesis developed a novel analysis of elite American football stemming from the identification of playing styles. A systematic review of literature revealed that unsupervised dimensionality reduction techniques were predominantly used in team invasion sports (84.21%). Therefore, principal component analysis (PCA) conducted on 5 seasons (2018-2022) of NFL event data, identified 8 distinct styles of play which accounted for exclusively offensive ($n = 5$) and defensive ($n = 3$) areas of performance. Subsequent cluster analysis of playing style efficacies demonstrated that teams could be classified into 2-3 proficiency clusters, to varying group membership distributions. Furthermore, multiple linear regression showed that 6 playing style proficiencies were significant predictors of regular season success. Conditional inference tree (CREE) models established the relative importance of 4 playing style proficiencies (passing efficacy, scrambling, rushing and deep passing) in predicting the likelihood of teams executing a pass, run or scramble play. Enhanced by playing styles efficacy among several situational variables, this machine learning approach, to develop a model of playing style efficacy, as a function of team quality, improved the applicability, and most notably, the accuracy, of decision trees as an analytical tool for use in the NFL by approximately 3.35-5.2% respectively for models including and excluding scrambling eventualities.

Acknowledgements

To my mother, Julie, to whom I must express my deepest thanks and appreciation for providing me with endless and unconditional moral support throughout my life, academic and professional pursuits. To my father, Stephen, who has continued to challenge, motivate, and focus my attention and efforts towards achieving any goals I set myself by always providing me with the support I needed, and prioritising what was best for me and my family, regardless of personal sacrifice. To my grandparents, for giving me their unreserved love, help and home away from home whenever I was in need.

To Dr. Nimai Parmar who has allowed me to develop as an academic and analyst through his efforts to present students with valuable work placements, organising interesting and impactful talks from field leading professionals, in addition to informative tutorials - allowing upcoming sport performance analysts to fulfil their potential. I cannot thank Nimai enough for the benefit I have received from his guidance through my journey - starting by analysing the Middlesex men's rugby team, through to becoming a first team analyst at Tottenham Hotspur and reaching my goal of being appointed to an academic post following the completion of my PhD. To Professor Nic James, I thank for his continued and valuable supervision, instruction and sharing his extensive wealth of knowledge in research alongside his experience in academia.

Contents

Abstract	2
Acknowledgements	3
Contents	4
List of Tables and Figures	7
Appendices	8
1 Introduction	9
1.1 Thesis Objectives	11
1.2 Thesis Outline	12
2 A systematic review of methods concerning the identification of playing styles in team invasion sports and their applicability in the National Football League	13
Abstract	13
2.1 Introduction	14
2.1.1 Rationale	14
2.1.2 Objectives	16
2.2 Methods	17
2.2.1 Eligibility criteria	17
2.2.2 Data collection and management	17
2.2.3 Selection process	18
2.2.4 Quality and data extraction	18
2.3 Results	20
2.3.1 Information sources and search strategy	21
2.3.2 Methodological quality	21
2.3.3 Sample and data types	21
2.3.4 Methodologies	22
2.3.5 Dimensionality reduction	23
2.3.5.1 Feature extraction	23
2.3.5.2 Principal component analysis	24
2.3.5.2.1 Sampling adequacy	24
2.3.5.2.2 Rotation	25
2.3.5.2.3 Communalities and factor loadings	25
2.4 Discussion	25
2.4.1 Identification of playing styles	26
2.4.1.1 Appraisal of methodologies	28
2.4.1.2 Principal component analysis	30
2.4.1.2.1 Suitability	30

2.4.1.2.2 Factor extraction	31
2.4.1.2.3 Factor loadings	32
2.4.1.2.4 Feature rotation	33
2.4.2 Interpreting styles of play	34
2.4.3 Application to the National Football League (NFL)	35
2.5 Conclusion	36
3 Identification and Interpretation of Play Styles in the National Football League	38
Abstract	38
3.1 Introduction	39
3.1.1 Rationale	39
3.1.2 Objectives	43
3.2 Methods	43
3.2.1 Sample	43
3.2.2 Procedure	43
3.2.3 Statistical analysis	44
3.2.4 Overview of statistical methodologies	46
3.3 Results	46
3.4 Discussion	51
3.4.1 Limitations and future research	54
3.5 Conclusion	55
4 Evaluating playing style proficiencies as a function of team quality and their contributions toward regular season wins in the National Football League (NFL)	56
Abstract	56
4.1 Introduction	57
4.1.1 Rationale	57
4.1.2 Objectives	60
4.2 Methods	60
4.2.1 Sample	60
4.2.2 Procedure	60
4.2.3 Statistical analysis	61
4.3 Results	64
4.4 Discussion	68
4.4.1 Limitations and future research	71
4.5 Conclusion	72
5 The influence of playing style proficiencies on play calling tendencies in the National Football League (NFL).	73

Abstract	73
5.1 Introduction	74
5.1.1 Rationale	74
5.1.1.1 Model selection	76
5.1.1.2 Variable selection	78
5.1.2 Objectives	79
5.2 Methods	79
5.2.1 Sample	80
5.2.2 Procedure	80
5.2.3 Statistical analysis	80
5.3 Results	82
5.3.1 Model accuracies	82
5.3.1.1 Overall model accuracy	82
5.3.1.2 Play-calling predictability	83
5.3.2 Variable importance	85
5.3.3 Case study	86
5.4 Discussion	87
5.4.1 Overall model accuracies	88
5.4.2 Situational predictability	89
5.4.3 Play-calling tendencies	90
5.4.4 Application of conditional Inference Trees	91
5.4.5 Limitations and future research	92
5.5 Conclusion	92
6 Discussion	94
6.1 Summary of findings	94
6.2 Limitations and recommendations for future research	95
6.3 Applications	96
References	98

List of Tables and Figures

2 A systematic review of methods concerning the identification of playing styles in team invasion sports and their applicability in the National Football League (NFL)	
Figure 1. Article filtering according to eligibility, inclusion, and exclusion criteria.	20
Table 1. Methodological quality assessment using the critical review form for quantitative studies	21
Table 2. Methodological quality assessment using the critical review form for systematic reviews	22
3 Identification and Interpretation of Play Styles in the National Football League (NFL)	
Figure 2. Scree test with non-graphical solutions (component retention PCA)	45
Figure 3. Overview of statistical methodologies used to derive styles of play	46
Table 3. Variance accounted for by principal components	46
Table 4. Factor loadings of principal component analysis with oblique (direct oblimin) rotation	48
Figure 4. Marker variables and proportion of variance explained by playing style	49
Table 5. Principal component correlation matrix	50
4 Evaluating playing style proficiencies as a function of team quality and their contributions toward regular season wins in the National Football League (NFL)	
Table 6. Clusters per style of play selected by majority rule	62
Figure 5. Overview of statistical methodologies used to cluster efficacy within styles of play as a function of team quality and assess their association with success	63
Figure 6. Playing style proficiency distributions	64
Figure 7. Boxplots of playing style proficiency classifications	65
Table 7. Multiple linear regression: Playing style proficiencies as predictors of success (regular season wins)	66
Figure 8. Multiple linear regression coefficient estimates (β) per playing style	66
Figure 9. Teams clustered according to (dis)similarity in playing style proficiencies	67
5 The influence of playing style proficiencies on play calling tendencies in the National Football League (NFL)	
Figure 10. Formulation of accuracy, precision, and sensitivity	81

Table 8. Conditional inference tree accuracy with(out) scrambling outcomes	82
Figure 11. PR precision and sensitivity per down	83
Figure 12. PR precision and sensitivity per quarter	84
Figure 13. Variable importance in the PRS conditional inference tree	85
Figure 14. PRS CTREE play probabilities (case study)	86

Appendices

A Keyword searches and combination of Boolean operators	122
B Automated duplicate removal and filtering script	122
C Summary of literary findings	123
D Performance variable operational definitions	127
E Styles of play and composite variable weightings (factor loadings)	130
F Example of cluster validation index used for majority rule	131
G Situational variable operational definitions	132
H Conditional inference tree (CTREE)	133
I Ethical approval letter	136

1. Introduction

Maintaining a high level of performance is considered the ultimate goal for teams performing in invasion sports (Sampaio, Drinkwater and Leite, 2010). Hence, while coaches' recollection of events both during and after matches has been highlighted to be limited to approximately 42% accurate, the use of data in sport has allowed for teams to review performance in an increasingly holistic manner (Franks and Miller, 1986). However, in order to accomplish this complex objective, notational analysis is frequently employed both in practical and research environments to quantify technical performance through the coding of behaviours, actions, events and outcomes (Carling, Williams and Reilly, 2005; Hughes and Bartlett, 2002; Hughes and Franks, 2004; Taylor, Mellalieu, James and Shearer, 2008). Indeed, with the aim of understanding game structure through the identification of important behaviours (Lames and McGarry, 2007), Vogelbein Nopp, and Hökelmann (2014) argued that performance indicators - defined as variables which must depict some or all aspects of performance (Hughes and Bartlett, 2002), provide a vital contribution to the objective quantification of sport performance. In turn, Hughes and Bartlett (2002) suggest that the collection of performance indicators (PIs) allows for comparative analysis to be conducted by analysts and coaches to retrospectively or prospectively assess performance (Vogelbein et al., 2014; Parmar, James, Hearne and Jones, 2018). Furthermore, key performance indicators (KPIs) are regularly employed to differentiate between winning and losing teams and can serve to ratify strategies, produce rationale for tactical decisions and enhance team performance (Csataljay, O'Donoghue, Hughes and Dancs, 2009).

Previous approaches to the analysis of American football may be considered a reflection of competition structure, whereby many performance variables collected predominantly detail behaviours exhibited by the team with possession of the ball (Dutta, Yurko and Ventura, 2020). However, contrary to many team invasion sports, teams are comprised of players with exclusively offensive or defensive roles requiring a wide range of skill sets (Robbins, 2011). Whereas the offensive team is awarded possession and an opportunity to sufficiently advance the ball or score within a series of set plays, the objective of the defensive team is to prevent them (Osorio, 2011). As such, an understanding of opposition behaviours, manifesting in the form of run or pass attempt

with the ball (Emara, Owens, Smith and Wilmer, 2017), are key to navigating team tactics and strategy in the NFL. Therefore, the predominant focus of investigations exploring performance in American football has been the prediction of offensive and defensive tendencies in the run and pass game, with the intent to inform formation and personnel selection - suggested to hold a decisive impact on match outcomes (Lee et al., 2016). As such, the prediction of play type has been the objective of predictive algorithms trained on contextualised plays using situational variables including score-line, down, yards to go, time remaining, formation and game location in addition to limited profile information such as prior pass completion and average yards per pass or run (Joash Fernandes et al., 2020; Lee et al., 2016; Goyal, 2020). However, notwithstanding the high level of randomness inherent to American football within these models (Blaikie, Abud, David and Pasteur, 2011), predictability between teams has been depicted as a limitation due to high variation in accuracy, suggested to originate from disparities in styles of play (Lee et al., 2016; Joash Fernandes et al., 2020). Whereas the presence of playing styles have been hypothesised, styles of play are yet to be empirically established in American football. Moreover, to date, outlines of playing styles in the NFL have been anecdotal and descriptive of overall approaches to the design of offensive and defensive plays such as 'west coast', 'wildcat' and '3-4' (Berri and Burke, 2011; Davis, Fodor, Pfahl and Stoner, 2014). However, Osorio (2011) suggests that the pre-coordination of plays defines styles of play for each team, what's more, that the ability for a team to succeed may depend on the extent to which teams can effectively perform under specific offensive and defensive styles of play. Therefore, an objective identification of playing styles, in addition to their hypothesised association with success may produce a previously unexplored key performance indicators in American football.

Though few, parallel efforts in association football have sought to determine the role of efficacy over several styles of play on team performance (Fernandez-Navarro, Fradua, Zubillaga and McRobert, 2019). Indeed, the findings observed by Fernandez-Navarro et al. (2019) indicated that the relative quality of the opposition was the only contextual factor that influenced all playing styles, and as such, teams exhibited lesser efficacy when faced with stronger opponents, in contrast to greater efficacy against weaker teams. Albeit, when applying this theory in American football, where teams are comprised of exclusively offensive and defensive players (Joash Fernandes et al., 2020) and offences dictate play through possession of the ball, it can be inferred that competition

between teams manifests a result of a defence's efficacy in preventing, and an offence's ability to impose a play style. Moreover, understanding the importance of playing styles as a function of team quality in specific situations may help coaches contextualise decision making during performances (Fernandez-Navarro et al., 2019). In the same way, what could be interpreted as playing style proficiencies may be of practical use when considered as situational variables - previously demonstrated to influence and interact with technical performance in team invasion sports (Taylor, Mellalieu, James and Shearer, 2008; Sarmiento et al., 2014; Gollan et al., 2018). For example, O'Donoghue's Interacting Performances Theory (2009) asserts that the quality and type of opposition influence a team's performance profile, confirming reports by McGarry et al. (2002) indicating teams exhibit unique traits depending on the opponent. Consequently, an exploration of team and opposition quality within established styles of play can provide further support decision-making processes and the design of plays - selected and executed by a team based on their expected probability of success.

Coaches in many team sports have limited opportunities to assess the impact of strategies during performances (Page, Fellingham and Reese, 2007; Correia, Araújo, Vilar and Davids, 2013). Subsequently, this dynamic is suggested to hinder the identification and synthesis of playing patterns - a source of errors in coaches decision-making processes (Franks and Miller, 1986). Conversely the intermittent nature of American football affords coaches pauses in play (for periods of up to 40 seconds) to formulate a response to the current in-game situation (NFL, 2008). Therefore, this difference in game mechanics presents an ideal opportunity to investigate the interactive influence of teams varying in styles of play and playing style proficiencies through adjustments made in response to fluctuations in match conditions. Accordingly, developing an awareness of an oppositions offensive and defensive play-calling tendencies informs decisions in the NFL (Lee, Chen and Lakshman, 2016).

1.1 Thesis objectives

1. Discern an appropriate technique for the identification of playing styles in the NFL.
2. Evaluate the methodologies applied to analogous team invasion sports and asses their applicability to American football.

3. Address the hypothesised presence of playing styles in the NFL and American football.
4. Determine the strength of relationships between resultant styles of play.
5. Objectively classify teams according to (dis)similarity within the underlying data.
6. Assess the predictive power of playing style proficiencies on total regular season wins.
7. Quantify the impact of team offensive and opposition defensive playing style proficiencies on the probability of play type selection.
8. Establish the variation in accuracy between models when including the potential for scrambling outcomes.
9. Gauge the performance of conditional inference trees for use as an analytical tool for practitioners in elite American football.

1.2 Thesis outline

To assess the presence of playing styles and their influence on team proficiencies in relation to play calling predictability in the NFL, several studies were conducted. To date, no previous investigations have sought to objectively establish styles of play in American football, therefore, the PRISMA protocol (Page et al., 2020) was exercised to undertake a systematic review of methodologies used to identify playing styles in team invasion sports and determine their applicability to the fundamental game mechanics of American football. This was followed by a study using dimensionality reduction techniques to develop and interpret a series of playing styles to reveal the underlying structure within NFL data and outline a novel framework for analysing technical performance in American football. An empirical definition of playing styles in the NFL thus produced a platform to compare teams and evaluate performance within latent areas of play as a function of team quality or proficiency and their ability to effectively operate under specific styles. Furthermore, the classification of playing styles as strengths or weaknesses for teams was shown to predict season success for many styles of play. The final section of the thesis was concerned with the influence of playing style proficiencies on the predictability of team play calling tendencies. As a result of implementing playing style proficiencies among other situational variables to predict play calling, substantial improvements in testing accuracy among interpretable machine learning models were observed.

2. A systematic review of methods concerning the identification of playing styles in team invasion sports and their applicability in the National Football League (NFL).

Abstract

Few studies have analysed team performance in the NFL. Indeed, no investigations into the identification of playing styles exist within literature published on American football. Therefore, the aim of this systematic review was to evaluate methods utilised to derive styles of play and assess their applicability to the NFL using the PRISMA protocol. A search of electronic resources returned 1543 articles, which, once filtered for duplicates and relevance, resulted in 21 studies. Analogous football codes demonstrated the necessity to transform high-dimensional data into feature subsets to quantify playing styles. Differences in dimensionality reduction techniques were observed, however, unsupervised techniques were predominant (84.21%). Principal Component Analysis (PCA) was the most prevalent approach, with its adoption supported within wider literature as an effective and accurate methodology. Albeit, concurring with extant evaluations of exploratory factor analysis (EFA) in sport science, procedural review demonstrated little variation in feature extraction (eigenvalues greater than one), rotation (varimax) and loading criteria (>0.6 and >0.7). Due to the game dynamics constituting American football, establishing styles of play requires greater consideration of valid rotation techniques. Furthermore, contrary to the ubiquitous application of orthogonal rotation in team invasion sports, potential correlations between styles of play in American football may necessitate the use of oblique rotations.

Keywords: Style of Play, Playing Style, Performance, PRISMA, Systematic Review, NFL, American football, Exploratory factor analysis

2.1 Introduction

2.1.1 Rationale

The increasing availability of large play-by-play datasets in sport has greatly augmented the range of accessible research questions (Han, Pei and Kamber, 2011). This presents a methodological challenge in sports such as American football afflicted by unsubstantiated theory concerning the relationship between variables, alongside an abundance of raw data pertinent to actions and strategies with the potential to reveal valuable insights (Zuccolotto, Manisera and Sandri, 2018). According to Jeon and Park (2021), data collected on complex practices in sport have become increasingly detailed and comprehensive, enabling unprecedented observations of novel microscopic behaviours. Notwithstanding, behaviours have been previously evaluated under the form of performance indicators (PIs) - variables accounting for certain or all facets of performance (Hughes and Bartlett, 2002). Moreover, these PIs allow for the exploration and classification of key performance indicators (KPIs) which differentiate teams based upon relationships with successful match outcomes (Csataljay, O'Donoghue, Hughes and Dancs, 2009) and enable analysts and coaches to objectively distinguish the most important areas of a given sport to guide further analysis (Lames and McGarry, 2007, Vogelbein et al., 2014).

However, while this form of univariate analysis permits an understanding of substantial differences between successful and unsuccessful teams, as well as variation among teams playing at a similar level or quality - it may oversimplify the complexity and minutiae observed in on-field performance (Brito Souza et al., 2019). As such, there may be potential for the collective influence of variables impactful performance indicators to be omitted or overlooked (Lames and McGarry, 2007). Therefore, investigating play in invasion sports in a multidimensional and context-inclusive manner may enhance understanding of the complex collective systems that embody team strategy (Gollan, Ferrar and Norton, 2018; Low et al., 2020). Progressively, patterns in play exhibited by teams through a combination of tactics and behaviours over time become recognisable and generate team-specific behaviours (Fernandez-Navarro, Fradua, Zubillaga and McRobert, 2018; Greenham, Hewitt and Norton, 2017) - defined as a game or play style (Gollan et al., 2018). Hence, PIs can be compiled, and serve as a metric to objectively determine and evaluate dependence on

specific styles of play. Furthermore, such profiles can be maintained as a reference from which teams may deviate from or adapt to, given a dynamic opposition (Fernandez-Navarro et al., 2016; Lago-Peñas, Gómez-Ruano and Yang, 2017; Gómez, Mitrotasios, Armatas and Lago-Peñas, 2018). Indeed, profiling styles of play may be utilised to identify dominant opposition behaviours and further employed to manipulate strategies and expose weaknesses to improve the likelihood of offensive and defensive success (Fernandez-Navarro et al., 2016; Lago-Peñas et al., 2017; McLean et al., 2017). Equally important is the consideration of situational variables which have been found to influence team performance and playing patterns (Fernandez-Navarro et al., 2018).

In American football, distinctive playbook-oriented styles of play based on certain personnel, positional permutations and formations (Teramoto, Petron, Cross and Willick, 2015) have been utilised to categorise playing styles in an attempt to derive relative injury risks. These structures from which styles have been interpreted include preconceived systems (devised by coaches prior to, or during the 1960s and 1970s) of play such as the West Coast or Air Coryell offences, as well as the 3-4 or 4-3 defences (Carroll, 1999). Albeit, while rudimentary styles of play have been utilised to categorised data (Teramoto et al., 2015), they do not present an empirical approach to differentiating teams according to collective patterns in play. Nonetheless, these ‘philosophies’ presented a framework surrounding the short and long pass attacks, in addition to aggressive and preventative defending strategies. Notwithstanding, in other team invasion sports, styles manifest through changes in behaviour due to teams being afforded larger timeframes to develop distinct spatiotemporal patterns (Low et al., 2020). Further distinctions can be made between playing styles, often illustrated by the actions and speed of, or intended ball and player progression patterns - categorised as elaborate and direct buildup possession characteristics in association football (Tenga, Ronglan and Bahr, 2010; Bradley, Lago-Peñas, Rey and Gomez Diaz, 2013, Collet, 2013; da Mota, Thiengo, Gimenes and Bradley, 2016). However, in American football, the classification and quantification of behavioural characteristics can be measured in substantially shorter intervals (Goyal, 2020). Whereas a majority of individual plays in analogous sports are organised on an ad-hoc basis over the duration of a possession, all plays in American football are coordinated, and frequently systematically formulated in response to contextual factors such as down, yards to go, and points differential (Joash Fernandes et al., 2020).

Given the abundance of action variables proliferating numerous aspects of NFL performance, characterised as having a high dimensionality (Johnstone and Titterington, 2009), certain statistical procedures serve to transform technical indicators into a meaningful representation of the data - most notably dimensionality reduction techniques (van der Maaten, Postma and van den Herik, 2009). For example, Cohea and Payton (2011) used factor analysis (principal component analysis) on 67 variables in American football to reduce the data to 14 principal components or features within the data. However, as the study aimed to evaluate the relationship between player, or collective team actions and match outcomes, the findings produced several factors representative of outcomes achieved per performance rather than accounting for how teams achieved the resultant values. While the findings developed a preliminary outline of the game mechanics in the NFL, little elaboration surrounding a team's performance characteristics were represented within the sampled PIs. While this demonstrates a potential limitation within the sample, research by Cohea and Payton (2011) reflects the relative importance of factors that can be reflective of team style of play or latent strategic information. Furthermore, greater granularity concerning situational detail corresponding with the sampled data, most notably the down upon which a pass, run or defence such as those included within models predicting of play types (Joash Fernandes et al., 2020; Jordan, Melouk and Perry, 2009; Lee et al., 2016; Teich, Lutz and Kassarnig, 2016) and match outcomes (Blaikie, Abud, David and Pasteur, 2011; David et al., 2011; Anyama and Igiri, 2015; Uzoma and Nwachukwu, 2015) may enhance the practical applications of analysis in professional American football.

2.1.2 Objectives

To date, styles of play in the National Football League are yet to be objectively established. An empirical development of playing styles may enable practitioners to ratify areas for improvement or manipulate team and opposition strengths and weaknesses (Fernandez-Navarro et al., 2016; Lago-Peñas et al., 2017; McLean et al., 2017). However, due to immediate differences in game play between American football and other team invasion sports (Goyal et al., 2020), this systematic review of literature aims to discern an appropriate technique for the identification of playing styles in the National Football League (NFL). To achieve this objective, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2020) will be

engaged to evaluate the methodologies applied to analogous team invasion sports and assess their applicability to American football.

2.2 Methods

This review of literature was conducted according to the protocols outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2020). The initial results produced over numerous databases and search engines were then processed with consideration towards their accordance with the inclusion criteria. No date restrictions were applied.

The studies ascertained for analysis and review were identified for extraction by adopting an amalgamation of keywords or terms: ‘playing style’; ‘style of play’; ‘collective systems’; ‘performance’; ‘analysis’; ‘sport’; ‘strategy’; ‘tactic’. Moreover, these were combined with the use of boolean operators: ‘AND’, ‘OR’ to refine and optimise searches of the aforementioned electronic resources - computing inter- and intra-group dependencies of key search identifiers respectively (Rojas-Valverde, Pino-Ortega, Gómez-Carmona and Rico-González, 2020).

2.2.1 Eligibility Criteria

Articles included within the sample were studies: (1) pertaining to the empirical identification/development of playing styles (i.e., no use of pre-determined styles of play); (2) analysing of team invasion sports; (3) sampling full matches (e.g., no small-sided games); (4) comprising of cross-sectional (systematic) reviews of literature; (5) written in or translated into the English language; (6) published in a scientific journal.

2.2.2 Data Collection and management

Once the primary results were returned from the electronic resources, if the database or search engine in use did not provide a function to export articles on mass (available in BibTex: .bib, Research Information Systems: .Ris, or direct export to RefWorks (refworks.proquest.com)),

individual citations were distinguished and scraped thereafter, through the use of the Mendeley Web Importer (2021, version 3.2.78), which features the ability to compute all information containing reference materials was used to extract pertinent data.

2.2.3 Selection Process

Post collation of search results, the citations were available to be extracted (in the .Ris format) along with ancillary information including the title, keywords, abstract, etc. These citations were then subject to automated filtering using functions from the base R and revtools libraries through the RStudio IDE (integrated development environment), wherein the references were tabulated through the R Project for Statistical Computing (R Core Team (2023), version 4.3.1). Prior to being exported in a BibTex file, the articles were refined through titles containing any of the following keyword or title string matches (Appendix B):

team behaviour; playing style; style of play; collective systems; game style sport; tactic; team; tactical; strategy; strategic; performance; league; match; indicators; interaction; player; play; game; technical; sports.

With the remaining articles prepared for title and abstract review, the remaining articles were re-imported to the Mendeley reference manager for manual examination and flagging for inclusion: firstly, reducing the results according to their value towards the research objectives and secondly, a more thorough inspection of the articles abstracts and their pertinence to the issue.

2.2.4 Quality and data extraction

To evaluate the quality of quantitative designs included within the systematic literature review, the guidelines for critical review form developed by Law et al. (1998) were used where appropriate. This format rates publications on up to 16 criteria, yet for the purposes of this review, the items were reduced to 11: Clear statement of purpose and/or research question (1), review of background literature (2), adequacy of design relative to the question (3), description in detail of sample (4), justification of sample size (5), reliability of results (6), validity of outcomes (7), reporting of

statistical significance (8), pertinent method selection (9), importance and impact (10), and result appropriate conclusions and implications (11). Each item appropriated from the aforementioned guidelines were allocated a standardised binary score (1 or 0), with a final percentage calculated - producing a representative rating for every study. Previous systematic reviews have exercised cut-off values to discriminate between 'excellent' (>75%), 'good' ($\leq 75\%$ and $> 50\%$) and 'low' ($\leq 50\%$) respective methodological qualities, however, these have historically been applied arbitrarily by authors citing the original guidelines (Te Wierike et al., 2013, Faber et al., 2016, Sarmento et al., 2018). Furthermore, due to the nature of the guidelines and variation in study design, answers for many items are possibly not applicable, therefore, with each item retaining an equal weighting, the overall quality score can easily be skewed with an impact on summary categorisations. As such, this literature review did not further analyse studies categorically according to methodological quality.

To critique the quality of extant systematic reviews, the Measurement Tool to Assess Systematic Reviews 2 or AMSTAR 2 (Shea et al., 2017) appraisal tool was consulted to examine their quality. This device enables an assessment to be made based on 16 items: PICO (Population, Intervention, Control group and Outcome) inclusion and description (1), explicit statement of review methodology (2), study design inclusion explanation (3), comprehensive search strategy description (4), duplicate author study selection (5), data extraction inter-reliability (6), index of excluded studies (7), adequacy of detail included and relevancy to research questions (8), satisfactory assessment for risk of bias (9), reporting sources of funding (10), appropriate statistical methodology for meta-analyses (11), assessing the potential impact of risk of bias on meta-analysis results (12), accounting for risk-of-bias in meta-analysis interpretation/discussion (13), if present, a coherent explanation for heterogeneity in the results of the review (14), investigation of publication bias (15), reporting issues including conflicts of interest and funding for conducting the review (16).

2.3 Results

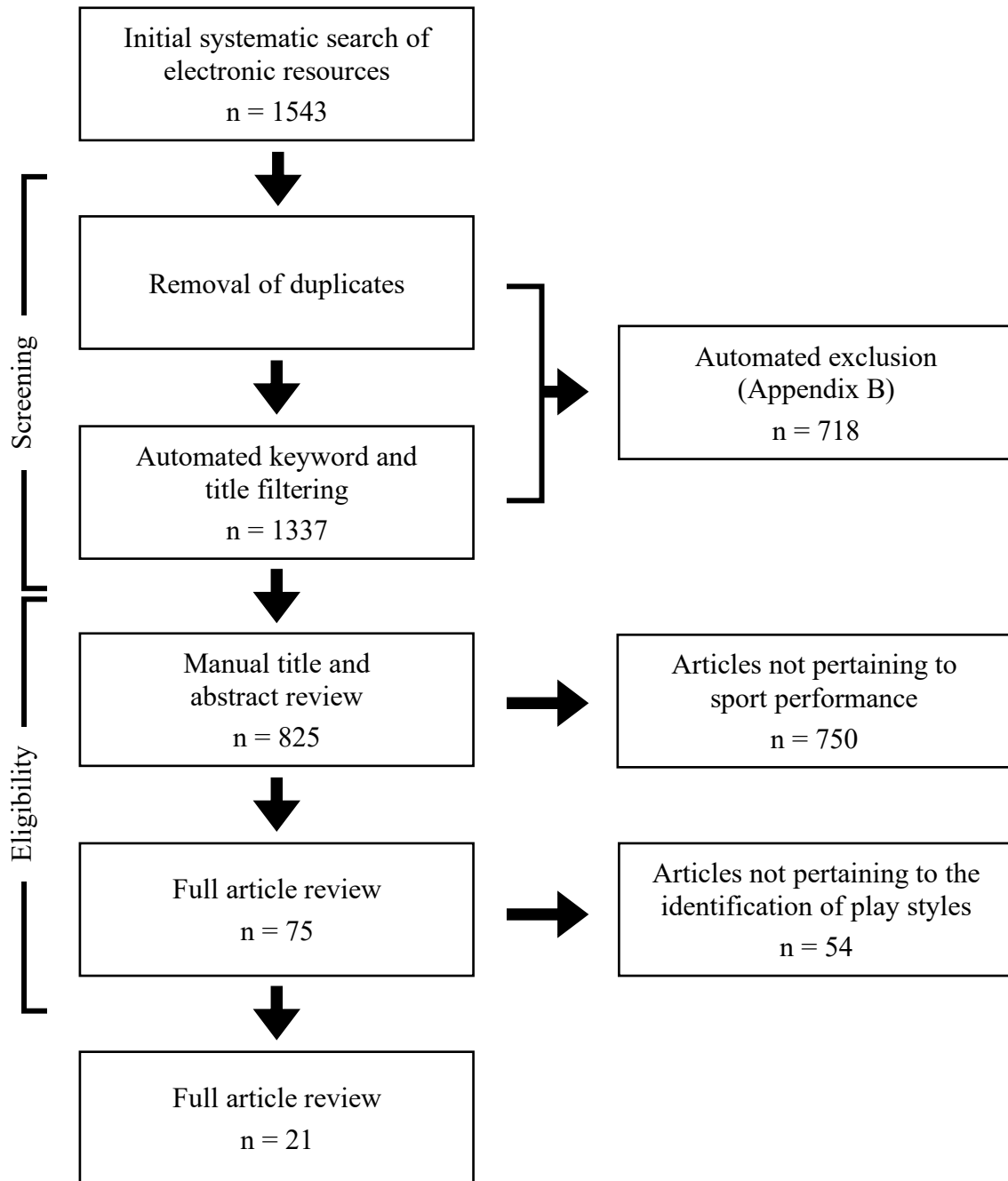


Figure 1. Article filtering according to eligibility, inclusion, and exclusion criteria.

2.3.1 Information Sources and Search Strategy

In order to extract research publications related to the identification of playing styles in team invasion sports, five disparate electronic resources were consulted for the identification of relevant articles, including four databases: Institute of Electrical and Electronics Engineers (IEEE) (n = 222), ScienceDirect (n = 231), Association for Computing Machinery (ACM) (n = 323), in addition to one search engine: Google Scholar (n = 769) were used. The final search was conducted in June 2023. The initial search returned a total of 1545 resultant articles. However, after the removal of duplicate studies (n = 206), automated title and keyword filtering (n = 512), manual title, abstract (n = 750) and full article reviews (n = 54), 21 publications were deemed eligible for systematic analysis (Figure 1). Of the resulting studies, 19 cross-sectional analyses were assessed according to the critical review form for quantitative studies (Law et al., 1998), with the remaining 2

Table 1. Methodological quality assessment using the critical review form for quantitative studies (Law et al., 1998).

Study	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	MQ
Amatria et al., 2021	1	1	1	1	0	1	0	1	1	0	1	73%
Cintia and Pappalardo, 2021	1	0	1	0	0	0	0	0	1	1	1	45%
Clijmans et al., 2022	1	1	1	0	0	0	0	0	1	1	1	55%
Decroos and Davis, 2020	1	0	1	0	0	0	0	0	0	1	1	36%
Fernandez-Navarro et al., 2016	1	1	1	1	1	0	0	0	1	1	1	73%
García-Aliaga et al., 2022	1	1	1	1	1	1	1	0	1	1	1	91%
Gollan et al., 2018	1	1	1	1	1	1	0	1	0	1	1	82%
Gómez et al., 2018	1	1	1	1	1	1	0	1	1	1	1	91%
He et al, 2023	1	1	1	1	0	1	0	1	1	1	1	82%
He et al., 2021	1	0	1	0	0	0	0	1	1	1	1	55%
Kong et al., 2022	1	1	1	1	1	1	0	1	1	1	1	91%
Lago-Peñas et al., 2017	1	1	1	1	1	1	0	1	1	1	1	91%
Plakias et al., 2023	1	1	1	1	1	1	0	1	1	1	1	91%
Ruan et al., 2022	1	1	1	1	0	1	0	1	1	1	1	82%
Sheehan et al., 2022a	1	1	1	1	0	1	0	1	1	1	1	82%
Sheehan et al., 2022b	1	1	1	1	0	1	0	0	1	1	1	73%
Sheehan et al., 2020	1	1	1	1	0	1	0	1	1	1	1	82%
Wedding et al., 2021a	1	1	1	1	0	1	0	1	1	1	1	82%
Wedding et al., 2021b	1	1	1	1	1	1	0	0	1	1	1	82%

C = criterion; C1 = clear statement of purpose and/or research question; C2 = review of background literature; C3 = adequacy of design relative to the question; C4 = description in detail of sample; C5 = justification of sample size; C6 = reliability of results; C7 = validity of outcomes; C8 = reporting of statistical significance; C9 = pertinent method selection; C10 = importance and impact; C11 = result appropriate conclusions and implications; MQ = aggregate methodological quality score.

systematic literature reviews assessed according to the AMSTAR 2 appraisal tool (Shea et al., 2017).

2.3.2 Methodological quality

The quality of cross-sectional studies (Table 1) included in the sample (n = 19) revealed that 12 articles' navigation of methods was 'excellent' (>75%), with 5 research designs classified as 'good' (≤75% and ≥50%), and 2 publications categorised as presenting with a 'low' (<50%) scoring methodology. An evaluation of systematic reviews (Table 2) meeting the eligibility criteria (n = 2) highlighted 1 study as 'good' (MQ = 73%) and 1 study as 'poor' (MQ = 45%).

Table 2. Methodological quality assessment using the critical review form for systematic reviews (Shea et al., 2017).

Study	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	MQ
Plakias et al., 2023	1	1	1	1	1	0	0	1	0	1	-	-	-	1	-	-	73%
Lord et al., 2020	1	1	1	1	1	0	0	0	0	1	-	-	-	-	-	1	64%

C = criterion; C1 = PICO (Population, Intervention, Control group and Outcome) inclusion and description; C2 = explicit statement of review methodology; C3 = study design inclusion explanation; C4 = comprehensive search strategy description; C5 = duplicate author study selection; C6 = data extraction reliability; C7 = index of excluded studies; C8 = adequacy of detail included and relevancy to research questions; C9 = satisfactory assessment for risk of bias; C10 = reporting sources of funding; C11 = appropriate statistical methodology for meta-analyses; C12 = assessing the potential impact of risk of bias on meta-analysis results; C13 = accounting for risk-of-bias in meta-analysis interpretation/discussion; C14 = if present, a coherent explanation for heterogeneity in the results of the review; C15 = investigation of publication bias; C16 = reporting issues including conflicts of interest and funding for conducting the review; MQ = aggregate methodological quality score.

A percentage error test was conducted to evaluate the inter-rater reliability of the methodological quality assessment. Repetition of the analyses was undertaken one month after the original analysis was exercised on approximately 15% of the articles included for review (n = 3). The results indicated 96.96% agreement between measures.

2.3.3 Sample and data types

The cross-sectional study characteristics indicated that literature published on playing styles in team invasion sports was comprised of research conducted in association football (n = 14), Australian rules football (n = 2) and rugby league (n = 2). Additionally, no cross-sectional articles meeting the inclusion criteria analysed more than one sport. However, the publications varied

greatly in the average competitions (2.37 ± 2.61), seasons (3.47 ± 1.9) and total matches sampled (1265.29 ± 2192.95). Interestingly, 68.42% of studies incorporated three or more seasons' data, whereas only 26.32% of analyses encompassed three or more competitions within samples (Appendix C).

The variable types employed to devise playing styles were constituted of predominantly event data ($n = 10$) and match performance indicators ($n = 9$). Exceptions to this data selection theme manifested in experimental study designs seeking to aggregate features such as contextual ($n = 2$) and physical ($n = 2$) variables, along with metrics describing network, sequence and spatiotemporal characteristics.

2.3.4 Methodologies

The techniques used to evaluate playing styles structurally underlying team invasion sports were widely comprised of principal component ($n = 12$) and cluster analysis ($n = 4$). The implementations of dimensionality reduction procedures were found to be well-suited to the classifications of playing styles, with representation in 63% and 26% of research designs respectively. Due to their respective procedures, the application of principal component analysis and cluster analysis allows for the development of game styles - providing simple and interpretable results. Comparably less established methodologies for the identification of playing styles - namely polar coordinate analysis ($n = 1$), discrete-time Markov chain ($n = 1$), t-distributed stochastic neighbour embedding ($n = 1$) and unsupervised artificial neural networks ($n = 1$) highlighted difficulties in simplicity of interpretation. Indeed, as methodologies to derive styles of play increase in complexity, the applicability and generalisability appear to decrease, thus hindering the potential for insights.

2.3.5 Dimensionality reduction

2.3.5.1 Feature extraction

A variety of means were exercised to ascertain an appropriate number of game styles - classified as either graphical ($n = 3$) or non-graphical ($n = 12$). Graphical factor retention solutions were

resolved through the plotting eigenvalues for PCA (Sheehan et al., 2022a; Sheehan et al., 2022b) or the sum of square errors for clustering data (Cintia and Pappalardo, 2021) on a y-axis along with the number of factors on an x-axis in order to identify a distinct point at which the curve plateaus - known as the elbow, denoting the elbow method.

Non-graphical factor extraction procedures were increasingly diverse. Eigenvalues were used as a feature cut-off, namely through the k1 criterion ($n = 5$) originally devised by Kaiser (1974), where only factors with eigenvalues >1 are retained. Nonetheless, studies by Wedding et al. (2021a; 2021b) extended this criterion to eigenvalues >1.2 in order to optimise model accuracy and decrease the quantity of resulting factors, albeit potentially to the detriment of variance explained by the extracted features (approximately 51-54% between articles). Moreover, co-consideration of eigenvalues ($k>1$) and variance explained was exercised at two criteria, $>60\%$ (Kong et al., 2022) and $>70\%$ (Fernandez-Navarro et al., 2016).

Determining a fitting number of clusters through non-graphical means was also achieved by combining Akaike and Bayesian information criteria (AIC and BIC respectively) and were adopted in unsupervised Gaussian mixture models (GMM) published by He et al. (2021; 2023). These procedures iterate over the data to produce models to strike a balance between model selection specialising in specificity, potentially overfitting (AIC) or simplicity, increasing the likelihood of underfitting (BIC) the data (Dziak et al., 2020). Whereas, in employing supervised k-means clustering, Gollan et al. (2020) used an exploratory technique, wherein an a priori number of clusters must be manually defined when conducting iterative clustering - relying on researcher interpretation to identify the most meaningful set of clusters, or styles of play.

2.3.5.2 Principal component analysis

2.3.5.2.1 Sampling adequacy

The Kaiser-Meyer-Olkin (KMO) test ($n = 10$) and Bartlett's test of sphericity ($n = 8$) are methods applied to determine the suitability of the data for factor analysis. Each publication addressing Bartlett's test of sphericity found significant differences between correlation matrices and null

correlation matrices, thus there were strong enough correlations within the data to reduce in dimensionality. KMO tests were conducted to assess partial correlations among variables, with results ranging on a scale of 0 to 1 where a minimum value of 0.5 may be deemed acceptable. According to Kaiser (1974), of the 9 studies reporting KMO values, 2 articles found ‘miserable’ values (<0.6), 3 articles were classified as ‘mediocre’ (<0.7), and the remaining KMO findings were ‘middling’ (<0.8).

2.3.5.2.2 Rotation

Studies reporting component rotation employed limited variation in their methodologies, with each study selecting the orthogonal varimax technique ($n = 10$). However, few articles rationalised rotation selection, yet 2 publications performed component correlation matrices to determine the appropriate rotation solution (Fernandez-Navarro et al., 2016; Plakias et al., 2023). This research design resulted in findings of negligible to no correlation between factors - justifying performing an orthogonal rotation such as varimax.

2.3.5.2.3 Communalities and factor loadings

Cut-offs in the sum of square component loadings or communalities were applied in the majority of articles utilising PCA to investigate the underlying structure of team invasion sports ($n = 7$). Findings from the review sample included studies retaining variables with minimum values of >0.4 ($n = 3$) - ranging to >0.5 ($n = 2$), and >0.6 ($n = 1$), with the greatest criterion of communalities of >0.7 observed among research designs. Within the residual components and their constituent variables, the most important features were predominantly derived using a cut-off value $>|0.6|$ ($n = 7$). One study set the retention criterion to loadings $>|0.7|$ (Fernandez-Navarro et al., 2016), yet 3 studies (Sheenan et al., 2020; Sheenan et al., 2022a; Sheenan et al., 2022b) deemed a lower level ($>|0.4|$) acceptable.

2.4 Discussion

This review aimed to identify literature pertaining to distinguishing playing styles in the National Football League (NFL), American football or analogous team invasion sports. The findings of the systematic search revealed a nihility of publications concerning styles of play in American football, whilst highlighting the interest garnered from sport scientists toward the analysis of association football ($n=15$) and a lesser emphasis on Australian rules football ($n = 3$) and rugby league ($n = 2$). Though few similarities in game dynamics are easily transferrable, the methodologies applied to the derivation of playing styles are applicable. Moreover, the nature of American football lends itself to evaluating the variation in style of play between teams and their uses in practice. Indeed, if identified validly and reliably, game styles in the NFL stand to provide a foundation for increasingly complex statistical procedures. Conventional means of capturing, managing and analysing data increasing in complexity and variability struggle to contend with the technological demands and speed of growth, materialising in ‘big data’ (Costa, 2014). The necessity for alternative procedures - most notably to accommodate time-efficient quantifications of training and competition performance through the reduction of ‘big data’ should be exercised (Rojas-Valverde et al., 2020). To achieve this in sport performance analysis, multivariate procedures such as factor analysis are adopted to explain myriad variables through a smaller number of extracted features (Federolf et al., 2014).

2.4.1 Identification of playing styles

Phases and sub-phases of a sport such as styles of play are representative of the underlying correlations between performance indicators. Due to the abundance of data available to analysts in the modern era, data reduction practices mitigate the issues faced in producing actionable insights from ‘big data’ (Plakias et al., 2023). To achieve this, a variety of methodologies were observed, yet a theme apparent among the sample studies was the use of dimensionality reduction (Appendix C). However, prior to the extraction of features or playing styles, a statistical procedure must be selected in accordance with the study design and sampled variables. These methods are broadly categorised as supervised, semi-supervised and unsupervised (Ang, Mirzal, Haron and Hamed, 2015). In cases where a feature subset has been pre-selected or labelled contingent on

meeting criteria, existing theory or hypotheses, a supervised process is undertaken to quantify the importance and relevance of a feature set in training a model (Tang, Alelyani and Liu, 2014). Contrarily, an unsupervised approach to feature selection involves the exploration of intrinsic and unlabelled structures within a data set described by characteristics such as distribution, correlations and variance to produce combinative features of the data (Gheyas and Smith, 2010). Alternatively, semi-supervised feature selection incorporates both labelled and unlabelled elements to improve or guide an unsupervised technique (Ang et al., 2015).

Applying supervised feature selection to playing styles in sport performance analysis cultivates challenges due to innate subjectivity manifesting from external knowledge originating from academics, analysts, coaches and other stakeholders within a sports organisation - resulting in a potentially unreliable, time-exhaustive, and costly process (Bø and Jonassen, 2002). The risks incurred through supervised feature selection include over-fitting a training model, thus limiting the generalisability of the factors or styles of play by unintentionally omitting relevant features or prioritising irrelevant features (Ang et al., 2015). Nonetheless, in practice, coaches may intend to analyse a sport through their own framework - designated as moments in play by Gollan, Ferrar and Norton (2018). Consequently, irrespective of empirical associations between variables, features may be promoted in alignment with coaching philosophy depending on their reliability and contextualisation within a corresponding moment of play (Hewitt et al., 2016). Appropriate procedures for the analysis of these structures depend upon the extent to which pre-designated features represent a majority or minority within the data classifies feature selection as semi-supervised or semi-unsupervised respectively (Ang et al., 2015). Empirically unassisted feature extraction is favoured by researchers in many fields as it allows for the inclusion of large data sets rather than restricting sampled variables based on preconceptions (Fernandez-Navarro et al., 2016; Lago-Penas et al., 2018; Winter & Pfeiffer, 2016; Gollan et al., 2018). Despite being considered more challenging than supervised and partially supervised approaches, unsupervised feature identification produces an unbiased and, in many cases, an enhanced account for the variance present within the data. Moreover, findings by Ang et al. (2015) in gene expression analysis revealed that partial labelling or no labelling of features was able to produce a good prediction of model performance - further suggesting that these approaches should be considered promising for future study designs.

In an indiscrete sample of studies attempting to reduce high-dimensional data, findings by Zebari et al. (2020) aligned with findings in the present review where authors commonly employed methodologies based upon intent to confirm existing theory, hypothesised structures or to explore data in search of associations between previously unassociated variables. These approaches focus on the optimisation of variance explained and prediction accuracy. For example, in feature selection - the outcome of supervised techniques, support vector machine (SVM) and k-nearest neighbours (KNN) achieved the best accuracy. For feature extraction - the output from unsupervised approaches, convolutional neural networks (CNN) and deep neural networks (DNN) perform well in classification tasks, albeit with challenges in interpretability and being susceptible to outliers and chance effects, similar to certain supervised learning methods (Livingstone, Manallack and Tetko, 1997). However, principal component analysis (PCA) remained more widely adopted for feature extraction (Appendix C). Furthermore, if performed correctly, PCA may achieve more effective results in prediction accuracy, efficiency in computational time and validity in the reduction of features (Zebari et al., 2020). Moreover, its ability to eliminate redundant information or noise from a dataset (Hasan and Abdulazeez, 2021) has led to the widespread use of PCA to quantify technique in sports (Gløersen et al., 2018; Parmar et al., 2018). Nonetheless, in cases where selected or extracted features have been refined from broader data sets, subsequent findings may be clustered or used in regression analysis to enhance understanding of sport behaviours (Oliva-Lozano et al., 2021; Rojas-Valverde et al., 2019).

2.4.1.1 Appraisal of methodologies

Multiple research designs elected to utilise cluster analysis ($n = 4$) - both supervised (Cintia and Pappalardo, 2021; Gollan et al., 2018) and unsupervised (He et al., 2021; He et al., 2023) techniques allowed for meaningful groups to be discovered within a dataset. However, when used in isolation, these procedures prohibit the researcher's ability to understand the relative power of predictors to classify members into groups or quantify their association within them (Satre-Meloy, Diakonova and Grünewald, 2020). Regardless of potential ambiguity, Cintia and Pappalardo (2021) were able to create a relational tool to assess similarities in coaching tactics requiring pre-existing knowledge to appropriately interpret.

Whereas PCA aims to maximise the variance present within the data, auxiliary methodologies such as t-SNE (t-distributed stochastic neighbour embedding) were represented within the sample studies as a mechanism for reducing the dimensionality of team performance (García-Aliaga et al., 2022). However, this method has been described as computationally complex and increasingly less efficient, commensurate with increases in sample size - becoming impractical for real-world use (Van der Maaten and Hinton, 2008). Furthermore, t-SNE is primarily utilised as a means of visualisation for grouping data according to similarity or dissimilarity and was incorporated by García-Aliaga et al. (2022) to discriminate team performance profiles, yet, due to its visual manner of representing differences, further comprehension of discrete features representative of playing styles may necessitate contributions from supplementary statistical procedures. Therefore, though t-SNE is established as a dimensionality reduction technique (Van der Maaten et al., 2009), as applied to the identification of playing style, it could be considered a clustering method and limited in its power to explore factors. As reflected in the results of this review, numerous techniques have been used to extract factors or game styles (Appendix C). Despite this, the results corroborate conclusions from Williams, Osman and Brown (2010), that PCA is one of the most used statistical techniques in sport - extending to the identification of playing styles, with 63% of authors utilising PCA as their primary approach. Notwithstanding the relative homogeneity in technique, many variations in the implementation of principal component analysis were apparent. To appraise the application of exploratory factor analyses (EFA) such as PCA, a 5-step Exploratory Factor Analysis Protocol developed by Williams et al. (2010) provides a platform to determine the suitability of the data, the quantity of, and criteria for factors to be extracted, in addition to the appropriate rotation for factors. Accordingly, in a review of articles using EFA in sport (n = 45), Rojas-Valverde et al. (2020) found that 100% of studies used PCA - representing a far greater proportion than research attempting to identify styles of play.

2.4.1.2 Principal component analysis

2.4.1.2.1 Suitability

It was proposed that a correlation matrix should be exercised to evaluate relationships between individual variables (Williams et al., 2010) and was found to be one of the most popular initial steps in EFA procedures (Henson. and Roberts, 2006). A review of the correlation matrix is used to determine the factorability (of R) of a dataset - allowing investigators to evaluate correlation coefficients and assess whether the factor analysis should be considered an appropriate statistical test. Tabachnick, Fidell and Ullman (2013) suggest a minimum of 0.3 for correlation coefficients indicating that a matrix is factorable and includes multiple sizeable correlations. Albeit, 0.3 was identified as a ‘minimal’ value by Hair, Black, Babin and Anderson (2010), with the authors presented further categorisations of 0.4 (‘important’) and 0.5 (‘practically significant’) as a rule of thumb. Contrary to the systematic review conducted by Rojas-Valverde et al. (2020) - finding numerous studies ($n = 10$) reporting factorability between 0.4 and 0.7 for variable selection, none of the articles undertaking EFA to define game styles presented correlation coefficients between independent variables. Alternatively, more sophisticated tests of establishing the factorability of the data were routinely practiced, potentially due to limitations in the capacity for bivariate correlations to reflect an underlying relationship between several variables simultaneously.

The Kaiser-Meyer-Olkin (1974) measure of sampling adequacy ($n = 10$) and Bartlett’s (1954) test of sphericity ($n = 8$) were administered in all but 2 and 4 study designs respectively. These methods were found to be implemented more frequently in the present review relative to their presence in approximately 50% of PCA procedures among sport science publications (Rojas-Valverde et al., 2020). However, Bartlett’s test of sphericity (1954) - testing a hypothesis that there are no correlations present in a correlation matrix is considered ‘highly sensitive’ with a commensurate likelihood to produce a significant result in substantial sample sizes (greater than 5 cases per variable), regardless of very low correlation coefficients (Tabachnick et al., 2013). Therefore, the implementation of Bartlett’s test of sphericity (1954) within studies analysing styles of play may be questionable. Use of the Kaiser-Meyer-Olkin test (1974) to demonstrate the adequacy of data for feature extraction may exhibit fewer shortcomings in application to the

derivation of playing styles as it formulates a ratio of the sum of squared correlations to the sum of squared correlations and the sum of squared partial correlations to produce an index or score ranging between 1 and 0 (with an adequate cut-off value of 0.5). In this systematic review, of 10 articles reporting Kaiser-Meyer-Olkin sampling adequacy, 5 articles found ‘miserable’ or ‘mediocre’ values (0.5 - 0.7), and the remaining studies ($n = 4$) indicated ‘middling’ (≥ 0.5) adequacy (Kaiser, 1974). This distribution of sampling adequacy was supported by Rojas-Valverde et al. (2020) indicating an even split of stated values by investigators in sport using EFA (>0.5 and >0.7).

2.4.1.2.2 Factor extraction

The outcomes of factor analysis are contingent on the quantity of factors (linear components calculated through the computation of eigenvalues of a matrix) retained (Field, 2013). Methods for determining the stopping point during factoring are subject to uncertainty and debate, albeit, extracted factors should be limited to those that account for non-trivial variance within the data (Rummel, 1970). The use of cumulative variance explained at a proportion of 50% was suggested as a cut-off for factor extraction by Streiner (1994). To this end, this systematic review highlighted 2 instances of articles using cumulative variance (ranging from >60 to $>70\%$). However, these were employed alongside the $k1$ criterion (Kaiser, 1970) - a method designed to ensure that the variance accounted for by the extracted factors is sufficient. Nonetheless, Kaiser’s eigenvalues above one criterion is considered as another rule of thumb but has been demonstrated to often provide incorrect estimates in larger scale study designs (Tucker, Koopman and Linn, 1969). Despite this, the $k1$ criterion’s use is widespread within research identifying playing styles, with 5 investigations (41.67%) limiting factors to eigenvalues >1 and an additional 2 studies (16.67%) selecting features with eigenvalues >1.2 . These findings were substantiated in analysing sport using EFA wherein 62.22% of investigations retained factors with eigenvalues >1 (Rojas-Valverde et al., 2020).

While non-graphical solutions such as cumulative variance and eigenvalues greater than one have historically received support as an appropriate approach (Kim & Mueller, 1978), Catell's (1966) scree test plots eigenvalues with their associated factors descending in magnitude to

produce a graphical criterion - arguing that factors preceding a sharp inflexion in the eigenvalues represent the most important features (Field, 2013). Though this technique was found to be somewhat reliable when strong factors arise (Stevens, 2002; Zwick & Velicer, 1982) and was exhibited in 2 studies sampled in this review, uniquely graphical solutions are susceptible to subjective application (Ford et al., 1986). Therefore, a technique called parallel analysis (PA) was formulated to distinguish factors with real eigenvalues greater than those expected from randomised data as an apt number of factors to retain (Humphreys and Montanelli, 1974). This method effectively compares features with real underlying structures with simulated data with no underlying structure (Ford, MacCallum and Tait, 1986). Though a majority of feature extraction techniques allow for investigators to contemplate which solution to favour, the implementation of imprecise and ambiguous techniques in team sport and game styles research is potentially problematic in producing valid analyses of game dynamics.

2.4.1.2.3 Factor loadings

Factor loadings are computed from eigenvalue-derived eigenvectors, thus, gauging the relative importance of variables within a feature (Field, 2013). The extent to which these loadings are deemed meaningful in EFA is contingent upon the cut-off specified by the investigators - deemed to be a matter of preference according to Tabachnick et al. (2013). However, Comrey and Lee's (1992) descriptions of factor loading cut-offs are frequently used as a standard for evaluating variable contributions to features. Under these criteria, loadings below 0.32 are not retained, however, values exceeding 0.32 are considered poor, above 0.45 fair, greater than 0.55 good, over 0.63 very good and in excess of 0.72 are considered excellent. Cases where cut-off values can be easily applied to naturally occurring gaps in loadings, renders applying criteria simple, however, in circumstances where the interpretation of features is complicated by a desire for greater loading quality, a variety of loading cut-offs may be expected - a trend observed in this systematic review (Appendix C). Variables with absolute factor loadings greater than 0.6 were the most common among attempts to derive playing styles (approximately 64% of publications), whereas >0.7, approaching excellent value were rarely utilised (n = 1). Other studies necessitated the use of a lower cut-off in order to achieve interpretable styles of play (n = 3). Though this sample may be too small to be generalisable, findings from PCA applications in sport science literature revealed

that loading cutoffs >0.6 and >0.7 were the most prevalent and equally adopted (Rojas-Valverde et al., 2020).

2.4.1.2.4 Feature rotation

Unrotated factors often complicate attempts to categorise related variables - resulting in many variables being grouped in the first feature (Richman, 1886). However, investigators have an opportunity to maximise high loadings and minimise low variable loading in other factors - optimising the likelihood of easily interpreting features and spreading loadings across several factors (Plonsky, 2015). The adoption of feature rotation techniques allows researchers to apply solutions to determine the most appropriate representation of patterns within the sampled variables (Guadagnoli and Velicer, 1988). Factor rotation methods can be categorised as either orthogonal, wherein factors are assumed to be uncorrelated, or oblique where rotated factors may be correlated (Fabrigar and Wegener, 2012). Common usages of solutions for orthogonal (e.g., Varimax, Quartimax or Equimax) and oblique (e.g., Direct Oblimin and Promax) rotations potentially stem from their availability on statistical software packages (Plonsky, 2015). The present review determined that all studies using rotated techniques employed orthogonal Varimax rotation to their factors ($n = 10$). Moreover, applied more widely to sport science, studies reporting factor rotation (55.55%; $n = 25$) implemented orthogonal Varimax rotation on 88% ($n = 22$) of occasions, whereas oblique Direct Oblimin ($n = 2$) or unspecified non-orthogonal ($n = 1$) techniques comprised the remaining 12% (Rojas-Valverde et al., 2020). In effect, most research in sport performance analysis appear to assume that all latent variables or features will be independent and uncorrelated. With respect to distinguishing styles of play in team invasion sports, this assumption extends the hypothesis that teams may or may not enact more than one playing style concurrently, which may be problematic in practice. Given the behaviours exhibited by teams must be conducive with laws of the NFL (e.g., offensively opting to attempt a pass or run play, defensively contriving a blitz, man or zone coverage), attacking and defending teams devise plays according to decisions, either predetermined by coaches or improvised by players (Goyal, 2020). Thus, adapting to collective behaviours observed from the opposition (Dutta, Yurko and Ventura, 2020; Joash Fernandes et al., 2020). Accordingly, it may be hypothesised that play style features may demonstrate correlations

- indicating that the prevailing selection of orthogonal Varimax rotation may not be appropriate for the study of game styles in American football.

2.4.2 Interpreting styles of play

Concerns reported by Sheehan et al. (2023) outlined potential asymmetries in the theoretical application of PCA and its use in practice, primarily manifesting in issues with the interpretation of EFA features (i.e., naming or labelling). However, a solution for such limitations by consulting the guidance of experts such as coaching staff to elucidate the findings of supervised, semi-supervised and unsupervised approaches has been suggested (Sheehan et al., 2023). Nonetheless, whereas univariate PIs allow for performance analysis to produce insights, PCA enables an explanation of original datasets proportional to the variance explained in the resulting principal components (Jolliffe, 2002). Furthermore, the sole consideration of technical performance indicators has been characterised as inadequate in sensitivity to distinguish playing styles in team sport, their constituent behaviours and organisation (Wedding et al., 2021). Additional context surrounding the intra-match variation in styles of play has been proposed as a solution for optimising the likelihood of successful match outcomes during performances (Wedding et al., 2021). However, caution should be emphasised in surmising game styles due to the potential for unfavourable patterns of behaviour, for example, an indicator of poor defensive performance for teams controlling a large proportion possession in a match (Sheehan et al., 2022a). Notwithstanding, Sheehan et al. (2022a) proposed that collecting data that accounts for the efficacy of attacking or defensive play has been proposed as a solution to accommodate such contexts. Albeit highly important in adapting to evolving match dynamics, the analysis of one-off matches increases opportunities for finding biases within the quantification of playing style (He et al., 2021). In a study of game style variability, He et al. (2021) associated style of play flexibility with team success - most notably with offensive outcomes, further suggesting that in practice, both attacking and defensive styles of play require a team to respond to the playing style enacted by their opponent.

He et al. (2023) elaborated on these findings in a study analysing teams according to perceived attempts to either predominantly impose or react to playing styles. However, teams

favouring either of these states exhibited decreases in performance, whereas teams showing an ability to adapt between imposing and reactive play styles depending on the strategic environment find greater success. Contrary to perceptions that defensive consistency and resistance to opposition or external perturbations lead to favourable performance outcomes, flexibility in defensive game styles has also been positively associated with KPIs and successful match outcomes (Zurloni et al., 2014; Hee et al., 2021). Furthermore, stability or consistency in playing style is hypothesised to lead to poor performance, rendering teams highly predictable.

2.4.3 Application in the National Football League (NFL)

As prescribed to rugby league (Wedding et al., 2021), the exploration of playing styles in American football may enhance analyses and allow teams to understand collective strategic behaviours as well as their variability and formulate approaches and responses to evolving match dynamics geared towards manipulating the likelihood of successful outcomes in the NFL. Albeit, the limitations and methodological restrictions highlighted in this systematic review may elucidate suitable study designs for the reduction of increasingly high-dimensional data collected to account for events and actions occurring in team invasion sport.

The importance of the data collected and included within samples should evidence adaptations in player behaviours to match conditions (Fernandez-Navarro et al., 2018). Moreover, to ensure the applicability of findings, more performances, competitions and variables should be incorporated into EFA research designs (Fernandez-Navarro et al., 2016; Lago-Peñas, Gómez-Ruano and Yang, 2017). Most notably, representations of influences relating to context detailing match venue, status and elapsed time are suggested to retain the potential to measure team proficiency in certain styles of play (Ruan et al., 2022). However, exercised in the NFL, these contextual factors extend further with down, distance to goal, and yards to go which are considered highly influential (Joash Fernandes et al., 2020). An issue present in many studies attempting to derive styles of play in addition to extant literature is the sole consideration of either offensive or defensive play, with omissions of this nature recommended to be addressed in future research (Lago-Peñas et al., 2017). Furthermore, the complexity of team actions can be considered the result of player interactions, therefore, competition between two teams is a result of complex systems

interacting, thus, neglecting opposition performance should be avoided in designs pursuing an accurate reflection of game styles (Sheehan et al., 2023). While EFA serves to empirically identify playing styles, the correct interpretation of features necessitates great attention, but can result in valid applications and practical use-cases in team invasion sports and football codes (Plakias et al., 2023). Indeed, an appropriate selection of features stands to improve the generalisation and comprehension of game styles and the prediction of performance (Ang et al., 2015). With very few analyses of technical performance indicators in the NFL despite myriad data collected from first and third parties, harnessing EFA (e.g., PCA) may play an important role in American football performance development (Ang et al., 2015) whilst embodying a fast, efficient and cost-effective solution in practical environments (Gutkin, Shamir and Dror, 2009).

2.5 Conclusion

Extant literature has concluded that combinations of performance variables enable the identification of playing styles and the description of structures underlying team sports. Discovering styles of play is enhanced by an abundance of data - necessitating dimensionally reduction techniques, the most prevalent of which was found to be principal component analysis (PCA). Unsupervised approaches like PCA augment the objective and empirical classification of game styles. However, alongside variable inclusion limitations, these methodologies have seen little apparent justification concerning the selection of feature extraction, rotation and loading. Nonetheless, once distinguished, resultant components may be utilised as integral performance characteristics for further analyses.

Manifesting as one of the most established forms of exploratory factor analysis, PCA can be appropriated to distinguish styles of play in the NFL. Notwithstanding, special attention to the methodological parameters requires consideration of the game dynamics evident in, and somewhat unique to American football. Notably, prevailing feature rotation techniques may not be pertinent due to the constraints incurred by exercising orthogonal rotation - assuming the independence of game styles. These assumptions may be problematic due to the dichotomous nature of non-special teams' play selection (i.e., run or pass), wherein play-calling tendencies may lead to positive or

negative correlations between potential styles of play. Albeit, these facets of study design demand precision in their execution to optimise the interpretability of results.

To complement variables selected according to professional or athletic experience, evidence-based measurements of playing styles allow coaches, analysts and scouts to evaluate the profiles and strategy - assisting in augmenting training design and preparation to face specific opponents. Objective classifications of team behaviours may be challenging to implement within practical environments. As such, developing an awareness of the impact caused by, and interactions between playing styles may be essential to effective in-game tactical adjustments.

3 Identification & Interpretation of Play Styles in the National Football League (NFL).

Abstract

In the NFL, the ability to predict opposition behaviours informs critical decision making. Moreover, understanding differences in offensive and defensive patterns of play has emerged as a limitation in American Football. Whilst the assimilation of match conditions, situational variables and game state have played a crucial role in predicting offensive and defensive play design, the absence of an empirical foundation from which to identify styles of play confound current perceptions of team interactions. Principal component analysis was used to explore underlying structures responsible for correlations among variables over a period of five NFL seasons (2018-2022, 2622 team observations). Optimal coordinates extraction derived 8 distinct features accounting for 57% of the total variance within the dataset. Direct oblimin (oblique) rotation identified a low, yet meaningful positive correlation between the short and long passing latent variables (.35). Offensively, yards before and after the catch or contact alongside third down success were associated with passing, rushing and scrambling related playing styles. Defensively, the inclusion of exclusively defensive action variables led to 3 factors interpreted as pass game suppression, defensive penetration and tackling proficiency. These findings provide a means by which team performance may be further analysed and potentially aid in predicting and rationalising play design.

Keywords: Style of Play, Playing Style, Identification, Definition, NFL, Sport, American Football, Performance Analysis

3.1 Introduction

3.1.1 Rationale

Advancements in granularity and availability of data collected in sport continues to enhance investigations - increasing the scope of potential research questions (Han, Pei and Kamber, 2011). A focus of contemporary analyses of American football have been aimed at understanding the outcomes of decisions in relation to match outcome and formulating novel variables to assess likelihood of drive and performance success (Biro and Walker, 2022). Dissimilar to many alternative football codes such as association football, rugby union or league, differences in game mechanics exist wherein possessions are comprised of individual plays interrupted for a period of 40 seconds. Therefore, researchers have sought to increase prediction accuracy in quantifying scoring or win potential using particular metrics, including Expected Points (Carter and Machol, 1971), Win Probability (Lock and Nettleton, 2014), Wins Above Replacement (Yurko, Ventura and Horowitz, 2019). Albeit, in practice, these methods serve predominantly as a retrospective means of contextualising the influence of on-field actions, developing a statistical approach to the prediction of behaviours in American football may enhance the rational underpinning play-calling in the NFL (Emara et al., 2017; Joash Fernandes et al., 2020; Lee, Chen and Lakshman, 2016; Teich, Lutz and Kassarnig, 2016).

Throughout the National Football League, understanding the factors contributing to the predictability of a team's actions provides coaches with actionable insights about the likelihood of collective opposition behaviours (Joash Fernandes et al., 2020). Whereas previous attempts to evaluate optimal tactical approaches to play-calling have compared teams' proficiencies in outperforming their opponents in either passing or running efficiency (May et al., 2010), such research designs, using outcome data (e.g., total passing/rushing yards, yards per pass/rush), may be limited in their predictive capacity, and as indicators of team tendencies. However, evaluating propensities in play type (i.e., passing or running game) has been the subject of several predictive models (Lee et al., 2016; Emara et al., 2017; Joash Fernandes et al., 2020; Goyal, 2020), predominantly attempting to incorporate situational variables (e.g., down, yards to go, match status, points differential, location and game clock), in addition to team profile characteristics over

the duration of a season (e.g., pass/run attempts %, pass completion %). Notwithstanding, these algorithms have demonstrated substantial inter-performance and inter-season variability (Lee et al., 2016).

Unprecedented and novel observations of novel microscopic behaviours have been enabled by increasingly comprehensive and abundant data (Jeon and Park, 2021). However, sports such as American football have limited research on performance analysis (Cohea and Payton, 2011) with a consequent paucity of substantiated theory regarding relationships between performance variables (Zuccolotto, Manisera and Sandri, 2018). Moreover, the exploration of complex behavioural data could be used to determine styles of play - hypothesised to account for the variability in play type prediction accuracy (Joash Fernandes et al., 2020). Playing styles are a manifestation of patterns in performance, systematically or tactically implemented by teams over the duration of matchups, competitions and seasons (Fernandez-Navarro et al., 2018; Gollan, Ferrar and Norton, 2018; Greenham, Hewitt and Norton, 2017). In equivalent football codes, styles of play have been established through the empirical reduction of indicators into latent variables (Decroos and Davis, 2020; Fernandez-Navarro, et al., 2016; Gómez, et al., 2018; Kong et al., 2022; Lago-Peñas et al., 2017; Ruan et al., 2021; Sheenan et al., 2022; Wedding et al., 2021), an objective technique to derive underlying structures in data or correlated areas of performance. Teams' relative dependence or proclivities towards these areas of play can be sustained as a profile from which teams may aim to deviate from or adapt to in accordance with contextual demands of a match and their opponent. Whereas examining differences in these features of offensive and defensive performance between teams has been highlighted as a limitation in American Football (Fokoue and Foehrenbach, 2013; Joash Fernandes et al., 2020), it has been a recent objective in association football - projected to assist in developing a comprehensive overview of play style dynamics (Lago-Penas et al., 2017; Gomez et al., 2018; Gollan et al., 2020). Nonetheless, personnel, formation and positional permutations have previously been used to categorise NFL teams according to playing styles and quantify relative injury risks (Teramoto, Petron, Cross and Willick, 2015). However, these styles of play were ascribed to teams according to perceived similarity with somewhat antiquated systems of play such as the West Coast or Air Coryell offences, as well as the 3-4 or 4-3 defences formulated by coaches in the 1960s and 1970s (Carroll, 1999). Though Teramoto et al., (2015) were able to contextualise their findings through a

‘philosophical’ framework, reflective of researcher interpretation of play designs (e.g., short or long pass attacks, aggressive and preventative defences), they were unable to empirically account for the game styles present in NFL performance.

Further challenges have been elucidated in the analysis of American football performance, namely in study design wherein defensive play is afforded proportionally scant consideration among action variables, overlooking apparent interactions between teams and potentially omitting half of the events during a match (Fokoué and Foehrenbach, 2013). The effect has been associated with a tendency to regard the dynamics of American football as though the defence is considered constrained to reactionary behaviours and the offence dictates play through possession of the ball (Joash Fernandes et al., 2020). Nonetheless, while it may be inferred that defensive and offensive game styles are subject to team interactions, offensive and defensive play-by-play metrics collected in the NFL are limited to observations of direct player involvements (Dutta, Yurko and Ventura, 2020). Cohea and Payton (2011) assessed 67 variables, reduced into 14 features of performance, with the aim of determining their association with success. However, this research design demonstrated limited practical applicability as the sampled variables were quantified in terms of outcomes without consideration of the actions occurring during the plays. Additional constraints of granularity for within play measurements has been demonstrated in the capacity of models to correctly predict scrambling attempts. For example, Lee, Chen and Lakshman (2016) identified that a substantial number of mistakes developed in their algorithm concerned teams with ‘mobile’ quarterbacks. These errors, amounting to 77% of occasions, originated on plays that were designed for passing attempts and ended with the quarterback running the ball.

Several approaches to the depiction of playing styles in team invasion sports stem from variation among input variables and their respective data structures (Carter, James and Parmar, under review^a). This has resulted in attempts to derive distinct phases of play over the duration of possessions through the use of spatiotemporal information to identify combinations of events such as ball movements, positional progressions and player involvement in sequences (Amatria et al., 2021; Cintia and Pappalardo, 2021; Clijmans, Van Roy and Davis, 2022). Findings from such approaches in association football have established patterns of play and elucidated strategies (e.g., elaborate and direct buildup). However, the classification of behaviours using spatiotemporal data

in the NFL can be measured over substantially shorter intervals, research has described elements of offensive and defensive play including route running trajectories, blocking vectors and receiver separation (Deshpande and Evans, 2020; Dutta, Yurko and Ventura., 2020; Haar et al., 2019; Ibrahim, 2021; Yurko et al., 2020; Nguyen, Yurko and Matthews, 2023). Albeit, while in analogous sports, individuals are able to implement or re-initiate a coach's game style throughout sequences, comparable plays in American football are coordinated - allowing for a systematic formulation of a response to contextual factors (Joash Fernandes et al., 2020). However, once plays have been implemented, post-snap (ball in play) actions quickly become unique with individuals reacting dynamically according to their respective or irrespective of attacking and defensive responsibilities (Dutta et al., 2020). Moreover, these methods often rely on the accessibility of GPS (Global Positioning System) or LPS (Local Positioning System) measurement of movements, presenting an impracticality in data collection due to limitations in team willingness to share resources (Lord, Pyne, Welvaert and Mara, 2020). Furthermore, according to Lord et al. (2020), the complexity of spatiotemporal metrics renders the applicability and ease of interpretation challenging for coaching purposes.

Contrarily, the simplicity and universal familiarity with technical variables allows for strategy and tactics to be ratified according to context (Lord et al. 2020). Notwithstanding, exclusive use of unidimensional performance variables has been characterised as inadequate with regard to sensitivity for identifying playing styles in team sports (Wedding et al., 2021). To accommodate for this effect, research on alternative football codes, namely association football, have engaged numerous methods to discern styles play (Carter et al., under review^a). The most notable and prevalent among these techniques is Principal Components Analysis (Gløersen et al., 2018; Parmar et al., 2018; Fernandez-Navarro et al., 2016; Lago-Penas et al., 2018; Winter & Pfeiffer, 2016; Gollan et al., 2018), an unsupervised approach to dimensionality reduction found to produce effective results in prediction accuracy, efficiency in computational time and validity in reduction of features (Zebari et al., 2020). Therefore, to understand styles of play and their respective compositions, technical data, used in conjunction with dimensionality reduction techniques, may produce an objective and easily interpretable exploration of playing styles in professional American football.

3.1.2 Objectives

The aim of this study was to address the hypothesised presence of playing styles in the NFL and American football, whilst reducing technical measures accounting for a greater degree of granularity than previous investigations. Moreover, this research design sampled dynamic offensive and defensive technical performance variables - incorporating several exclusively defensive variables to avoid offensive bias in game style interpretation. Furthermore, this study sought to determine the strength of relationships between resultant styles of play.

3.2. Method

3.2.1 Sample

Game-related play-by-play data from 1311 regular season games from the 2018 to 2022 seasons of the National Football League resulted in 2622 team observations. The data was obtained from third-party public information sources. These resources included nflfastR (2023) and pro-football-reference (2023) - repositories of scraped first-party official game summaries published by the NFL (2023). The original text-based data was comprised of raw play-by-play descriptions of on-field events and actions. Data scraping technique were exercised on NFL game summaries to generate performance and situational variables (Appendix D). Whereas certain variables published by nflfastR (Sebastian and Baldwin, 2024) were made available through the nflreadr (Ho and Carl, 2023) package for R, the remaining proportion of the data was ethically web-scraped from pro-football-reference (2023) on the public domain.

3.2.2 Procedure

Performance and situational variables ($n = 46$) were collated from the nflfastR and pro-football-reference datasets, aggregated, then statistically analysed using the RStudio Integrated Development Environment (IDE) for R (Posit team, 2023), powered by the statistical computing language R (version 4.0.5, R Core Team, 2023). Operational definitions were elucidated through the appraisal of definitions supplied by third parties in accordance the NFL official glossary of

terms (Appendix D). Ethical approval for this study was granted by the London Sport Institute Research Ethics Committee (REC) on 24/11/2021 (Appendix I).

3.2.3 Statistical analysis

Principal component analysis (PCA) is a form of exploratory factor analysis (EFA) frequently used in attempts to investigate the underlying processes responsible for generating correlations among variables, and to reduce the dimensionality of large, interrelated data sets to effectively summarise the correlations between the original variables, whilst minimising any potential loss of variation (Jolliffe, 2002; Tabachnick and Fidell, 2012). To achieve this, a correlation matrix must be computed to identify linear combinations of variables (eigenvectors), with eigenvalues quantifying the variance accounted for these latent variables or components within the matrix. Corresponding eigenvectors or factor loadings for each variable represent the strength of relationship, be it positive or negative to the associated feature (Comrey and Lee, 2013). Moreover, in team invasion sports, PCA has been exercised to identify the presence of, and interpret playing styles from the underlying structures of team performance (Decroos and Davis, 2020; Fernandez-Navarro, et al., 2016; Gómez, et al., 2018; Kong et al., 2022; Lago-Peñas et al., 2017; Ruan et al., 2021; Sheenan et al., 2022; Wedding et al., 2021).

The suitability of the data for dimensionality reduction techniques was assessed by the Kaiser-Meyer-Olkin (KMO) test of sampling adequacy (Kaiser, 1974), in addition to Bartlett's (1954) test of sphericity. Two non-graphical criteria, parallel analysis (Horn, 1965), simulating sampling error and optimal coordinates (Raïche et al., 2006), providing an objective interpretation of Cattell's (1966) scree plot were used to gauge the appropriate number of components to retain for analysis (Figure 2). Furthermore, the sampling adequacy measured using the Kaiser-Meyer-Olkin (KMO) test resulted in an overall value of 0.74 (>0.7) - classified as mediocre (Kaiser, 1974). Bartlett's (1954) test of sphericity demonstrated that the data was suitable ($\chi^2 = 1248.3$, $p < 0.001$, $df = 46$) for exploratory factor analysis (EFA).

Rotation methods were employed to maximise factor loadings within individual components, whilst minimising the corresponding variable in others to optimise the interpretability

of principal components (Plonsky, 2015). However, oblique rotation solutions should be explored in cases where rationale exists to hypothesise a relationship between components such as teams' proclivity in attempting certain play types relative to others (Field, 2007), therefore, the components were rotated using Direct Oblimin. Comrey and Lee's (1992) classifications of contribution quality were used to evaluate variables' association to the identified PCs - excluding variables with loadings lower than $|0.32|$ - accounting for approximately 10% of the variance within a component. In accordance with this method, scores exceeding $|0.32|$ were considered poor, above $|0.45|$ fair, greater than $|0.55|$ good, over $|0.63|$ very good and in excess of $|0.72|$ were categorised as excellent.

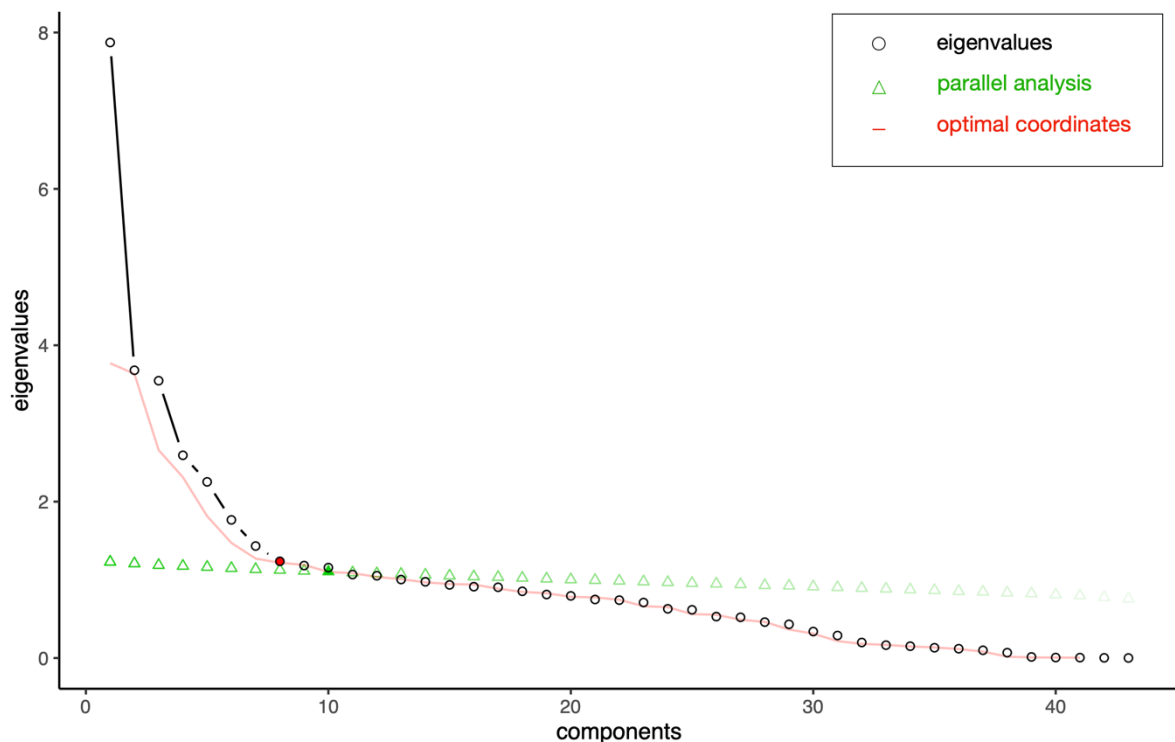


Figure 2. Scree test with non-graphical solutions (PCA component retention)

3.2.4 Overview of statistical methodologies

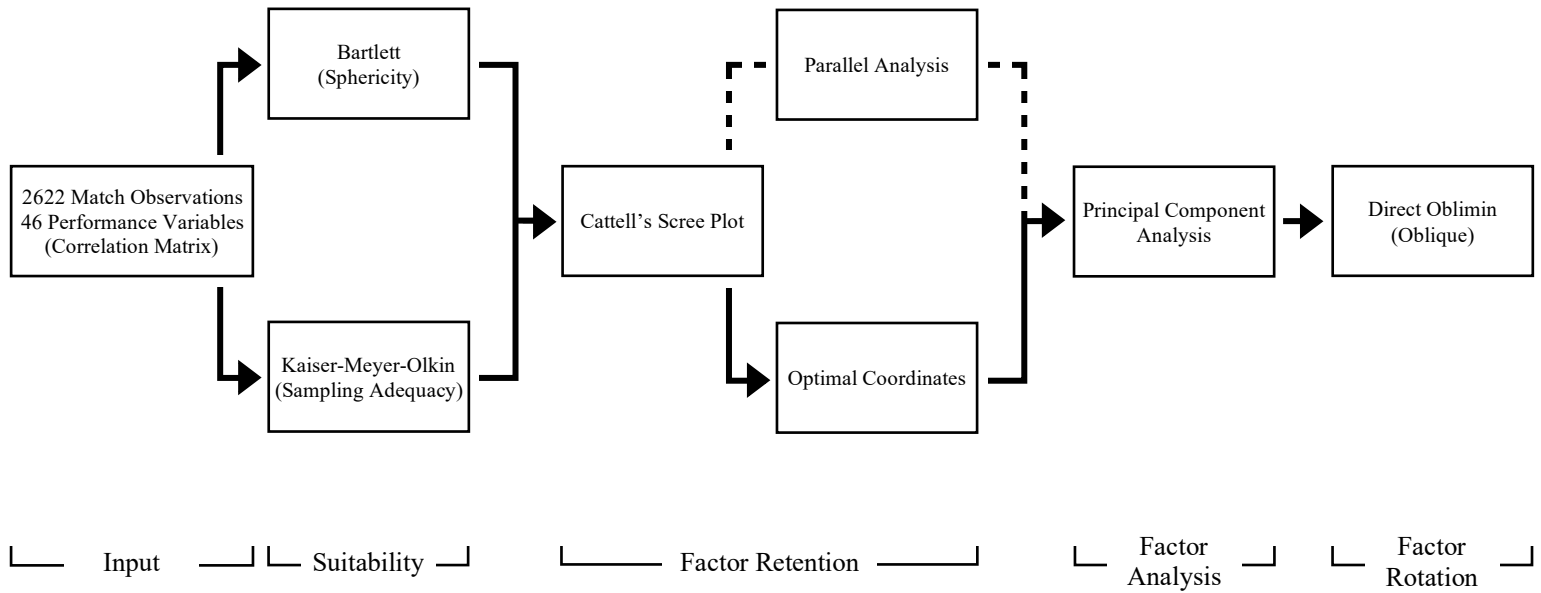


Figure 3. Overview of statistical methodologies used to derive styles of play in the NFL.

3.3 Results

Parallel analysis in the scree test for non-graphical solutions identified 10 distinct areas of play or playing styles (Figure 2), whereas, optimal coordinates, determined by a series of linear equations quantifying the difference between observed and expected eigenvalues demonstrated a clearer delineation ($n = 8$) between meaningful and unambiguous components (Ruscio and Roche, 2012). These factors accounted for 57% of the variance present within the match data (Table 3), while the 10 features determined by the parallel analysis explained a combined 2% additional variance.

Table 3. Variance accounted for by principal components

Variance	TC2	TC6	TC3	TC4	TC1	TC5	TC7	TC8
Sum of square loadings	5.05	4.01	3.40	3.54	2.78	2.50	1.73	1.38
Proportional variance	0.12	0.09	0.08	0.08	0.06	0.06	0.04	0.03
Cumulative variance	0.12	0.21	0.29	0.37	0.44	0.49	0.53	0.57
Proportion explained	0.21	0.16	0.14	0.15	0.11	0.10	0.07	0.06
Cumulative proportion	0.21	0.37	0.51	0.66	0.77	0.87	0.94	1.00

The first principal component (TC2) accounting for the greatest proportion of variance present among the match variables (12%) was comprised of predominantly passing related performance variables. This extensive passing feature included 5 variables with 'excellent' loadings or values $>|0.72|$ (completed passes, passing first downs, passing yards after catch, short passes and passing yards) and 2 variables providing good contributions $>|0.55|$ (pass attempts and quarterback drops). Lesser contributory action variables achieved fair loadings $>|0.45|$ (successful passing third downs and receiving tackles broken), in addition to a negative, yet poor association ($>|0.32|$) with combined defensive tackles (Table 4). A secondary passing factor (TC2) accounted for 9% of the variance in the PCA. However, contrary to TC2 with excellent contributions from short passes, 3 of 5 variables (deep passes, intended passing air yards, and passing air yards completed) constituting TC6 demonstrated particularly strong associations within the feature $>|0.9|$. Furthermore, these variables and excellent loadings were rotated with inaccurate and incomplete passes displaying good and fair values respectively.

The third principal component explaining 8% of the variance extant within the data incorporated 5 scrambling variables. Scramble yards, attempts and first downs produced the greatest contributions ($>|0.9|$) towards the factor (TC3). Additional variables concerning scrambling on third down attempts showed reduced importance to the component, albeit with disparate factor scores. The rotation resulted in contribution values classifying successful scrambling third downs as excellent, whereas unsuccessful scrambling third downs was represented as poor. Seemingly unrelated to the passing game, TC4 - accounting for 8% of proportional variance was comprised of rushing variables. Rushing yards, first downs, attempts, yards before represented the running game feature with excellent factor loadings. While they did not meet Comrey and Lee's (1992) highest criterion ($>|0.72|$), rushing yards after contact, successful and unsuccessful third downs provided a very good contribution. Further aspects constituting the rushing game style were tackles broken on running plays (0.38), while combined defensive tackles were negatively associated with the rushing component to a similar degree as TC2.

Table 4. Factor loadings of PCA with an oblique, Direct Oblimin rotation.

Variables	TC2	TC6	TC3	TC4	TC1	TC5	TC7	TC8
completed passes	0.90							
passing first downs	0.83							
passing yards after catch	0.82							
short passes	0.77				0.39			
passing yards	0.76							
pass attempts	0.61				0.38			
quarterback drops	0.61				0.38			
successful passing third downs	0.48							
receiving tackles broken	0.46							0.41
deep passes		0.97						
intended passing air yards		0.92						
passing air yards completed		0.92						
inaccurate passes		0.58			0.42			
scramble yards			0.92					
scramble attempts			0.92					
scrambling first downs			0.91					
successful scrambling third downs			0.72					
unsuccessful scrambling third downs			0.38					
rushing yards				0.94				
rushing first downs				0.89				
rushing attempts				0.79				
rushing yards before contact				0.74				
rushing yards after contact				0.71				
successful rushing third downs				0.61				
unsuccessful rushing third downs				0.50				
unsuccessful passing third downs					0.81			
incomplete passes		0.54			0.58			
quarterback hits					0.49			
unsuccessful passing fourth downs					0.39			
receiving dropped passes					0.35			
defensive completed air yards allowed						0.66		
defensive yards after catch allowed						0.64		
defensive quarterback knockdowns						0.49	0.39	
defensive combined tackles	-0.37			-0.41		0.47		
defensive blitzes						0.39	0.36	
defensive sacks							0.61	
defensive tackles for loss							0.51	
defensive interceptions							0.47	
rushing tackles broken				0.38				0.60
defensive missed tackles								0.58
defensive hurries							0.36	0.42

Principal components ordered according to proportional variance explained. Variables with no association to styles of play: successful passing fourth downs, successful scrambling fourth downs, successful rushing fourth down, unsuccessful scrambling fourth downs, unsuccessful rushing fourth down.

TC1 exhibited associations between passing and receiving related indicators - forming 6% of the variance within the unreduced data. Unsuccessful passing third downs (0.81) manifested as its greatest constituent; however, incomplete passes and quarterback hits were associated with TC1 with good (0.58) and fair (0.49) associations respectively. This feature was further loaded with meaningful, albeit poor ($>|0.32|$) contributions from inaccurate passes (0.42), unsuccessful passing fourth downs (0.39), short passes (0.39), pass attempts (0.38), quarterback drops (0.38), receiving dropped passes (0.35). Though only explaining 3% of the total variance, broken tackles during rushing attempts (0.6) and defensive missed tackles (0.58) were established as the most meaningful variable in TC8. However, defensive hurries and receiving tackles broken produced lesser (or poor) factor scores.

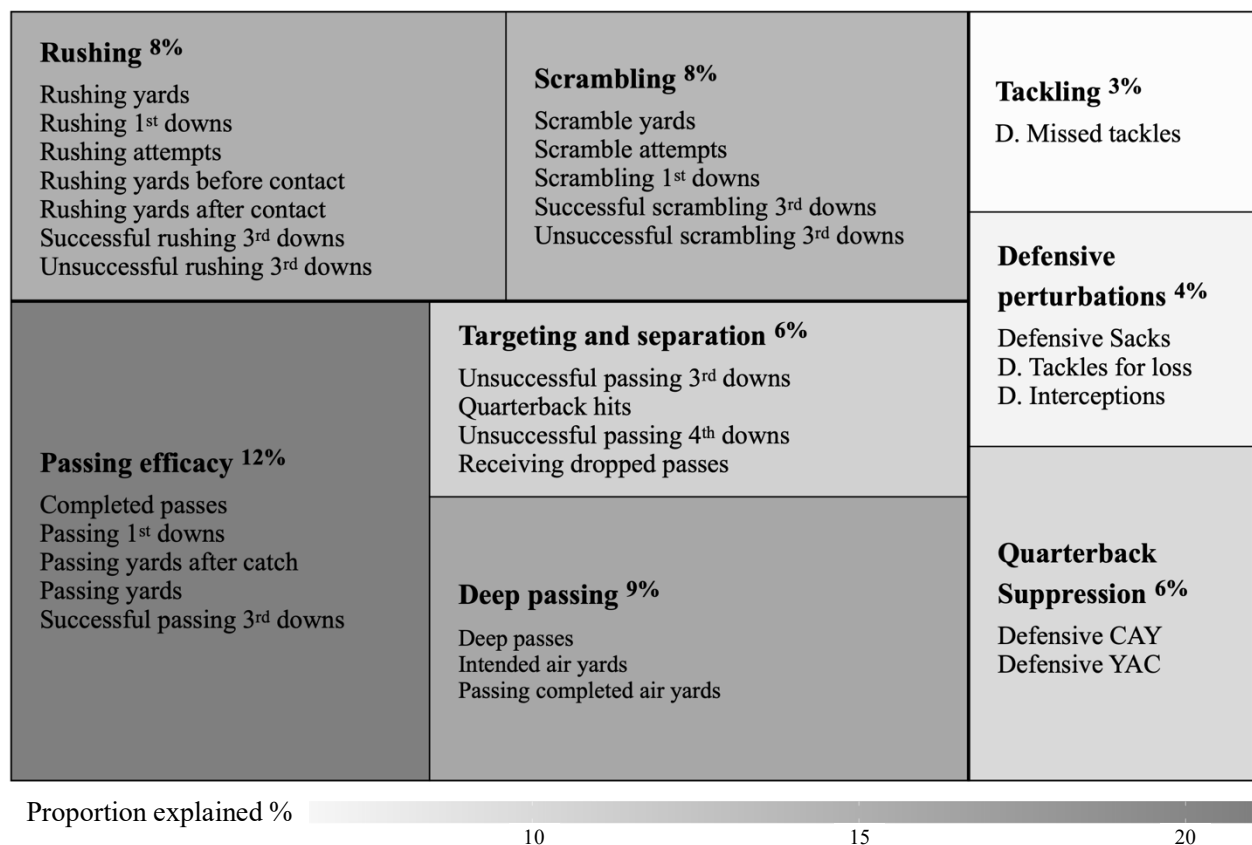


Figure 4. Marker variables, proportion of variance and proportion explained of identified principal components with playing style interpretations.

CAY: Completed Air Yards; YAC: Yards after catch.

The primary defensive component aggregated defensive completed air yards and yards after the catch conceded - enhancing the feature with very good loading values. Moreover, defensive combined tackles were associated poorly with the principal component (TC5). Notwithstanding, ancillary inputs demonstrating intent to reaching the quarterback including defensive quarterback knockdowns (0.49) and defensive blitzes (0.39) were represented within this latent variable. The second underlying defensive structure (TC7) retained variables associated with perturbations. Most notable among these these defensive behaviours sacks (0.61), tackles for loss (0.51), and interceptions (0.47). Furthermore, TC7 these defensive perturbations were substantiated by the positive association of defensive quarterback knockdowns, blitzes, and hurries as well as the negative loading of unsuccessful passing fourth downs exceeding the $|0.32|$ factor score criterion.

Oblique rotation (direct oblimin) conducted on the extracted features revealed the presence of one meaningful positive correlation (.35) between two passing features (Table 5). This correlation demonstrated a low effect between efficacy on short passing patterns in play (TC2) and the deep pass feature (TC6). Though the remaining factors were considered uncorrelated, negative relationships were observed between the running style of play (TC4) and passing related factors (TC2, TC6, TC1).

Table 5. Principal components correlation matrix.

PC	TC2	TC6	TC3	TC4	TC1	TC5	TC7	TC8
TC2		0.35	0.02	-0.16	0.12	-0.07	-0.05	-0.04
TC6			0.06	-0.25	0.25	0.08	-0.11	-0.03
TC3				0.08	0.09	-0.03	-0.08	-0.04
TC4					-0.21	-0.12	0.19	0.04
TC1						-0.11	-0.11	-0.05
TC5							0.00	0.07
TC7								0.04
TC8								

Where TC2 = passing efficacy; TC6 = deep passing; TC3 = scrambling; TC4 = rushing; TC1 = targeting and separation; TC5 = quarterback suppression; TC7 = defensive perturbations and TC8 = tackling.

3.4 Discussion

Through the reduction of performance variables, this study confirmed the presence of offensive ($n = 5$) and defensive ($n = 3$) playing styles hypothesised to exist in professional American football (Joash Fernandes et al., 2020) between the 2018 and 2022 National Football League seasons. These principal components were derived from 46 performance variables and explained a majority (57%) of the variance presence within the 2622 match observations - allowing for the identification and description of patterns in play reflecting discrete playing styles (Rein & Memmert, 2016). Regardless of feature rotation, certain performance indicators, outlined as marker variables, represent purer and exclusive measurements of factors relative to others (Tabachnick and Fidell, 2012). Furthermore, marker variables can be used to clearly define the nature of a factor and decrease ambiguity in their interpretation. The findings of the principal components analysis revealed three components related to the passing performance, two factors representing running performance and three defensive features.

Two passing components (TC2 and TC6) accounted for the greatest proportion of variance within the extracted features and data yet demonstrated clear differences in patterns of play. TC2 was characterised by marker variables including completed passes, passing first downs, passings yards after the catch, passing yards and successful passing first downs - performance indicators associated with passing efficacy. Whereas TC2 included playing style characteristics including short passes (<15 yards), quarterback drops and receiving tackles broken, TC6 presented with pure measures of deep passes (>15 yards), intended air yards and passing air yards completed, in addition to inaccurate and incomplete passes, albeit representing non-marker variables. Despite recording metrics including calculated measures and their composite variables (e.g., offensive passing completions, offensive passing attempts and offensive passing completion %) with inherent correlations, Cohea and Payton (2011) identified a factor labelled as 'offensive passing and total offence' including passing yards, first downs and completions which exhibited an association with passing efficacy. However, the present analysis enables a greater illustration of the patterns in play influencing passing performance - distinguishing two novel components within passing offence and delineating variables associated with short passing and efficacy (TC2), and long passing and accuracy (TC6). Moreover, these two passing features accounting for 38% of the

proportional variance were associated with one another, resulting in a low positive correlation (.35). Meanwhile, an uncorrelated third passing related factor marked by unsuccessful passing third and fourth downs, quarterback hits, and receiving dropped passes was established (TC1). With additional contributions to the feature including unsuccessful passing fourth downs as well as several performance indicators from the other passing styles of play. TC1 exemplifies a relationship between passing in pressing scenarios and the inability to complete passes. Indeed, as the likelihood of a pass play occurring supersedes alternatives on third (79.3%) and fourth (67%) down (Joash Fernandes et al., 2020), the predictability of play type selection appears to reduce the capacity for the passer and the receiver to create separation from defensive personnel and make a catch (Hear, 2019) - resulting in a feature representing receiver targeting and separation.

Quarterback scrambles typically occur on failed passing plays resulting in an unplanned and unforeseeable event (Reyers and Swartz, 2023). Yet, despite findings suggesting that quarterback rushing and scrambling ability is a key determinant of NFL success for college players, Craig and Winchester (2021) discovered that scrambling ability is systematically undervalued and overlooked throughout the scouting process. However, this study, incorporating scrambling related indicators in exploratory factor analysis, has revealed a feature exhibiting 5 marker variables denoting total scrambling yards, attempts, first downs achieved and success on third down as pure measures of scrambling or quarterback rushing ability (TC3). Similarly, the rushing factor (TC4) produced 7 marker variables. Moreover, yards achieved before and after contact exhibited a greater association with rushing performance than success on third down attempts - further corroborated by a negative correlation with opponent defensive combined tackles and a positive correlation with rushing tackles broken, albeit contributing 'poor' factor loadings. Though findings by Walsh (2021) accounted for avoided tackles and yards after contact among other action variables through an overlap of extraction techniques designed to describe running back positional profiles, rushing yards before contact provides a novel and empirically established characteristic of rushing performance in the NFL.

Many quantifications of professional American football performance omit defensive analysis, fewer still observe uniquely defensive measures (Dutta et al., 2020; Walsh, 2021). In cases where defensive actions are represented within a sample, they are often simply parallels of

offensive metrics (Cohea and Payton, 2011). While there may be a disassociation between offensive and defensive intra-team players, defensive interactions have been anecdotally construed as reactionary - inhibiting teams' capacity to impose a defensive style (Joash Fernandes et al., 2020). However, this study elucidates the presence of 3 defensive styles of play accounting for 23% of variance explained by the retained principal components. TC5 was associated with two marker variables comprising of completed air yards and yards after the catch conceded alongside lesser contributions from defensive quarterback knockdowns, combined tackles and blitzes which may be interpreted as quarterback passing suppression. Opposing the suppression of the quarterback in the passing game, a pattern in play comprising of defensive perturbations was established (TC7). Markers for this pattern of play manifested predominantly in sacks, tackles for loss and interceptions. Moreover, this feature comprised meaningful contributions from blitzes, hurries and quarterback knockdowns - penetrative defensive actions possibly leading to interception or turnover opportunities. Teams' capacity to contain ball carriers materialised in the final defensive factor (TC8). While Berri and Burke's (2011) appropriation of James' (1974) range factor has been proposed to predict and evaluate individual tackling efficacy (tackle factor), variability according to positional differences renders the metric impractical and limited in generalisability. Therefore, a pure measure of missed tackles associated with rushing and passing broken tackles in addition to defensive hurries provide a first depiction of team tackling performance in the NFL.

An understanding of composite performance indicator contributions has been hypothesised to elucidate the variability in observed strategic behaviours (Wedding et al., 2021). This exploration of playing styles in the National football league has demonstrated the presence of underlying structures in play based upon technical variables. These styles of play may enable the navigation of evolving match dynamics and provide a means by which tactics may be manipulated to increase the likelihood of successful outcomes. Indeed, whereas previous research has determined the importance of total offensive and defensive prowess in the pass and run game (Cohea and Payton, 2011, May et al., 2010), this study has investigated the characteristics of these patterns in play. Offensively, team efficacy on third down on pass, run and scrambling opportunities were associated with play types while demonstrating a clear importance within their respective styles of play. These findings contradict previous research by Cafarelli, Rigdon and

Rigdon (2012), which identified third down success rates as almost meaningless. Notwithstanding, the ability to gain yards before and after contact as well as achieving yards after the catch provided meaningful additions to game styles. With the inclusion of uniquely defensive actions within exploratory factor analysis, teams appear to adopt three clear playing patterns wherein efforts are made towards the suppression of the opposition passing game, containing ball-carriers in open play or penetrative efforts towards the quarterback - manifesting in perturbations. It has been established that mixed strategies are employed by NFL teams, confirming that game styles are not executed without regard for opponent actions (McGarrity and Linnen, 2010). Therefore, further analysis aimed towards evaluating teams' ability to implement a variety of playing styles such as those identified in the present study may be used to quantify opposition strengths and weaknesses in specific areas of play. Moreover, a combination of multivariate solutions, used to provide easily interpretable results may be used to inform strategic preparation for matchups, or adjust tactics during games.

3.4.1 Limitations and future research

With passing styles accounting for 48% of the proportional variance, scrambling and rushing factors 29%, and defensive features explaining 23%, greater detail and equal representations of performance concerning the variables included within analyses may result in a more elaborate explanation of playing styles in American football. The relative predictive power of the identified playing styles may therefore be restricted by the granularity of the data collected. This effect is particularly notable in scrambling and defensive areas of play. Whereas Byanna and Klabjan (2016) produced spatiotemporal measures to quantify blocking performance, there is a dearth of publicly accessible technical metrics regarding offensive and defensive output. Furthermore, the complexity of these analyses remains challenging for practical implementation (Lord et al., 2020). This effect has been noted by Dutta et al. (2020), wherein a majority of data collected on NFL events is constrained to actions performed by, or towards the ball carrier - omitting essential information.

While equivalent approaches are required to corroborate these findings in college football, the use of unsupervised techniques to discern the mechanisms underlying the variation in performance patterns provides a platform to understanding team interactions in professional

American football. However, adopting these styles and the relative quality within them exhibited by teams may be harnessed to enhance evaluations of team performance in the NFL. Equally important, engaged alongside situational variables (e.g. match location, yard line, down) and game states, quality exhibited under several playing styles may inform the predictability of team and opposition behaviours with respect to realistic in-game scenarios (Gollan et al., 2020).

3.5 Conclusion

Whereas alternative football codes have engaged numerous methods to discern patterns in play and performance tendencies, no such investigations concerning American football have been published. This study empirically derived eight playing styles in the National Football League, accounting for several areas of play through associations between technical variables. These predominantly uncorrelated latent variables provide a simple and interpretable structure from which team performance may be further analysed - providing a tangible tool for the design of specific training drills (Fernandez-Navarro et al., 2016), the accurate simulation of competition scenarios and the improvement of proactive offensive or defensive decision making by coaches (Sampaio et al., 2010).

4 Evaluating playing style proficiencies as a function of team quality and their contributions toward regular season wins in the National Football League (NFL).

Abstract

In elite American football, teams can be observed as a combination of sub-teams with players categorised according to offensive or defensive skillsets and responsibilities. Consequently, efforts to provide an overall measure of team quality may present an inadequate, invalid and potentially ineffective platform to improve team, or expose opposition weaknesses. Therefore, the aim of this study was to cluster styles of play ($n = 8$) to identify partitions among teams as a function of team quality over 5 seasons in the NFL. Cluster validation determined that by majority rule, 75% of playing style were best classified as a dichotomy (stronger or weaker), while 25% demonstrated a trichotomy (strong, medium, weak) as the optimal partition. Multiple linear regression was conducted on the resultant team playing style proficiencies ($n = 160$) to predict regular season wins. Deep passing and scrambling proficiency did not produce a significant prediction of wins ($p > 0.05$), while the strongest predictors ($p < 0.001$) were passing efficacy and targeting and separation. However, defensive perturbations was the only negative predictor of regular season wins ($p < 0.001$). This investigation developed a novel technique to evaluate teams according to quality through the use of playing styles. Moreover, these findings and approach may be implemented in practice to contextualise analyses.

Keywords: Team Quality, Playing Style, Cluster Analysis, Regression, NFL, Sport, American Football, Performance

4.1 Introduction

4.1.1 Rationale

In American football, the predominant focus of research has been to predict match outcomes (Blaikie, Abud, David and Pasteur, 2011; David et al., 2011; Anyama and Igiri, 2015; Uzoma and Nwachukwu, 2015), create discretionary metrics aimed at fan comprehension of play such as quarterback rating (Stimel, 2009; Mettling, 2014), or the likelihood of a particular play type being selected (Lee, Chen and Lakshman, 2016; Goyal, 2020; Joash Fernandes et al., 2020). However, in pursuing increased prediction accuracy, highly complex and obscure machine learning techniques used to predict success may conflict with efforts to understand the causal effects of performance indicators (Blaikie et al., 2011). Hence, in practice, many advanced means of modelling performance may not necessarily provide realistic examples of how organisations seek to adapt and design training techniques to address areas for improvement (Woods et al., 2020; Mackenzie and Cushion, 2013). On the other hand, oversimplifying analyses to a univariate level when establishing the difference between successful and unsuccessful teams may in turn cause subtle variations and minutiae in technical differences between teams to be overlooked (Brito Souza et al., 2019; Low et al., 2020).

In attempting to address this issue in association football, Yang, Leicht, Lago and Gómez (2018) established that little research had examined the patterns in play associated with success. Furthermore, among football codes, previous investigations have determined that technical performance may be a reflection of the playing standard and style of a team (Lago, 2009; Lago-Ballesteros, Lago-Peñas and Rey, 2012; Bradley, Lago-Peñas, Rey and Sampaio, 2014). More specifically, styles of play are thought to emerge through an amalgamation of strategy and tactics exhibited throughout performances (Greenham, Hewitt and Norton, 2017; Fernandez-Navarro, Fradua, Zubillaga and McRobert, 2018; Gollan, Ferrar and Norton, 2018). Albeit, in professional American football, notably little research has investigated the existence of game styles (Carter, James and Parmar, under review^a), with one example utilising pre-established play-book centred styles to distinguish potential for injury (Teramoto, Petron, Cross and Willick, 2015) including “West Coast” or “Air Coryell” offences, in addition to “3-4” or “4-3” defences formulated during

the 1960s and 1970s (Carroll, 1999). Notwithstanding, Carter, James and Parmar (under review^b) demonstrated the presence of playing styles in the NFL, in addition to identifying the contributions of specific performance indicators (PIs) within styles of play through the use of dimensionality reduction techniques. Moreover, the use of playing styles to identify dominant team behaviours has been proposed as a method for teams to contextualise their strategies in order to optimise the likelihood of successful outcomes (Fernandez-Navarro et al., 2016; Lago-Peñas, Gómez-Ruano and Yang, 2017; McLean et al., 2017). Nonetheless, while this may be the case, patterns in team behaviours have been revealed to be subject to the influence of situational variables, and as such, could hold an impact on match outcomes (Fernandez-Navarro et al., 2018).

In fact, in order to prevent the misinterpretation of findings, context must be considered to accommodate for limitations in isolated analyses of PIs (Hughes and Bartlett, 2002). Furthermore, the synthesis of contextual and performance indicators in preparation for upcoming opponents is recommended, as it allows teams to replicate the conditions under which a future performance may occur (Kormelink and Seeverens, 1999). Indeed, these situational variables may lead to a greater understanding of teams' strengths and weaknesses, and aid coaches in formulating responses to expose opposition weaknesses (Hughes and Bartlett, 2002; Lago, 2009). While the influence of situational variables has been analysed in several sports (Carling et al., 2005; Tucker, Mellalieu, James, & Taylor, 2005; Taylor et al., 2008; Lago, 2009; Csataljay, James, Hughes and Dancs, 2012; Trewin, Meylan, Varley and Cronin, 2017; Rico-González et al., 2022), their inclusion in studies investigating performance in the NFL has received little direct attention (Carter et al., under review^a). Albeit, certain situational variables such as quarter, down, yard line, game scenario (match status), time remaining, points differential and yards to go have been accounted for amongst research attempting to predict play types (Jordan, Melouk and Perry, 2009; Lee et al., 2016; Teich, Lutz and Kassarnig, 2016; Joash Fernandes et al., 2020). However, despite evidence that team and opposition quality influence key performance indicators (KPIs) which discriminating between winning and losing (Trninić, Dizdar and Lukšić, 2002; Drikos, Barzouka, Balasas and Sotiropoulos, 2022), scant exploration into their impact with professional American football has been attempted. More importantly, this dearth of analysis, alongside shortfalls in predictive accuracy hypothesised to stem from variation in styles of play (Joash Fernandes et al., 2020) may serve to develop a better understanding of team behaviours (Carter et al., under review^b).

To address the impact of differences in team and opposition quality, various means of have been used for classification, yet these approaches have been predominantly centred around qualifications of strong, medium and weak, quartile groups based on league position or end-of-season competition ranking (Taylor et al., 2008, Bradley et al., 2014; Yang et al., 2018, Drikos et al., 2022). However, in American football, teams can be seen as a combination of subsidiary groups - forming combinations of players according to offensive or defensive skillsets and responsibilities (Dutta, Yurko and Ventura, 2020), with further areas distinguished by styles of play (Carter et al., under review^b). Albeit, while rankings have been considered to account for specific areas of play (Lee et al., 2016; Goyal, 2020), differences in rankings may provide an inadequate representation of real differences between teams through an inability to holistically classify teams' performance. Subsequently, several investigations have identified a lack of sensitivity in many classification techniques - suggesting an exploration of methods which consider the current state of teams' relative quality at the time of a match (Taylor et al., 2008; Mao, Peng, Liu and Gómez, 2016), in addition to the use of objective techniques to categorise teams into apt groupings with clear differences (Drikos, Sotiropoulos, Barzouka and Papadopoulou, 2019; Drikos et al., 2022). Furthermore, achieving empirical categorisations of team quality may afford coaches and analysts an opportunity to comprehend actual differences between team quality profiles and devise adjustments to improve performance (Yang et al., 2018).

The potential outcome of these 'theories of successful play' (Trninić et al., 2002), wherein teams distinguished according to 'actual quality', particularly when conducted in accordance with performance in established styles of play (Sampaio, Drinkwater and Leite, 2010; Lopez-Valenciano et al., 2022; Carter et al., under review^b), may enhance actionable insight, as well as strategic development in a given matchup context (Trninić et al., 2002). Indeed, interactions between team and opposition strength have been found to influence the consistency of playing styles exhibited by teams (Sarmiento et al., 2014; Gollan et al., 2018).

4.1.2 Objectives

The aim of this study was to evaluate the impact of playing style proficiencies as a function of team quality on success in the regular season of the National football league by: (1) Using extant offensive and defensive styles of play ($n = 8$, Appendix B) to weight team performance indicators and objectively classify teams according to (dis)similarity within the underlying data and (2) to group teams by playing style proficiencies and determine their predictive power on total regular season wins.

4.2 Method

4.2.1 Sample

Match data recorded from play-by-play event data was collected from all regular season performances of the 2018 to 2022 seasons ($n = 5$) of the national football league (NFL). The sample criteria produced 1311 games, with one observation per team ($n = 2622$). The data originated from publicly available third-party repositories including nflfastR (Sebastian and Baldwin, 2024) and pro-football-reference (2024), derived from first-party official match summaries published by the NFL (2024). While the original play-by-play descriptions of on-field events, actions and context, scraping techniques were employed to generate performance and situational variables (Appendix D).

4.2.2 Procedure

Data collection was conducted through the use of text and ethical web-based scraping techniques. Action and situational variables retrieved by nflfastR were accessed through the nflreadr (Ho and Carl, 2024) package for R were aggregated with performance variables published on the public domain by pro-football-reference (2024) using the polite (Perepolkin, 2024) and rvest packages for R. The data was aggregated and prepared for statistical analysis using the RStudio Integrated Development Environment (IDE) for R (Posit team, 2023), powered by the statistical computing language R (version 4.3.2, R Core Team, 2023). The inclusion criterion for performance indicator

selection was undertaken in accordance with variables ($n = 41$) previously associated with playing styles ($n = 8$) in the NFL. Furthermore, prior to statistical analysis, each variable was summarised by a team averaged, then weighted (Appendix E) according to its respective playing style(s) and factor loading(s) - producing team playing style profiles ($n = 160$). The operational definitions were adopted in accordance the NFL official glossary of terms (Appendix D) as well as the pro-football-reference glossary of terms (pro-football-reference, 2024). Ethical approval for this study was granted by the London Sport Institute Research Ethics Committee (REC) on 24/11/2021 (Appendix I).

4.2.3 Statistical analysis

Clustering techniques generate several subsets or clusters comprised of closely related objects relative to those assigned to different clusters. This is achieved through partitioning according to the similarity or dissimilarity between data objects (Mondal and Choudhury, 2013). Moreover, such procedures allow for investigators to discover groups within data by empirical means rather than resorting to an arbitrary structure or design (Hodge and Petlichkoff, 2000). Indeed, cluster analysis affords the ability to distinguish subgroups which fit the data whilst maximising inter-cluster, and minimising intra-cluster differences (Hair, Anderson, Tatham and Black, 1995).

Clustering procedures are defined by partitioning or hierarchical groupings, however, in cases where no rationale serves to predict a hierarchical structure within the data, this technique becomes less suitable for analysis (Boslaugh, 2012). Therefore, non-hierarchical techniques such as k-means clustering provides an appropriate alternative which iteratively estimates cluster means from randomised centroids, assigns observations to clusters according to their proximity (Sankaran, 2014), and optimises cluster membership (Boslaugh, 2012). Albeit numerous methods serve to quantify the distance between an observation and potential cluster centroid, euclidian distancing has been found to outperform alternative distancing including Manhattan in efficacy or sum squared errors produced, and model efficiency (Thakare and Bagal, 2015). A principal requisite in clustering using non-hierarchical approaches is to define the number of clusters the algorithm should fit the data to - defined as cluster validity (Charrad, Ghazzali, Boiteau and Niknafs, 2014). To address the issue of cluster validity, numerous indices have been proposed to

quantify an optimal cut-off value (Appendix F) to partition the underlying data (Charrad et al., 2014). However, due to variation in algorithmic parameters leading to variation in cluster positioning, Charrad et al. (2014) developed a solution (NbClust: an R package for determining the relevant number of clusters in a data set) - enabling researchers to apply myriad cluster validity indices to a dataset and identify the optimal solution through comparison or draw conclusions by majority rule. Thus, empirically optimising cluster selection and reducing reliance on researcher interpretation in approaches used to identify visual cut-off points akin to the ‘elbow’ method.

Subsequently, in sport performance analysis, clustering performances through empirical or arbitrary means by match or profile may be interpreted as an indicator for the relative quality exhibited by a team, be it overall (Lago-Peñas and Lago-Ballesteros, 2011; Marcelino, Mesquita and Sampaio, 2011; Yang et al., 2018), according to characteristics (Sampaio, Drinkwater and Leite, 2010; Shelly et al., 2020), or within a certain playing style (Gómez et al., 2018; Gollan, Bellenger and Norton, 2020; Wedding et al., 2021). More importantly, observed variation in the cluster assignment within styles of play denoting proficiency may be evaluated alongside actual determinants of success and team quality such as wins. To analyse this effect, multiple linear regression was used to determine the contributions of game style proficiencies towards total wins per season - quantify the extent of linear relationships and predicting team success dependent on style proficiency cluster membership (Field, 2013).

Table 6. Number of clusters selected by majority rule (Charrad et al., 2014)

Style of play	Number of clusters (majority rule)	Proportion of indices (n = 27)
deep passing	2	52.85% (n = 14)
defensive perturbations	2	22.22% (n = 6)
passing efficacy	3	48.14% (n = 13)
quarterback suppression	3	40.74% (n = 11)
rushing	2	52.85% (n = 14)
scrambling	2	52.85% (n = 14)
tackling	2	29.62% (n = 8)
targeting and separation	2	48.14% (n = 13)

K-means clustering with euclidian distancing to determine the (dis)similarity of observations was conducted on data standardised using z-scores using the NbClust package for determining the relevant number of clusters in a data set (Charrad et al., 2014). The majority rule approach revealed the playing style resulted in clusters ranging from 2-3 per style of play. In determining the appropriate number of clusters to retain, few playing styles exhibited ‘indecision’ with regard to the majority rule, with the exception of defensive perturbations and tackling with only 22.22% ($n = 6$) and 29.62% ($n = 8$) of agreement between cluster validation indices ($n = 27$) respectively (Table 6). Whilst, in assessing the clusters of playing styles proficiencies, the data was analysed for homoscedasticity ($x^2 = 0.113$, $df = 1$, $p = 0.736$), multicollinearity (variance inflation factor (VIF) < 2.2 for all measures), and residual normality via a Shapiro-Wilk test ($W = 0.991$, $p = 0.399$) - meeting the requisite assumptions for performing multiple linear regression.

4.2.4 Overview of statistical methodologies

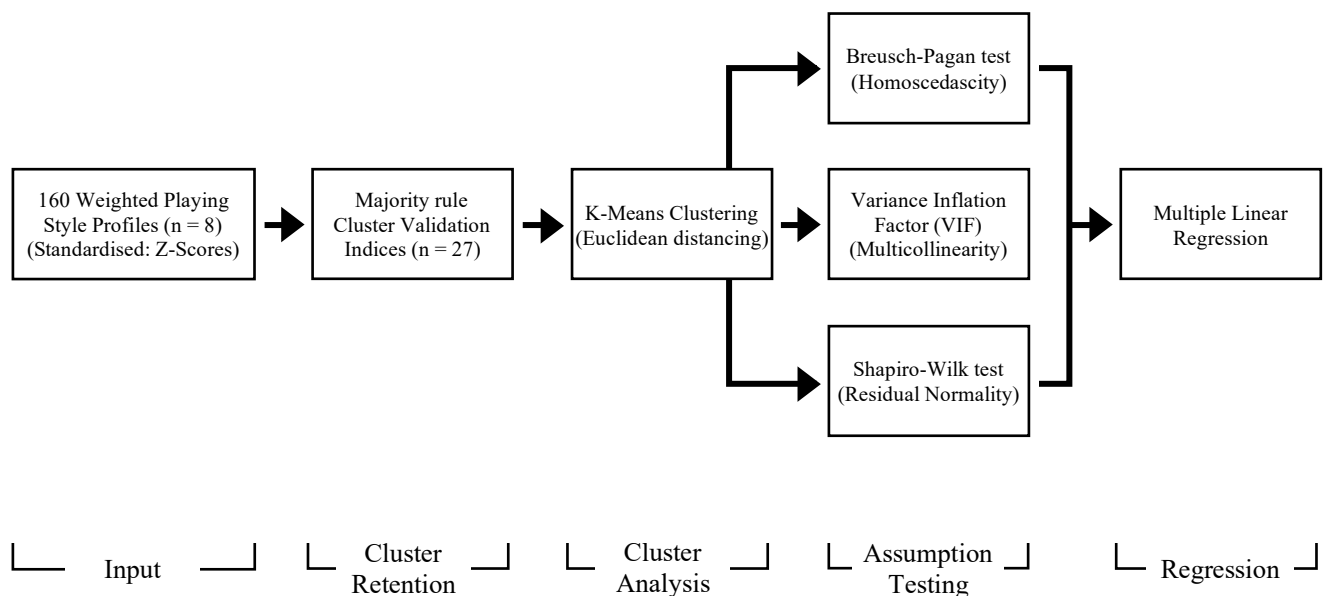


Figure 5. Overview of statistical methodologies used to cluster efficacy within styles of play as a function of team quality and assess their association with success in the NFL.

4.3 Results

Among the playing styles partitioned into two clusters ($n = 6$), three playing styles (deep passing, defensive perturbations and targeting and separation) demonstrated approximately even splits between the 160 team profiles, whereas within the remaining three styles of play (rushing, scrambling and tackling), a substantial difference in the proportion of teams exhibiting a high proficiency was notably reduced (Figure 6). Furthermore, amid game styles split around three cluster centroids, differences were apparent. This difference manifest in the percentage of teams depicted as middling in passing efficacy and quarterback suppression respectively, whilst teams clustered as middling in passing efficacy represented a greater proportion of the sample (50.6%), the quantity of teams middling in quarterback suppression accounted for 20%. Notwithstanding, the styles of play partitioned into three clusters displayed similarities in proportion of high and low proficiencies.

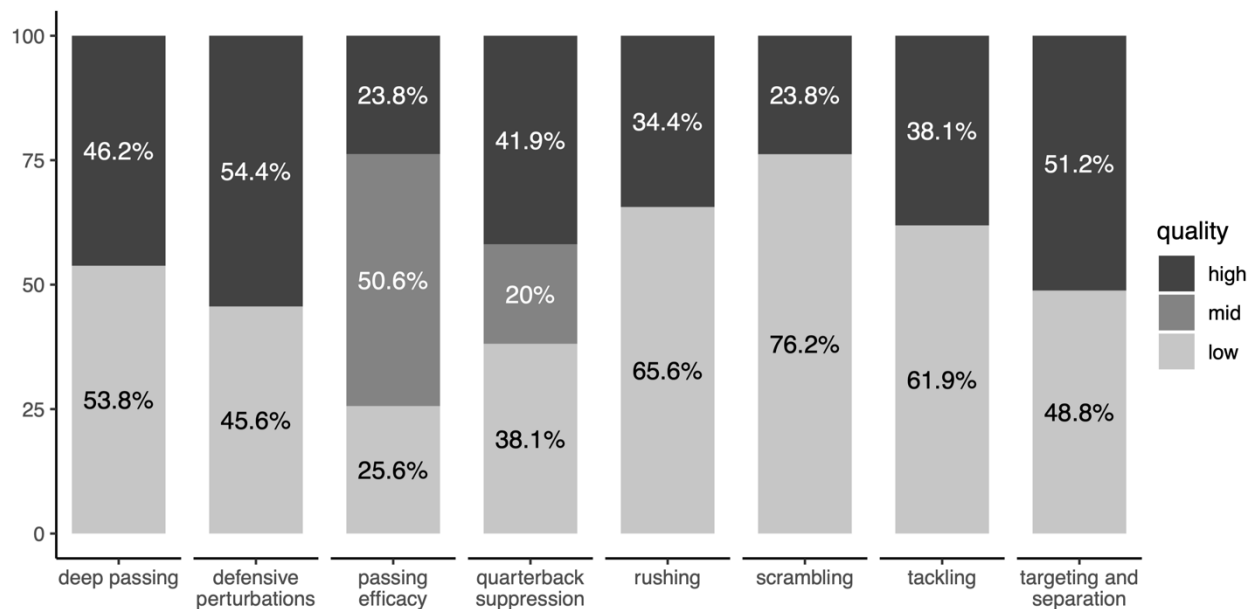


Figure 6. Playing style proficiency team distributions.

Figures 6 and 9 demonstrate the relative (dis)similarity and variation between teams in game style proficiency clusters. In 5 of the clustered styles of play, the cluster identified as exhibiting the highest quality varied more than to those belonging to lesser proficiency clusters. Interestingly, this effect did not apply to the remaining 3 styles of play, which were all defensive

in nature (defensive perturbations, quarterback suppression and tackling). Furthermore, the box plots illustrated that despite greater variation present in many of the clusters, outliers in higher team proficiencies among playing styles were more unlikely to occur than teams classified in middling or low clusters (Figure 7).

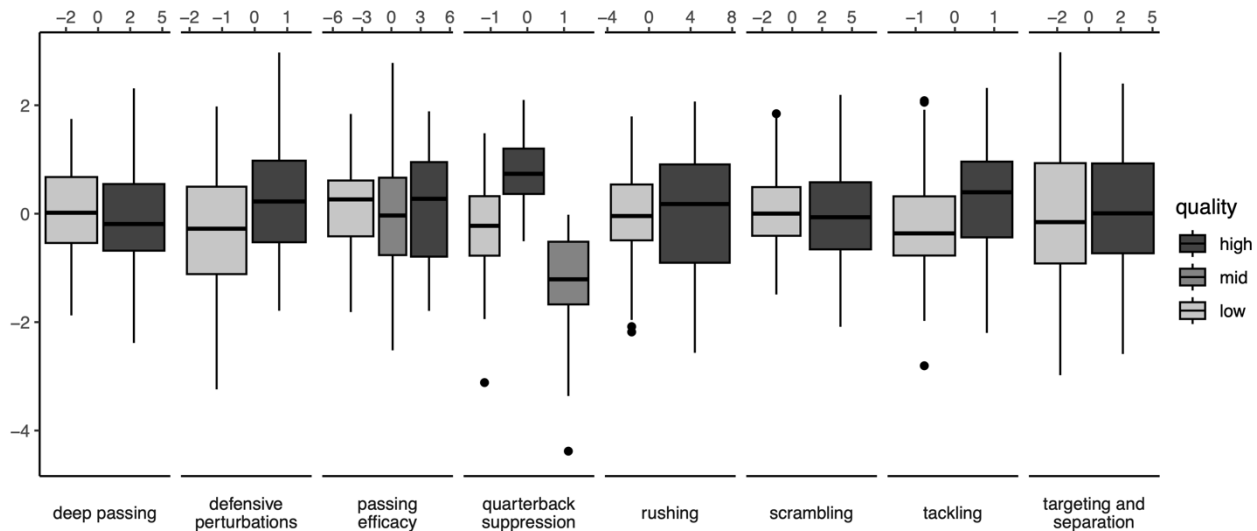


Figure 7. Boxplots of playing style proficiency classifications

The results of the playing style classifications were found to have a significant predictive power (Table 7) on season wins ($F(8, 151) = 18.37$, $R^2 = 0.493$, $p < .001$). In the passing game, passing efficacy, in addition to targeting and separation playing style proficiencies were the most significant positive predictors of season wins ($\beta = 1.75$, $t = 5.47$, $p < .0001$ and $\beta = 1.9$, $t = 3.62$, $p < .0001$ respectively), while differences in deep passing clusters presented no significant predictive power towards wins. Though to a lesser degree of significance (Figure 8), in pass defence, quarterback suppression proficiency was a positively significant predictor of success in the regular season ($\beta = 0.6$, $t = 2.86$, $p = 0.005$). Similarly, tackling proficiency showed a positive and significant, yet low estimate in predicting wins ($\beta = 0.92$, $t = 2.25$, $p = 0.026$). Conversely, defensive perturbations were identified as a highly significant negative predictor of wins ($\beta = -2.08$, $t = -5.44$, $p < .0001$). Meanwhile, running the ball offensively showed a disparity in the predictive power of a designed running style, or rushing proficiency and scrambling quality ($\beta = 1.4$, $t = 3.07$, $p = 0.003$ and $\beta = 0.82$, $t = 1.89$, $p = 0.06$ respectively), albeit, among statistically

insignificant styles of play in modelling season wins (Table 7), a difference in significance was apparent between scrambling ($p < .1$) and deep passing ($p = 0.97$).

Table 7. Results of multiple linear regression on season wins

	β	SE	t	2.5% CI	97.5% CI	p
intercept (constant)	-0.35	2.07	-0.17	-4.44	3.75	0.87
deep passing	0.02	0.51	0.04	0.04	0.97	0.97
defensive perturbations	-2.08	0.38	-5.44	-2.84	-1.33	<0.000
passing efficacy	1.75	0.32	5.47	1.12	2.38	<0.000
quarterback suppression	0.6	0.21	2.86	0.18	1.01	0.005
rushing	1.4	0.46	3.07	0.5	2.3	0.003
scrambling	0.82	0.43	1.89	-0.04	1.67	0.06
tackling	0.92	0.41	2.25	0.11	1.74	0.026
targeting and separation	1.9	0.52	3.62	0.86	2.93	<0.000

β = estimated regression coefficient; SE = standard error; t = test statistic; CI = confidence interval; degrees of significance (p): < .0001 = '***'; < 0.01 = '**'; < 0.05 = '*'.

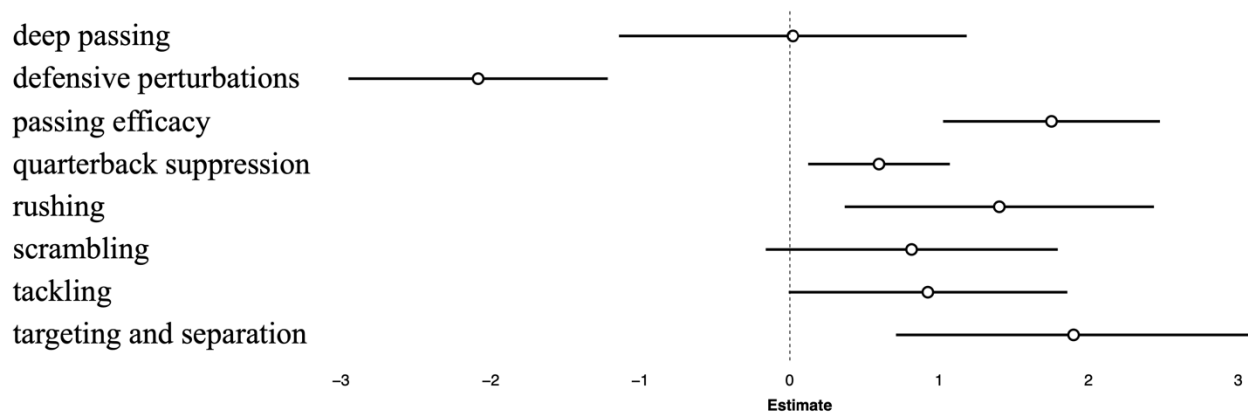


Figure 8. Multiple linear regression coefficient estimates (β) per playing style

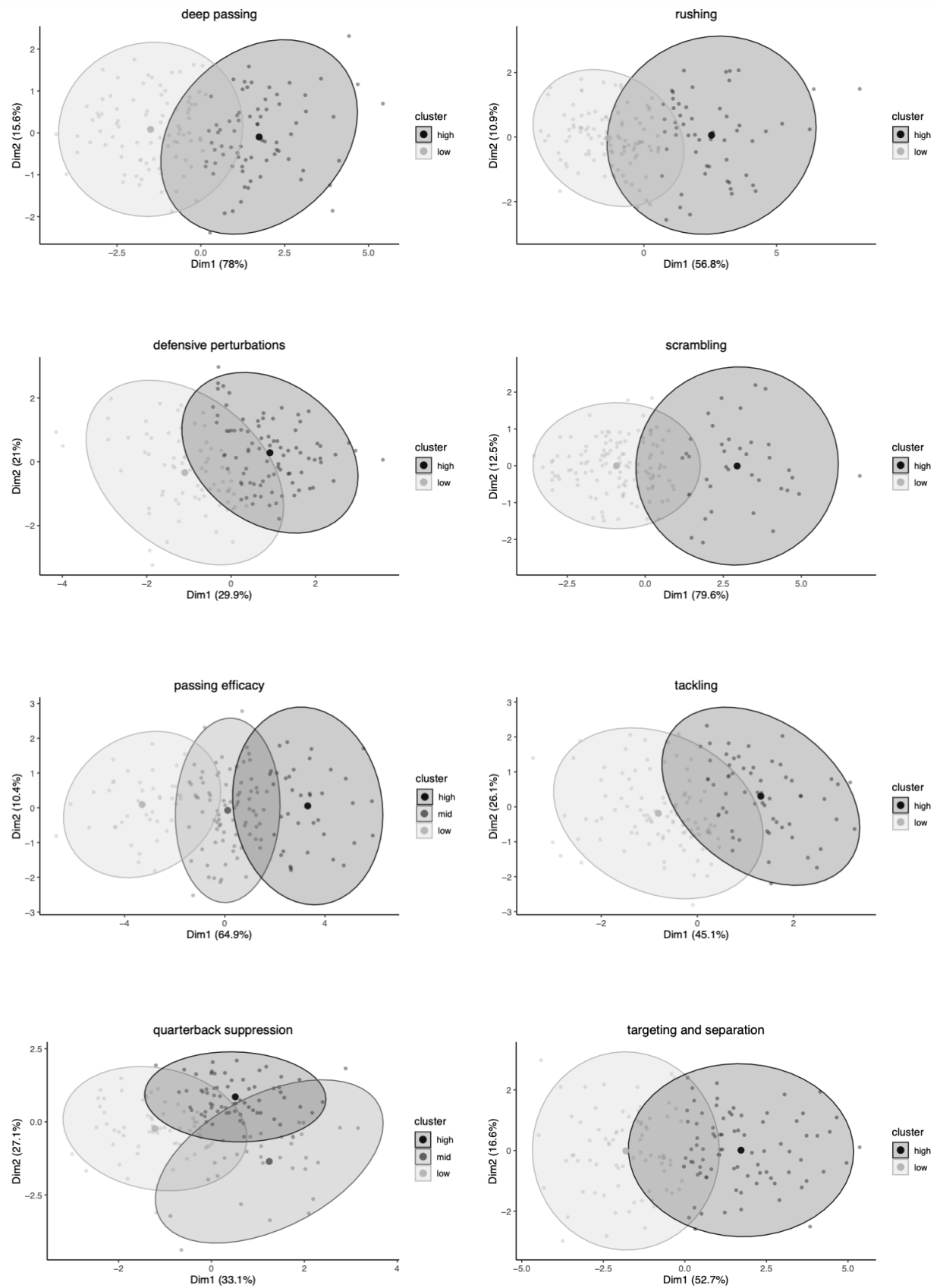


Figure 9. Teams clustered according to (dis)similarity in playing style proficiencies.

4.4 Discussion

The present study has identified a novel facet of team quality, allowing teams to be analysed according to their ability to impose particular styles of play. Previous studies investigating the interaction between team quality and game styles in association football have established that stronger teams exhibit longer possession-based attacking build-up patterns (Fernandez-Navarro et al., 2018), whereas weaker teams are more likely to operate under counterattacking or deep defensive styles of play (Lago, 2009; Bradley et al., 2014). Though teams may attempt to prolong possession time through the selection of running plays (Lee et al., 2016), American football game mechanics enforce a sustained a quasi-direct attacking style, with defensive responses limited to combination of preventative and perturbation-based measures (Joash Fernandes et al., 2020). However, in American football, the collective strategies implemented by players and teams can be measured over substantially shorter intervals (Carter et al., under review^a) - manifesting in passing, running or unintended scrambling attempts and reciprocal actions in defence (Goyal, 2020). Furthermore, as the majority of players exclusively play specific positions pertaining to offence or defence (Robbins, 2011), an overall team classification of quality may provide an insufficient account of performance in professional American football. Therefore, analysing team quality with regard to previously established styles of play enable a more detailed understanding of team strengths and weaknesses (Carter et al., under review^b) - allowing coaches to inform decision making with playing style proficiencies along with other situational variable to account for specific match context (González-Ródenas, Aranda and Aranda-Malaves, 2020).

Whereas certain studies have adopted categorisation techniques involving arbitrary dichotomous (i.e., strong or weak), trichotomous (i.e., high, middling, or low), and quartile partitions based on team rank at the end of, or during the season (Taylor et al., 2008, Bradley et al., 2014; Yang et al., 2018, Drikos et al., 2022), this study pursued and identified playing style-centred quantifications of team quality. Objective and adequately sensitive partitioning of teams have been hypothesised to produce a more appropriate solution for categorising teams (Yang et al., 2018; Drikos et al., 2019; Drikos et al., 2022), in addition to developing smaller and increasingly distinct groups (Yang et al., 2018). Indeed, to account for any potential inherent differences between teams in the execution of playing styles, the clustering conducted in this

investigation employed validation protocols through an exhaustive list of indices (Charrad et al., 2015) - meeting a key procedural criterion recommended for rigorous cluster analysis (Clatworthy et al., 2005). However, contrary to the proposed presence to smaller groupings among teams, results of the cluster validation and majority rule revealed that 75% ($n = 6$) of previously identified styles of play (Carter et al., under review^b; Appendix E) were most appropriately partitioned into a dichotomy of relatively stronger and weaker playing style quality. Notwithstanding, the two remaining game styles - interpreted as a both offensive and defensive representation of the passing game (passing efficacy and quarterback suppression) exhibited trichotomous partitions, albeit to varying degrees of clarity in (dis)similarity. In partitioning teams according to styles of play, the indices used to produce a majority rule for cluster validation demonstrated far greater indecision among defensive playing styles, with defensive perturbations (22.22%, $n = 6$), tackling (29.62%, $n = 8$) and quarterback suppression (40.74, $n = 11$) exhibiting a lesser majority for cluster validation relative to each of the offensive playing styles. It's possible to hypothesise that this effect may be caused by homogeneity in variance accounted for among the latent variables present within defensive styles of play (e.g., defensive sacks and tackles for loss constituting an equally important factor as interceptions and quarterback hurries) - rendering clustering as a function of proficiency more challenging relative to increasingly heterogeneous offensive counterparts.

While this may be the case, an interesting contrast in proportional representation of team proficiencies in passing efficacy and quarterback suppression. Despite embodying analogous areas of play differentiated by offence and defence, the greatest proportion of teams were classified as middling in passing efficacy (50.6%), whereas according to the present the interpretation of quarterback suppression proficiencies, teams partitioned as middling embodied the lowest proportion (20%) of sampled teams. Furthermore, among the styles of play where teams were found to most appropriately partition into two clusters ($n = 6$), three patterns in play, rushing, scrambling and tackling, displayed a substantially lower proportion of teams clustered with a high proficiency (34.4%; 23.8% and 38.1% respectively). Interestingly, the aforementioned styles of play may encapsulate the ability of an offensive player to evade their defensive opponent and vice versa. Though yards before and after contact have been identified as key contributors in rushing game styles (Carter et al., under review^b), certain investigations have suggested that many scouts do not to appropriately account for quarterback rushing ability (i.e., scrambling), despite having

been found to differentiate elite and very good quarterbacks in the NFL (Craig and Winchester, 2021). This confirmed findings by Yurko, Ventura and Horowitz (2019), describing mobility and scrambling as potentially overlooked, yet indicative of the ability to avoid sacks, thus tackles. The results of this study suggest that proficiency in evasiveness may extend to rushing, in addition to being reflected in tackling quality.

Efficacy in the performance of certain offensive styles of play in elite association football have been shown to be more important than defensive patterns of play (Lopez-Valenciano et al., 2022). Among the game style clusters sampled as a function of team quality, two offensive playing styles were identified as statistically insignificant predictors of season wins - resulting in an equal proportion of significant offensive and defensive predictors of success ($n = 6$). Moreover, despite constituting a key indicator of individual quarterback proficiency and differentiating teams according to 'elite' and 'very good' mobility and evasiveness, scrambling, provided a substantial, albeit statistically insignificant prediction of successful match outcomes ($p = 0.06$). Contrarily, teams' deep passing proficiencies ($p = 0.97$) demonstrated no predictive power. However, in accordance with themes observed by Lopez-Valenciano et al. (2022), high proficiency in the attack, with passing styles of play exhibiting the most significant positive predictors ($p < .001$) of successful matches outcomes per season (passing efficacy; targeting and separation). Moreover, results from multiple linear regression supported findings by May et al. (2010), whereby successful teams outperform their opponents in passing and rushing efficiency, though highlighting that efficiency rushing the ball was relatively less important ($p < 0.01$). Furthermore, while defensively the most significant predictor of wins was negative (defensive perturbations, $p < 0.001$), quarterback suppression ($p < .01$) and tackling ($p < .05$) were positive significant predictors of season wins. Thus, defensive approaches serving to prevent or react (quarterback suppression and tackling) to offensive opposition behaviours may lead to increased opportunities for successful season outcomes, as opposed to adopting pro-active (defensive perturbations) measures to pursue the ball, which potentially create space for teams to exploit in the passing game (Dutta, Yurko and Ventura, 2020).

To illustrate the association between team quality and the teams which produce the most successful match outcomes, though evident, Sampaio et al. (2010) highlighted that high quality

teams are most effective at scoring. More importantly, no differences were found between intermediate and weaker teams (Sampaio et al., 2010). In relation to performance in the NFL, efficacy in passing and pass defence have been distinguished as the most important factors which influence scoring (Cohea and Payton, 2011). Therefore, the findings of the present study indicating that higher proficiencies in passing and passing defence styles of play lead to a greater likelihood of success in elite American football. While an investigation analysing the impact of third down conversion rates on offensive performance determined that the metric is nearly meaningless (Cafarelli, Rigdon and Rigdon, 2012), third down success on passing and rushing attempts have been identified as marker variables within playing styles (Carter et al., under review^b) which have subsequently been found to be significant positive predictors of season wins. Notwithstanding, as described by Lopez-Valenciano et al. (2022), an understanding of playing styles associated with success allow coaches to quantify the importance of efficacy with the underlying constructs of a sport. Indeed, comprehension of these trends may allow coaches to adjust performance profiles according to team quality throughout a season (Sampaio et al., 2010).

4.4.1 Limitations and future research

Though the present findings highlight the impact of playing style proficiency, most notably the predictive power of team quality in pass offence and pass defence in regular season performances. This research design precluded data from the post-season (play-offs). Therefore, further analysis should be conducted to evaluate potential variation in playing styles, team proficiencies therein, and their potential impact in playoff environments. Moreover, the requisite number of season wins to qualify for the post-season varies between seasons due to the allocation of play-off positions per conference (i.e., AFC, NFC) and division (e.g., AFC North, NFC West, etc.) within each conference (Murray, 2018). Furthermore, conducting a disparate evaluation of the direct associations between offensive or defensive playing style proficiencies and points scored or conceded may produce a greater level of granularity for insights with regard to their practical implementations for coaches and analysts. Future investigations should consider the influence of proficiencies in styles of play alongside previously established contextual variables such as match location, in addition to in-game situational dynamics including match status, score differential, period, down, yards to go and timeouts remaining. Subsequently, the incorporation of playing style

proficiencies as a function of team quality to assess the interactive effects of team behaviours and match situations (Taylor et al., 2008) may elucidate current limitations in the analysis of NFL performance, particularly in a sport as strategic and coordinated as elite American football (Goyal, 2020).

4.5 Conclusion

To date, few studies have sought to evaluate playing style proficiency as a function of team quality in team invasion sports including professional American football. Using exploratory statistical methodologies, playing styles were objectively clustered as a function of team quality. This approach enabled teams to be partitioned according to inherent differences in performance proficiency present within the data rather than predetermined classifications. In modelling team quality among styles of play, only deep passing and scrambling were identified as insignificant predictors of regular season success in the NFL, while passing efficacy, as well as targeting and separation were the greatest positive and significant predictors of regular season wins. Notwithstanding, the game style interpreted as defensive perturbations produced a highly significant yet negative prediction of overall success. Though this area of play requires further investigation, in practical environments, playing style proficiencies may be utilised to quantify team and opposition strengths or weaknesses, thus contextualising and informing preparation for, and decision-making during performances.

5 The influence of playing style proficiencies on play calling tendencies in the National Football League (NFL).

Abstract

Unlike many team invasion sports, in the NFL, coaches are responsible for the coordination of almost every play. Hence, the ability to understand play-calling tendencies and predict opposition behaviours are considered key to performance outcomes. However, differences in predictability are suggested to originate from variation in style of play. Furthermore, teams' ability to execute certain playing styles has been shown to predict season success, yet their impact on play-calling has not been established. Therefore, through the use of Conditional inference trees (CTREEs), the aim of this study was to assess the influence of playing style proficiencies ($n = 8$) on play-calling predictability over the duration of 5 NFL seasons (2018-2022). The resultant models revealed that team proficiencies in passing efficacy, scrambling, rushing and deep passing among sampled situational variables ($n = 20$) improved play selection prediction accuracy ($>.1\%$). Relative to equivalent decision tree approaches (CART), CTREEs, with the contribution of playing style proficiencies demonstrated improved accuracy (3.35-5.2%). A comparison of models including and excluding scrambling eventualities highlighted differences in use-cases and practical implementations. Indeed, the visualisation, alongside statistical robustness of CTREEs enhance the interpretability of play prediction models and may inform decision making in the National football League by elucidating the likelihood of teams attempting to pass, run or scramble with the ball.

Keywords: American Football, NFL, Decision Tree, Machine Learning, Playing Style, Styles of Play, Team Quality, Play Calling, Conditional Inference Trees, CTREE

5.1 Introduction

5.1.1 Rationale

Elite-level sports competitions including the National football League (NFL) for American football have been described as highly complex environments with few everyday analogues (Reed, Critchfield and Martens, 2006). In turn, data collected to explain behaviour during matches has led to the use of numerous statistical methodologies to predict performance outcomes, in addition to the refinement of such techniques to accommodate an increasing demand for accuracy (Collins, Moir, Muresan, and Savage, 2017). In practice, the ability and attempts to predict events are said to constitute a key element of the game, with teams dedicating assets to identify team and opposition on-field tendencies (Lee, Chen and Lakshman, 2016). Indeed, understanding the interaction between play-calling and strategy is considered integral to American football (Teich, Lutz and Kassarnig, 2016). Notably, in sports environments, practitioners are tasked with identifying the necessity for, and implementing tactical adjustments according to team and opposition strengths and weaknesses (Hughes and Bartlett, 2002). More importantly, in American football, coaches are responsible for making on-field play-calling decisions (Teich et al, 2016).

Lee et al. (2016) contend that one of the most important areas analysed among NFL teams is predicting the likelihood of a team running or passing the ball according to specific situational parameters. However, unlike certain team invasion sports, tendencies may be reflective of the game mechanics introduced by the rules of American football (Section 9, Articles 1-3, NFL, 2024), wherein at the start of each possession, teams receive the ball with four attempts (downs) to advance to ‘the line to gain’ (10 yards) and receive another four downs, score, or turn-over possession. Consequently, given situations where teams have few attempts remaining to reach ‘the line to gain’, or need to overcome a challenging number of yards-to-go, tendencies towards passing the ball may predominate (Reed, Critchfield and Martens, 2006). Notwithstanding, under a ‘two-alternative operant choice’ framework, Stilling and Critchfield (2010) compared offensive decision making to a binary selection of play type between passing or running (alternatively rushing) the ball. However, with the application of generalised matching laws producing highly varying sensitivity to situational variables in play selection (Reed et al., 2006; Stilling and

Critchfield, 2010; Critchfield, Meeks and Stilling, 2014), alternative approaches have been undertaken by Lee et al. (2016), Goyal (2020) and Joash Fernandes et al. (2020), adopting several forms of machine learning (ML) techniques to compare and improve play selection prediction accuracy. Furthermore, notable efforts to utilise decision trees have been made to enhance the interpretability and intuitiveness, thus the applicability of play prediction models has provided a platform for refining current understanding of the influence of match context in decision making or play-calling in professional American football (Joash Fernandes et al., 2020).

Nonetheless, while Reed et al. (2006) acknowledged that most collated metrics record the outcome of plays resulting in a classification of pass or rush, the original intent of a play-call is not always encapsulated within collected data. For example, in modelling the sensitivity of plays called in a given context, instances where disparities assigned to player improvisation and player-coach influences led to a poor fit of the data. The case (Michael Vick) highlighted by Reed et al. (2006) demonstrated the potential impact of quarterback mobility or ability to scramble - subsequently presented as a distinguishing factor for identifying elite quarterbacks in the NFL (Craig and Winchester, 2021). Notwithstanding, the potential influence of mobile quarterbacks within attempts to predict play-calls has been omitted or transformed within sampled data to increase overall prediction accuracies by more than 1% (Lee et al., 2016; Joash Fernandes et al., 2020). Despite scant inclusion of scrambling measures in previous research (Yurko, Ventura and Horowitz, 2019; Carter, James and Parmar, under review^{a,b}), such aspects of performance may be extended to limitations reported by Joash Fernandes et al. (2020), emphasising the necessity to explore why some teams exhibit higher or lower predictability relative to others. Further hypothesising that coaches may demonstrate variability in play-calling consistency, which may manifest in ‘especially volatile’ situations such as plays occurring on third downs (Reed et al., 2006). Therefore, aside from high frequency in coaching staff changes resulting in lower prediction accuracies (Reed et al., 2006), differences in offensive styles reflected in play-calling patterns that become apparent over time have been proposed to partially account for up to 17.8% overall variation in predictability between teams (Reed et al., 2006; Joash Fernandes et al., 2020; Ötting, 2021).

In team invasion sports, styles of play are considered as a manifestation of tactics and strategy implemented over the duration of a performance or season, generating team-specific patterns in behaviour (Greenham, Hewitt and Norton, 2017; Fernandez-Navarro, Fradua, Zubillaga and McRobert, 2018; Gollan, Ferrar and Norton, 2018). More importantly, these playing styles can be used to outline team and opposition strengths or weaknesses by coaches in preparation for upcoming performances with the aim of manipulating the likelihood of successful outcomes (Fernandez-Navarro et al., 2016; Lago-Peñas et al., 2017; McLean et al., 2017). Within the context of NFL analytics, several playing styles have been identified through the use of dimensionality reduction techniques - forming latent variables representing areas of play such as passing, rushing, scrambling, in addition to distinctive defensive patterns (Carter, James and Parmar, under review^b). Moreover, established styles of play have been adopted to evaluate differences in team proficiency among playing styles, with the resulting classifications found to hold significantly positive or negative associations with season success (Carter, James and Parmar, under review^c). Furthermore, according to Gollan et al. (2020), additional research investigating styles of play may inform the predictability opposition behaviours within realistic in-game scenarios. However, to discern the mechanisms influencing the adaptation of performance patterns and enhance the current understanding of team interaction dynamics, playing styles must be studied under diverse situations (e.g., yard line, down or yards-to-go), alongside general sport contexts such as match location, scoreline, time remaining (Lago-Peñas and Gómez-López, 2014).

5.1.2 Model Selection

To derive probabilities concerning the likelihood of a given play type being called, a variety of statistical and machine learning procedures have been adopted (Lee et al., 2016; Goyal, 2020; Joash Fernandes et al., 2020) - resulting in a 10.4% range of predictive accuracies (65.3 to 75.7%). These approaches include random forests, gradient boost machine, k-nearest neighbours and neural networks - known as black-box systems due to their description as ‘opaque decision systems’ (Guidotti et al., 2018). However, despite embodying sophisticated, sensitive and easy-to-implement machine learning models to predict information (Goyal, 2020), further efforts are currently required to describe how such methods function and to explain decisions made within models (Guidotti et al., 2018). Additionally, the challenge to comprehend the influences forming

black-box predictions has been reported as potentially inappropriate for application in the NFL, as the relative importance of the features contributing towards a prediction itself may be more informative in practice than the accuracy of a model (Goyal, 2020; Joash Fernandes et al., 2020). Conversely, the application of more simplistic approaches like logistic regression and linear discriminant analysis have been suggested as more understandable, yet struggle in performance, namely in providing an accurate weighting for covariates, while being more susceptible to overfitting relative to ensemble tree procedures (Lee et al., 2020; Goyal, 2020).

In prioritising interpretability, coherence and trustworthiness for coaches, decision trees have been proposed to provide an acceptable combination of predictive accuracy and simplicity (Joash Fernandes et al., 2020). In addition to providing a clear visual representation of probabilities with 'real-game' applications, these techniques may allow for play-callers to assess the role of specific situational factors on the likelihood of a team passing or running the ball (Goyal, 2020). Nonetheless, widely practiced decision trees such as CART (Classification And Regression Trees) can be limited in accuracy due to overfitting data stemming from insufficient sample observations after splitting trees into myriad branches, thus leading to poor generalisability of the model (Goyal, 2020; Joash Fernandes et al., 2020). Moreover, many decision tree models may suffer from selection bias due to variable class, homogeneity and missing values potentially hindering the interpretability of tree-structures (Hothorn, Hornik and Zeileis, 2006). Hence, the conditional inference (CTREE) procedure was developed by Hothorn et al. (2006) and was found to match the predictive performance of exhaustive black-box search procedures by recursively partitioning data using stopping criteria based upon the most statistically significant differences among covariates. Indeed, Hothorn et al. (2006) further suggested that the implementation of conditional inference trees result in models which are easy to communicate with practitioners, while simultaneously being well-suited for the explanation and prediction of outcomes. Therefore, it may be hypothesised that CTREEs, in the context of NFL play prediction, may provide an enhanced approach to contextualise decision making by coaches when faced with any on-field situation.

5.1.3 Variable selection

In order to inform decisions in both offensive and defensive situations, coordinating coaching staff retain a level of communication with players (e.g., quarterbacks and middle line-backers) to make pre-snap adjustments for up to 15 seconds before the play clock (40 seconds) is exhausted (NFL, 2008). Thus, the input variables for modelling play types should be reflective of the information available during periods when contextual insights remain actionable (>15 seconds remaining on the play-clock).

A mixture of variables have been considered among previous research designs - defined as either static or dynamic. For example, Joash Fernandes et al. (2020) recommend to limit the inclusion of dynamic in-game metrics such as previous pass/run proportion and efficacy (Critchfield, Meeks and Stilling, 2014), as well as offensive and defensive players per position group (Lee et al., 2016) included in previous models due to the impracticality for coaches to perpetually update measurements and call plays during a short interval. This may be extended to the use of dynamic in-game probabilities, for instance, no-score, touchdown, field goal and win probabilities considered by Goyal (2020), which may not be available during performances. Nonetheless, despite a static variable criterion, metrics accounting for the formation or lineup positioning, potentially occurring below the 15 second threshold are commonly implemented (Joash Fernandes et al., 2020). Furthermore, whether the offence was in 'shotgun' formation was highlighted as one of the most influential contextual indicators in determining the likelihood of a pass or run play (Lee et al., 2016; Teich et al., 2016; Joash Fernandes et al., 2020; Ötting, 2021). Meanwhile, Ötting (2021) highlighted a clear and inherent bias indicating that play types are highly skewed towards passing outcomes on shotgun plays (1:2 to 1:3 run to pass ratio depending on down) which may potentially artificially augment prediction accuracies or confound models and render the inclusion of certain formations as static variables problematic among models.

Attempts to contextualise models with team quality have appropriated first (NFL, 2024) and third-party (Pro Football Focus, 2024) measures of team strength, including video game ranking of position groups and teams for overall and distinct areas of performance (i.e., passing and running offence or defence) without consideration of the players on the field in specific

situations (Lee et al., 2016; Goyal, 2020). Notwithstanding the potential validity or accuracy of these measures, the definition or formulation of these rankings are often omitted or disregarded (Lee et al., 2016; Goyal, 2020). Therefore, the use of previously established playing style proficiencies (Carter, James and Parmar, under review^b) as a measure of relative team quality in empirically derived features of NFL performance (Carter, James and Parmar, under review^a) may provide a fitting, transparent and static contextual factor in quantifying play-calling tendencies in elite American football.

5.1.4 Objectives

Due to ease of interpretation and intuitiveness, decision trees have been suggested as an appropriate technique to producing a tangible and practical account of play-calling tendencies (Goyal, 2020). As such, they may provide an optimal framework under which to analyse the influence of situational variables on play-calling tendencies in the National football league. Accordingly, the aim of this study was to apply previously established styles of play and teams' classifications therein as a function of team quality in order to (1) quantify the impact of team, in addition to opposition defensive playing style proficiencies on the predictability of play type selection, as well as their relative influence among other situational variables; (2) determine the variation in accuracy between models when including (Pass, run or scramble (PRS)) or excluding (Pass or run (PR)) the potential for scrambling outcomes and (3) evaluate the performance of conditional inference trees for use as an analytical tool for practitioners in American football.

5.2 Method

5.2.1 Sample

All play-by-play event data excluding special teams plays (i.e., punt, field-goal and extra point attempts) from regular season matchups ($n = 1311$) of the 2018 to 2022 national football league (NFL) seasons ($n = 5$) was collected for analysis, resulting in a total of 163882 recoded situational observations. More importantly, for machine learning purposes the data was further reduced into training (PRS: 114719; PR: 111736) and testing (PRS: 49163; PR: 47885) subsets of plays. The

data was sampled from the publicly available third-party repositories nflfastR (Sebastian and Baldwin, 2024) and pro-football-reference (2024), disseminating data distributed in official first-party match summaries by the NFL (2024). To achieve data collection for this study, a combination of web scraping and data manipulation techniques were performed to derive situational variables (Appendix G).

5.2.2 Procedure

Ethical text and web scraping techniques were conducted to collect situational and event data published on the public domain by the National football league (2024). The NFL official glossary of terms (2024) was consulted for the evaluation and inclusion of variables (Appendix G). Situational ($n = 20$) and play type ($n = 1$) variables (Appendix G) were collated from the nflfastR database (Sebastian and Baldwin, 2024), accessible via the nflreadr (Ho and Carl, 2024) package for R, while team playing styles ($n = 8$, Appendix F) and proficiencies were derived from previous research using a combination principal component analysis and cluster analysis (Carter, James and Parmar, under review^{a,b}). All data aggregation, manipulation and statistical analyses of the pass, rush or scramble (PRS) and pass or rush (PR) datasets as well as the resulting models were performed using the RStudio Integrated Development Environment (IDE) for R (Posit team, 2023), powered by the statistical computing language R (version 4.3.2, R Core Team, 2023). Ethical approval for this study was granted by the London Sport Institute Research Ethics Committee (REC) on 24/11/2021 (Appendix I).

5.2.3 Statistical analysis

With the aim of predicting play type (pass, run or scramble) given a number of situational measures ($n = 20$, Appendix G) in a systematic manner, a supervised machine learning approach was adopted (Joash Fernandes et al., 2020; Lee et al., 2016; Teich, Lutz and Kassarnig, 2016). Simply, such techniques task an algorithm with inferential learning from independent variables (input) within a training partition of a dataset, to predict dependent (output) outcomes (Louppe, 2014). Decision trees are used as a means of prediction and classification of outcomes under the form of recursive binary partitioning (Hothorn, Hornik and Zeileis, 2006), producing interpretable and intuitive

findings represented in a tree-like structure - enabling and facilitating the dissemination of results (De Ville, 2013). Specifically, conditional inference trees (CTREEs) were utilised in this study to ensure an unbiased, statistical significance-based selection of splitting covariates comprised of differing scales to produce optimally pruned decision trees which efficiently avoid overfitting classification tasks (Hothorn, Hornik & Zeileis, 2006). Furthermore, to increase the predictive capacity of the training data on the testing partition, 10-fold repeated ($n = 3$) cross-validation (CV) was conducted (Yadav and Shukla, 2016). The sampled data for this study was split into a training dataset for consisting of 70% of the analysed plays, while the remaining 30% of plays were designated for testing model prediction accuracy. Moreover, testing accuracy was additionally analysed for precision, as a measure of correct class prediction (accounting for false positives) and sensitivity, quantifying predictability over all instances of a class (underscoring the prevalence of false negatives) – exemplified in the prediction confusion matrix (Figure 10). All predictive models were generated through the ‘Caret’ package for building predictive models (Kuhn, M., 2008) and ‘partykit’ modular toolkit for recursive partitioning in R (Hothorn and Zeileis, 2015), while the resultant CTREEs were plotted using the ‘ggparty’ package for ggplot visualisations for ‘partykit’ (Borkovec and Madin, 2019).

	Actual Positive	Actual Negative	
Positive Prediction	TP	FP	Precision
Negative Prediction	FN	TN	
	Sensitivity		

Figure 10. Formulation of accuracy, sensitivity, and precision (Monaghan et al., 2021).

Where FN =False Negative; FP =False Positive; TN =True Negative; TP =True Positive;
 $Accuracy = (TP + TN) / (TP + TN + FP + FN)$; $Sensitivity = TP / (TP + FN)$; $Precision = TP / (TP + FP)$.

5.3 Results

5.3.1 Model accuracies

5.3.1.1 Overall model accuracy

The variation in accuracy between models attempting to predict play type with (PRS) and without (PR) the inclusion of scrambling plays (Table 8) varied by approximately 2% (68.65%, $p < .0001$ and 70.5%, $p < .0001$ respectively). Commensurate differences in accuracy were noted in pass/run and pass/run/scramble CTREES modelled according to specific downs and quarters when plays were executed (Figures 11 and 12). Furthermore, prediction accuracy was found to be substantially

Table 8. Conditional inference tree accuracy with(out) scrambling outcomes.

Model	Context	Accuracy %	95% CI	p-value
Pass, Run or Scramble (PRS)	Overall	68.65	(68.23, 69.06)	<.0001
	1 st down	61.2	(60.55, 61.58)	<.0001
	2 nd down	68.12	(67.4, 68.83)	<.0001
	3 rd down	83.82	(83.08, 84.54)	<.0001
	4 th down	80.14	(77.52, 82.58)	<.0001
	1 st quarter	65	(64.1, 65.88)	<.0001
	2 nd quarter	68.37	(67.57, 69.16)	<.0001
	3 rd quarter	66.05	(65.16, 66.92)	<.0001
	4 th quarter	74.73	(73.97, 75.47)	<.0001
	5 th quarter	63.74	(58.4, 68.85)	0.098
Pass or Run (PR)	Overall	70.5	(70.09, 70.91)	<.0001
	1 st down	63.23	(62.58, 63.88)	<.0001
	2 nd down	69.27	(68.55, 69.99)	<.0001
	3 rd down	87.54	(86.86, 88.2)	<.0001
	4 th down	82.74	(80.21, 85.08)	<.0001
	1 st quarter	66.35	(65.45, 67.24)	<.0001
	2 nd quarter	70.74	(69.95, 71.53)	<.0001
	3 rd quarter	67.51	(66.62, 68.38)	<.0001
	4 th quarter	77.06	(76.32, 77.79)	<.0001
	5 th quarter	66.07	(60.74, 71.12)	<.0001

CI: Confidence interval; 5th quarter: Overtime.

higher during plays occurring on 3rd and 4th downs (PRS: 83.82%, $p < .0001$ and 80.14%, $p < .0001$; PR: 87.54%, $p < .0001$ and 82.74%, $p < .0001$), whilst a similar increase in predictability was observed in the final quarters of each half of play - the 2nd (PRS: 68.37%, $p < .0001$; PR: 70.74, $p < .0001$) and 4th quarters (PRS: 74.73%, $p < .0001$; PR: 77.06, $p < .0001$). The inclusion of shotgun in PRS and PR models increased prediction accuracies by approximately 3% (PRS: 71.59%, $p < .0001$; PR: 73.55, $p < .0001$).

5.3.1.2 Play-calling predictability

The PRS model exhibited no sensitivity, nor precision in the prediction of scramble attempts. Hence predictability in play-calling tendencies were observed under the more accurate PR model (Table 8). The prediction of rushing attempts was least precise on 1st downs (59.11%), while less susceptible to false positives on 3rd and 4th downs, demonstrating 74.52% and 71.75% precision respectively (Figure 11). Elevated sensitivity for rushing plays on 1st (84.29%) and 4th downs

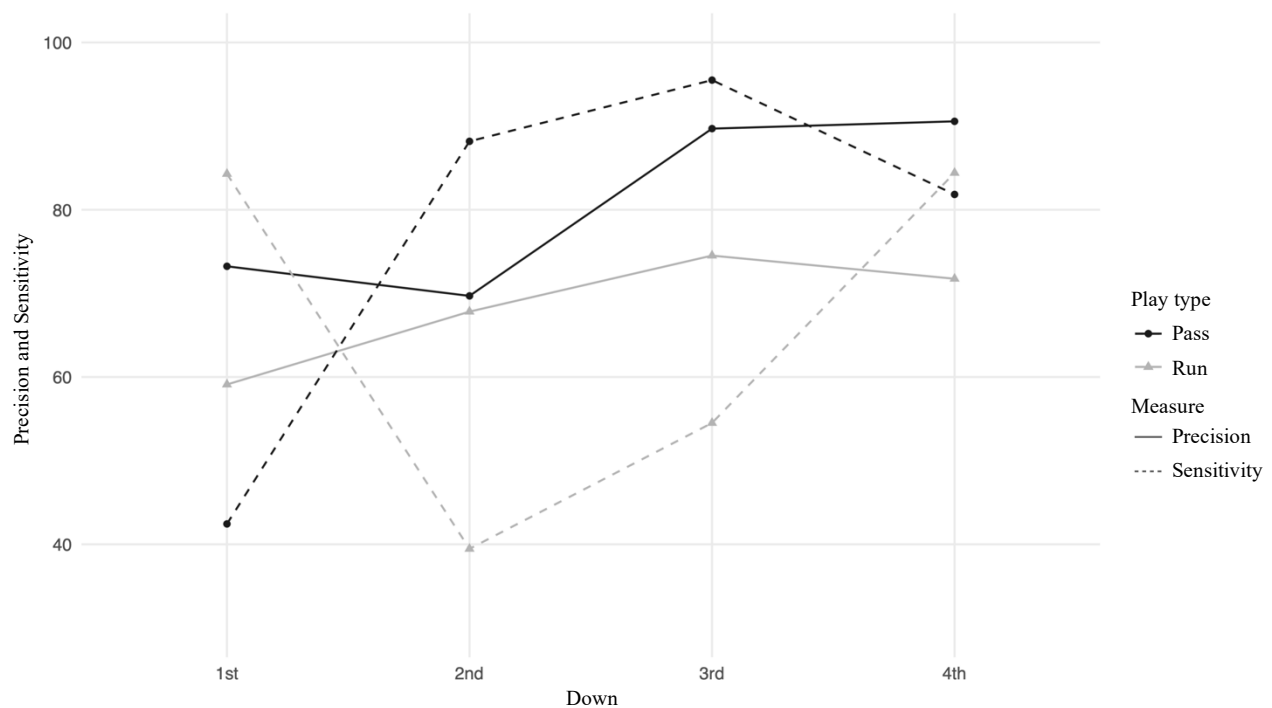


Figure 11. PR precision and sensitivity per down.

(84.41%) was demonstrated, yet 2nd (39.46%) and 3rd (54.51%) were exhibited notably high rates of false negative predictions. However, 1st, 2nd and 3rd downs appear to present an inverse relationship in precision and sensitivity - indicative of play calling tendencies in rushing and passing play selection (Figure 11). This effect was illustrated by the subsequent 73.24% precision in passing 1st down predictions, while a particularly low sensitivity (42.44%). Notwithstanding, passing plays were most precisely predicted on 3rd and 4th downs (89.7% and 90.57% respectively). Furthermore, pass attempts in latter down situations were also less susceptible to false negative misclassifications with high 3rd (95.5%) and 4th down (81.83%) sensitivity.

Play type predictability modelled according to quarter in play further indicated an inverse relationship in sensitivity between pass and run (Figure 12). While predictions of run attempts in the 1st (71.35%) and 3rd quarters (67.59%) resulted in fewer false negatives, a similar effect was observed for pass predictions in the 2nd and 4th quarters. Interestingly, similar trends were also reflected in the precision of pass predictability in the last quarters of each half. Contrarily, run play predictions were increasingly precise from the 1st quarter to the end of regulation time (4th quarter).

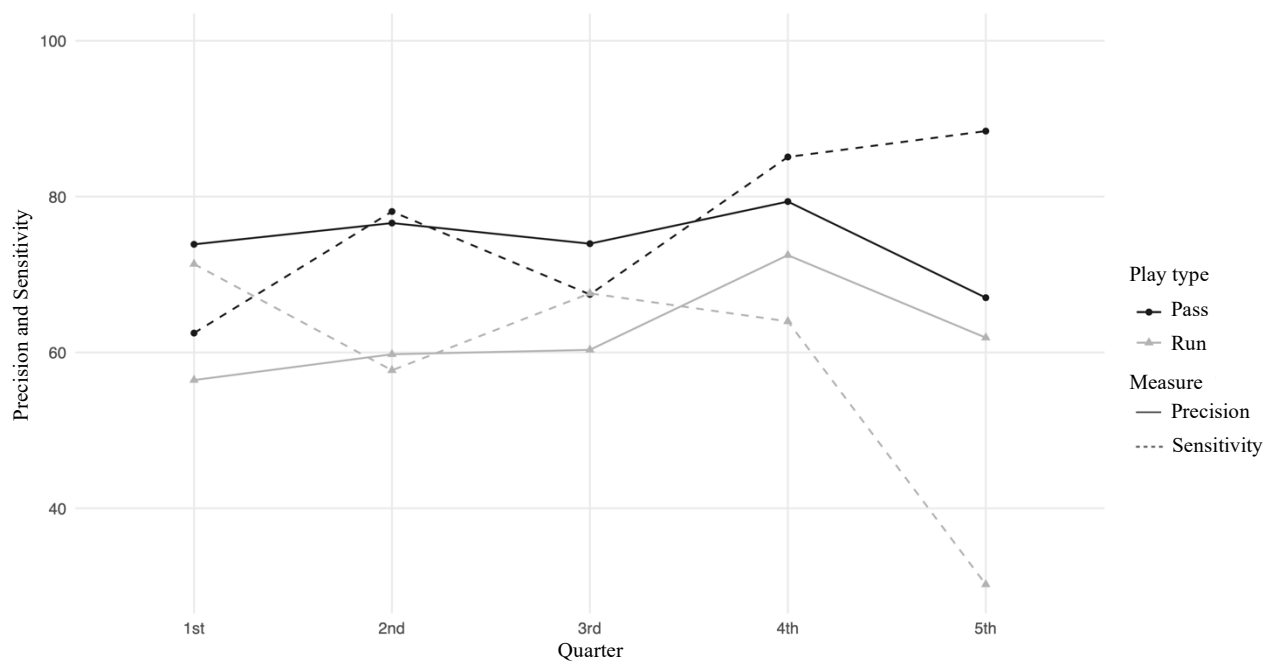


Figure 12. PR precision and sensitivity per quarter.

5.3.2 Variable importance

The relative variable importance (Figure 13) of situational covariates was represented by the average decrease in model accuracy when each feature is omitted (Strobl, Boulesteix, Zeileis and Hothorn, 2007). The situational variables exhibiting the greatest importance were down (20%), yards to go (11%), score differential (8%), half time remaining (5%) and quarter (4%). Though demonstrating lesser importance in refining the probabilities of play type outcomes, team proficiencies in passing efficacy (0.8%) and scrambling (0.7%) were approximately important as timeouts remaining (1%) and yard line (0.9%), while rushing proficiency (0.4%) and deep passing (0.2%) ability were found to split branches of the CTREE. Conversely, several features ($n = 7$) were presented negligible variable importance ($<0.1\%$) including the match location, divisional game and weather conditions, in addition to offensive targeting and separation and defensive quarterback suppression, defensive perturbations and tackling playing style proficiencies.

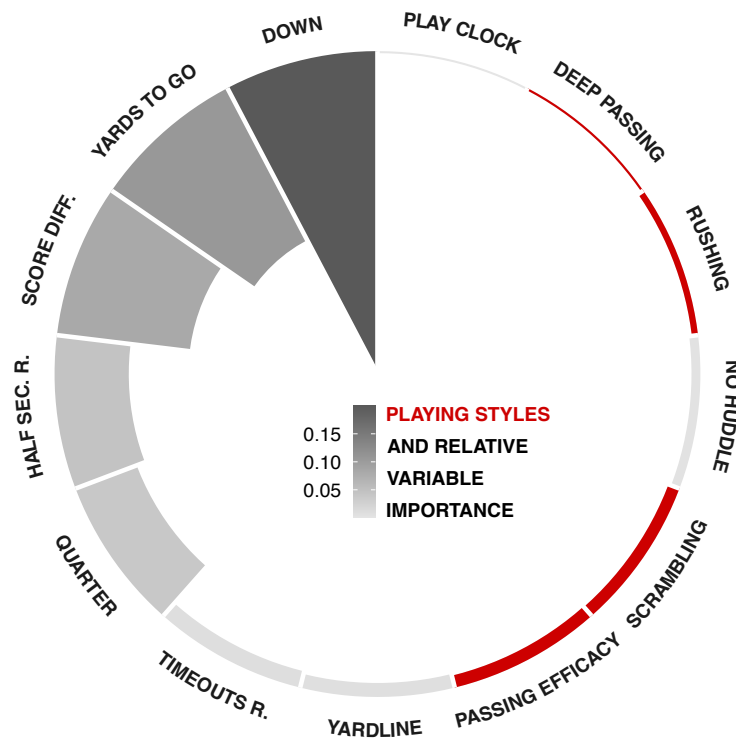


Figure 13. Variable importance in the PRS CTREE ($>0.1\%$).

5.3.3 Case study

Figure 14 illustrates the probabilities of play type outcomes given many situations. More specifically 3rd down scenarios, when the possession team had >1 yard-to-go to attain the line of gain, within one converted touchdown of their opponent (≤ 7 score differential) - the three most important variables for the prediction of play types (Figure 11). While not all significant splitting rules and variables were represented in order to provide a simplified depiction (Appendix H) of the PRS CTREE ($p < 0.05$, minimum split > 2000 and minimum terminal > 2000 plays), scrambling, rushing and deep passing playing style proficiencies were demonstrated as influential covariates in determining the probability of play type outcomes. Whereas a further difference in yards to go (≤ 2 , > 2) presented the most significant split with plays from ≤ 2 yards from the line of

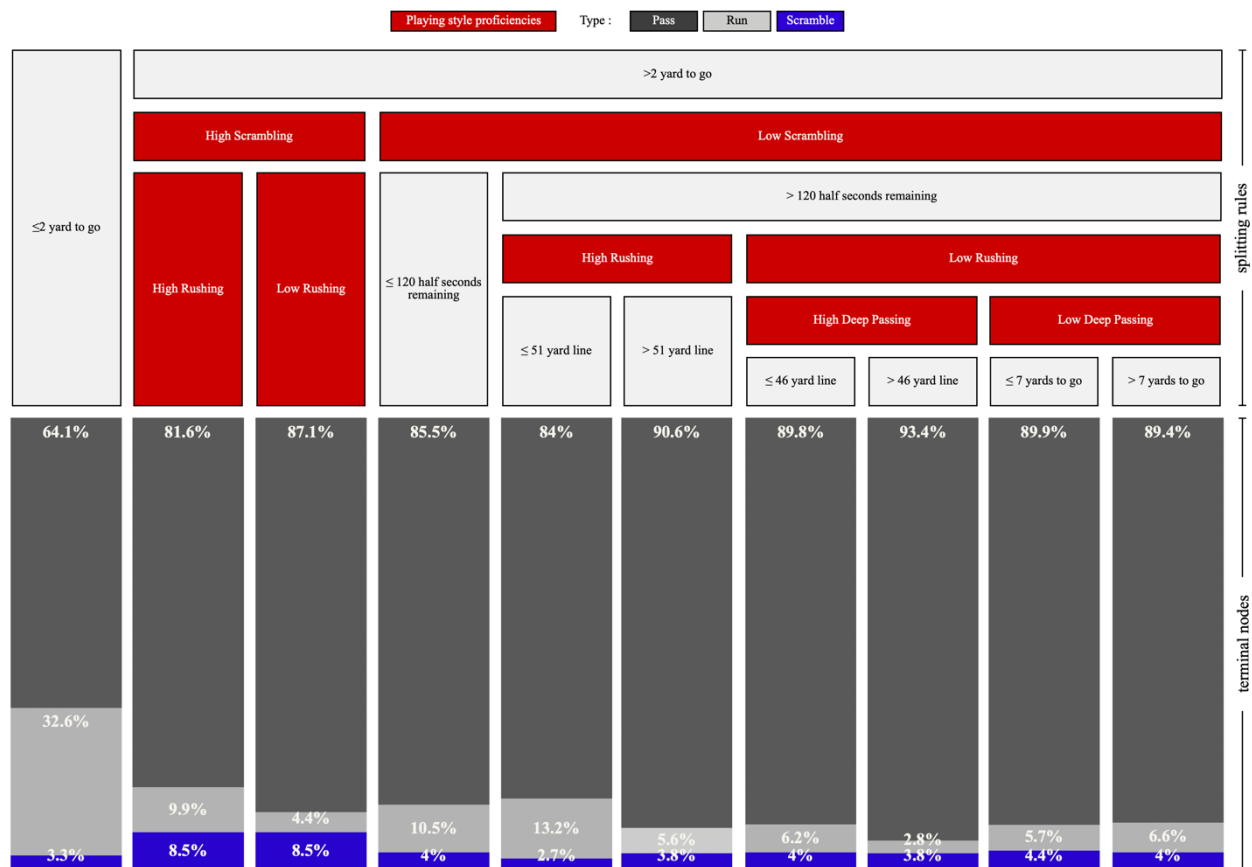


Figure 14. PRS CTREE probabilities of play types

On 3rd downs with more than 1 yard to go within a 7-point score differential: p -value < 0.01 differences in all splitting rules; terminal nodes: $n > 2000$ plays.

gain displaying a far greater probability to result in a running outcome (32.6%) than any similar context with >2 yards to go (mean = 7.21%). However, in situations with >2 yards to go the most significant subsequent split in the data was teams' scrambling proficiency (high and low), exhibiting differences in average scramble likelihood 4.69% greater for more proficient scrambling teams (8.5% $>$ 3.81%). Notwithstanding further splits in the data among proficient scrambling teams indicating that rushing team quality influences the probability of teams running the ball by 5.5%. For low scrambling proficiency teams, the greatest differences revolved around 2-minute warnings (half seconds remaining $>$ or $\leq$ 120 seconds), wherein teams were 4.02% less likely to attempt passing plays (89.52% $>$ 85.5%). Prior to 2-minute warnings, teams exhibiting low scrambling and high rushing qualities were partitioned according to field location, with plays performed 51 yards or closer to the goal line were 7.6% more likely to rush (13.2% $>$ 5.6%). Albeit, teams classified with low rushing proficiency were additionally distinguished according to deep passing quality whereby teams with high deep passing proficiencies on plays further than 46 yards from goal were 3.6% more likely to pass the ball relative to attempts from closer to the goal line ($\leq$ 46 yard line). Finally, teams with low deep passing quality were subsequently differentiated in scenarios with >7 and ≤ 7 yards to gain, displaying the least yet significant ($p<.001$) difference in passing, rushing and scrambling tendencies (Figure 12).

5.4 Discussion

Previous research has demonstrated the influence of situational variables towards predict play-calling tendencies in the NFL (Lee et al., 2016; Goyal, 2020; Joash Fernandes et al., 2020). However, in addressing hypothesised inaccuracies in prediction due to variation in style of play, the present study has demonstrated the relative importance of playing style proficiencies in modelling play selection tendencies. Furthermore, the impact of relative quality among styles of play were shown to tangibly alter the probability of pass, run and scramble attempts in any given scenario (Appendix H), notwithstanding variation in precision and sensitivity. Indeed, the inclusion of scrambling outcomes rendered the PRS model slightly less accurate in predictive efficacy than the PR model (68.65% $<$ 70.5%). Though, potentially due to the improvised nature of scramble attempts (Reed et al., 2006; Yurko et al., 2019; Craig and Winchester, 2021), the PRS model failed to predict their occurrence. Despite the foreseeable unpredictability of scramble

attempts under the conditional inference tree approach, CTREEs presented an effective visual technique with regard to the influence of each situational variable. Moreover, further analysis of variable importance highlighted several covariates - most notably quantifying the predominance of down, yards-to-go, score differential, time remaining (per half) and quarter, relative to the significant yet lesser splitting capacity of timeouts remaining, yard line, passing efficacy, scrambling, no huddle, rushing deep passing and play-clock variables.

5.4.1 Overall model accuracies

Within the context of previously adapted models, this research design utilised CTREEs, considered the appropriate variable selection, and included the possibility of scrambling play types in order to remain practical in real-game use-cases. However, when compared to more theoretical models achieving higher accuracies using ‘black box’ machine learning approaches (Guidotti et al., 2018), the PR model underperformed in overall accuracy relative to those observed by Lee et al. (2016) and Joash Fernandes et al. (2020) by approximately 3.3 to 5.4% (73.8-75.9%). However, it is important to note that a lack of practicality in variable selection could confound models and potentially cause resulting predictions to become retrospective rather than prospective. More importantly, broad and potentially indiscriminate consideration of variables may cause restraints in practice due to the computational demands of increasingly complex and opaque decision techniques (Rudin, 2019), further exaggerated by the time constraints for communication between coaches and players (NFL, 2008). For example, the addition of one variable excluded from the PRS and PR models - identifying whether a team were in shotgun formation at the snap (start of the play), improved the prediction accuracies by approximately 3% (71.59% and 73.55% respectively). Nonetheless, while these variables improve overall prediction accuracy, they may inadvertently cause meaningful decreases in usability of resulting models (Goyal, 2020).

Notwithstanding, the efficacy of decision trees for play type prediction have been evaluated by Goyal (2020) and Joash Fernandes et al. (2020) using classification and regression trees (CART), albeit with slight variations in methodology. For instance, while excluding 4th down attempts in a model similar to PR, while additionally incorporating several in-game probabilities, Goyal (2020) achieved a 0.7% higher prediction accuracy (71.2>70.5%). More interestingly, in an

attempt to reduce the influence of impractical variable selection, Joash Fernandes et al. (2020) proposed two decision tree models - distinguishing situational variables according to availability pre-communication cut-off and pre-snap with the aim of developing an analytical tool for use by both players and coaches alike. Moreover, this variation among inputs highlighted the influence on dynamic variables in predicting play-calling tendencies, increasing predictive power by 7% (Joash Fernandes et al., 2020). In contrast with a near static situational decision tree presented (65.3% accuracy) by Joash Fernandes et al. (2020), both PRS and PR models provided greater predictions accuracy by 3.35% and 5.2% respectively.

5.4.2 Situational predictability

The implementation of static measures (down and quarter) has not been consistent among research methodologies, with specific designs artificially reducing continuous variables (e.g., yards-to-go, time remaining, score difference and yard-line) into categories on the basis of anecdotal conventions (Goyal, 2020), thus potentially restricting their utility within, and restricting the fit of resultant models (Stilling and Critchfield, 2010). Nonetheless, existing frameworks to assess accuracy amid natively categorical variables such as quarter or down as a means of comparison (Lee et al., 2016; Goyal, 2020; Joash Fernandes et al., 2020) which exhibit the greatest variable importance among covariates within the present finding.

Play-calling predictability has been found to vary according to the quarter in which a play occurs. More specifically, consensus within present and previous research highlights the 1st quarter of play as the least predictable period (Lee et al., 2016; Goyal, 2020; Joash Fernandes et al., 2020). Whereas for the remaining quarters, Lee et al. (2016) and Goyal (2020) demonstrated increased predictions accuracies in the final quarters of each half (2nd and 4th) in accordance with the findings of both PRS and PR CTREEs, while Joash Fernandes et al. (2020) found an a more stable, yet lesser increase in test accuracy throughout the game. Albeit, outside of regulation quarters (1st-4th quarter), overtime periods (coded as 5th quarter) have only been included in black-box decision systems employed by Lee et al. (2016) revealing substantial decreases in predictability - corroborating values obtained through CTREE models and manifesting as the least predictable period of games. Nevertheless, while the methodology employed by Joash Fernandes et al. (2020)

provides the greatest equivalence (CART) to conditional inference tree models, prediction accuracy achieved by the PRS and PR CTREES superseded the performance of CART models by averaging 4.4% and 6.3% higher test accuracy respectively. This trend is perpetuated in testing accuracy per down for the PR (pass or run) CTREE relative to the aforementioned classification and regression tree (CART), with an average increase of 6.1% from 1st to 3rd downs, albeit approximately equal in predictive power on 4th downs (82.7<83.4%). Interestingly, the PRS (pass, run or scramble) CTREE outperformed the CART model proposed by Joash Fernandes et al. (2020) on 1st and 2nd downs, yet provided a slightly lower accuracies in scenarios occurring on latter downs, with a more pronounced difference on 4th downs (80.14<83.4%). On the other hand, despite disparities between CTREES and black-box techniques, similarities in findings indicate that play types are the least predictable on 1st downs, whilst play-calling becomes more predictable on 2nd downs, with subsequent 3rd downs exhibiting the greatest levels of predictability, decreasing slightly thereafter in 4th down situations (Lee et al., 2016).

5.4.3 Play calling tendencies

Tendencies in play selection relative to situational features have received limited direct analysis, albeit Reed et al. (2006) demonstrated that teams are biased in play-calling on 1st downs towards running the ball, whereas teams tend to pass on 3rd down attempts. Therefore, uniquely discussing the accuracy of the models may be insufficient in quantifying the performance of ML techniques in American football. Indeed, in determining the utility of a passing or rushing model, Lee et al. (2016) and Joash Fernandes (2020) further tested for the likelihood of false positive (precision) and false negative (sensitivity) predictions. While it has been argued that false negative (sensitivity) outcomes are more important in attempts to predict passing plays - allowing teams to more easily adapt to the run from defensive coordination intended for pass plays (Joash Fernandes et al., 2020), support for this notion is currently hypothetical. Nonetheless, in contextualising the accuracy of ML models the present study adopted the precision (false negative) and sensitivity (false positive) framework exercised by Lee et al. (2016). Yet, in evaluating the impact of false positives and negatives on predictions made by conditional inference trees, the pass or run (PR) model produced slightly greater sensitivity than precision (76.12>75.22%) on passing plays and a

slightly greater precision than sensitivity on running plays (63.12>61.98), while Lee et al. (2016) observed the opposite effect to a more pronounced extent.

Furthermore, several differences in precision and sensitivity in predicting specific play types according to the down and quarter were observed. For instance, excluding 5th quarters (overtime), the precision of pass and run play predictions appeared relatively stable between quarters, indicating lesser variation in false positive results with a notable decrease in false positive rate on running plays in the 4th quarter - potentially attributable to teams chasing or retaining a lead in score differential in the latter stages of games (Stilling and Critchfield, 2010; Lee et al., 2016). In contrast, passing predictability exhibited a greater frequency of false negatives in 1st and 3rd quarters (lower sensitivity), while fewer false negatives occurred were made in corresponding situations for predicted run attempts. This inverse relationship in sensitivity may be seen as a reflection of further biases in play calling tendencies - highlighting sensitivity as an ineffectual measure for the interpretation of play type predictability. Whereas sensitivity may be reliant on the tendencies for teams to run the ball earlier in each half of play and pass at the end of halves - in so-called 'volatile situations' (Reed et al., 2006), a similar effect was observed in prediction sensitivity for play types according to current down - reflecting and extending findings of running and passing biases by Reed et al. (2006). As such, greater sensitivity, thus a lower likelihood of false negatives was exhibited in the predictability of running plays on 1st and 4th down situations relative to passing predictions, while the opposite effect occurred on 2nd and 3rd downs - indicative of the prevalence of passing in these scenarios. Conversely, the utility of the precision measure in evaluating the efficacy of play type predictions may be less susceptible to tendency bias and provide greater insight. For example, the predictability of running plays increases in precision incrementally from 1st to 4th down, whereas predictability of pass attempts followed a similar trend in precision, however, 2nd down testing accuracy was the least precise.

5.4.4 Application of Conditional Inference Trees

The importance of visually interpretable decision trees, accompanied by clear differences in play type probabilities may aid in play-calling and identifying opposition behavioural trends. Moreover, play-calling tendencies in the NFL are suggested to hold a substantial influence over match

outcomes (Lee et al., 2016). Therefore, while play prediction models including scrambling play types are unable to accurately identify scrambling outcomes relative to passing or rushing attempts, visual approaches depicting the likelihood of each play type eventuality may be more appropriate in practice when compared to precision and sensitivity. For example, an increased likelihood of a quarterback scrambling (8.5%) due to high scrambling proficiency in situations with >2 yards to go within a 7-score differential may inform coaches decision to call for a quarterback spy (defensive man-to-man coverage on the quarterback). Moreover, given the same situation, coaches may devise adjustments to account for a 5.5% differential in running likelihood depending on opposition rushing proficiency.

5.4.5 Limitations and future research

The relative importance of four offensive playing style proficiencies were demonstrated to retain significant splitting rule power within the conditional inference tree models. However, removing opposition defensive qualities within established styles of play resulted in unsubstantial decreases (<0.1%) in overall model accuracy. Therefore, greater granularity in data collected on defensive behaviours to further inform defensive playing styles in addition to modelling defensive play selection (e.g., man or zone coverage and blitz, 3 or 4-man rushing the passer) may produce greater insight for offensive play-calling. Notwithstanding, while the models devised in this study present an analytical tool that may be prepared prior to performance for use in contextualising decision-making during games, they do not account for the likelihood of specific subcategories of run, passing plays or formation types from which they may stem. Although such detail may require advanced knowledge and expertise to define, additional attention could be of use to practitioners. Furthermore, implementation of the models proposed in this study may produce accurate results when complemented by in-season data to account for intra-season variation in playing style proficiency - providing a contemporary reflection of their influence on play calling tendencies.

5.5 Conclusion

Contrary to previous attempts to accommodate for team proficiency through the use of individual rankings in distinct areas of play, this study evaluated the impact of teams grouped according to

similarity in performance of play-calling tendencies in the NFL. Subsequently, the utilisation of easily interpretable decision trees, highlighted the influence of differences in style of play between teams on the likelihood of teams executing a pass, run or scramble attempt. Furthermore, conditional inference trees were found to be substantially better suited to identifying play selection tendencies relative to classification and regression trees containing prospective variable selection in prediction accuracy. Moreover, the resulting testing accuracy provided approximately equal results when compared to decision tree models trained on conceivably retrospective situational variables (e.g., formation, positional personnel count). Further analysis of variable importance revealed the relative influence of situational covariates in decision-making. Additionally, while less influential on play-type execution than certain situational variables (e.g., down, score differential, yards-to-go), playing style proficiencies were demonstrated to influence play-calling tendencies in the National football league.

6. General discussion and conclusions

6.1 Summary of findings

Developing an understanding of unpredictable and chaotic sports as a sum of their characteristics or key performance indicators has driven previous attempts to predict team performance and success (Mackenzie and Cushion, 2012). However, Mackenzie and Cushion (2012) suggest that for the enhancement of performance analysis, more advanced techniques should be used to augment current understanding in sport. In the present thesis, a mixture of unsupervised (principal component analysis and k-means cluster analysis) and supervised (multiple linear regression and conditional inference trees) machine learning approaches were used to identify and quantify the role of playing styles as a predictor of success and predictability in the National football league. Moreover, these techniques allowed for an analysis of previously unexplored and complex relationships in American football performance. More importantly, the findings from the modelled data allowed for the extraction of knowledge in an intelligible way (Louppe, 2014). The systematic and interpretable means of predicting phenomena from a set of measures utilised in this thesis may broaden the scope of research concerning performance in American football and approaches to analyses, while providing techniques that may be easily applied as analytical tools in professional practice. For instance, the development of playing styles by means of principal components analysis elucidates the contribution of each variable within respective styles of play in the form of factor loadings. Additionally, in conducting this methodology, a team may identify specific areas of play which may improve overall performance under a specific style. This was most notably demonstrated in components such as passing efficacy and rushing, where variables detailing how respective total yards, 1st, and 3rd downs were achieved through yards before and after contact or catch. However, while exploratory factor analyses depict the structures present among the sampled variables, it can be argued that their use in weighting team profiles according to efficacy - thus proficiency, allows for greater insight into relative team performance and provides a framework to analyse strengths and weaknesses. Moreover, determining actual differences between teams according to proficiencies in styles of play and their association with success in the NFL can be used to highlight areas for improvement and to expose opposition shortfalls. What's more, the use

of playing style proficiencies in predicting wins may facilitate and focus the formulation of training sessions and protocols. Indeed, measures quantifying team efficacy within playing styles may be used as situational variables among others used by coaches to prepare teams for upcoming opponents and scenarios which may occur (Kormelink and Seeverens, 1999).

Contrary to many football codes, where the influence of situational variables including location (i.e., home or away), match status (i.e., winning, drawing or losing) and team quality (i.e., strong, medium or weak) are the most impactful factors (Csataljay, James, Hughes and Dancs, 2012; Carling et al., 2005; Lago, 2009; Taylor et al., 2008; Tucker, Mellalieu, James, & Taylor, 2005) which may inform macro strategic approaches (Araújo et al., 2015), American football presents the ability to assess the influence of situational variables each play on a micro level. Furthermore, whereas the implementation of these parameters in training environments may not typically be pragmatic due to limitations in time and resources (Taylor et al., 2008), this may not apply in competitions such as the NFL. For example, teams employ practice squads to emulate opposition behaviour whilst providing suitable reserves in case of injuries (Calvetti Jr, 2023). Notwithstanding, the use of decision trees has previously been utilised to produce an analytical tool for coaches to identify the most likely behaviour in a given scenario (Goyal, 2020; Joash Fernandes et al., 2020), incorporating universal variables such as score differential and period of the game, in addition to sport specific metrics (e.g., down, yards to go, yard line etc.). However, with the inclusion of more advanced covariates derived from performance efficacy, the final study within this thesis demonstrated the relative importance of clearly delineated team proficiencies in playing styles among conventional situational variables in the prediction of offensive behaviours. Equally important within practice, is that the inclusion of playing style proficiencies further increased the prediction accuracy of easily interpretable models, in addition to illustrating the likelihood of each potential offensive behaviour (i.e., pass, run or scramble).

6.2 Limitations and recommendations for future research

Between the studies conducted in the present thesis, variation was apparent in the impact demonstrated by styles of play to predict success and behaviours in the National football League. Within the eight established playing styles, team quality in passing efficacy, rushing, targeting and

separation, tackling and quarterback suppression styles all produced significant positive predictions of regular season success. However, among playing style proficiencies influencing the likelihood of behavioural outcomes (i.e., pass, run or scramble), only passing efficacy, deep passing, rushing and scrambling were predictors of play type. While this may indicate that opposition efficacy among defensive playing styles does not significantly affect offensive play calling predictability or tendencies in the NFL, this disparity in influence of playing style proficiencies may stem from an inadequacy of defensive measures collected in American football. This issue was highlighted by the difference between the total of marker variables identified among components interpreted as representations of defensive areas of play relative to those observed in offensive features. Therefore, greater granularity in defensive data collection practices to equal those observed in offensive areas may provide a better understanding of the importance of playing styles for individual teams, as well as further comprehension of the interactions between teams. Indeed, the research designs presented within this thesis attempted to account for on-field behaviours and the influence of several situational variables, including team and player proficiencies within certain styles or areas of the game on play-calling, thus decision making. However, they do not necessarily account for individualised coaching philosophies and approaches to the game. Therefore, future research may seek to investigate head coaching consistency in the implementation and efficacy of playing styles in American football across multiple seasons and teams.

6.3 Applications

For practical applications to ensure both playing styles and team proficiencies constitute a true reflection of elite American football performance, further developments concerning style of play should conduct a hybrid data collection akin to suggestions by Joash Fernandes et al. (2020), with machine learning techniques to incorporate a combination of historical and recent data. On the other hand, prior suggestions to implement in-game predictive analytical tools have advised the development of web or mobile applications (Lee et al., 2016). Nonetheless, while the NFL provides teams with identically configured tablets, they are currently unable to install any software or features to gain a competitive edge over their opponent (Lee et al., 2016; NFL Football Operations, 2024). However, with the option to use still images or printouts to inform decision

making, teams could benefit from singular or multiple decision tree visualisations such as those demonstrated in the present thesis. Additionally, although these restrictions currently and apply to NCAA (National Collegiate Athletic Association) [American] football (Johnson, 2024) - a subsidiary and tributary competition for the NFL, models produced in the present research may prove advantageous in contextualising performance analyses outside of the NFL. Moreover, in future studies, these techniques may improve upon previously highlighted limitations in prediction accuracies relative to the NFL caused by a wider range of team and opposition strengths (Blaikie et al., 2011).

References

- Amatria, M., Maneiro, R., Casal, C.A., Papadopoulou, S., Sarmiento, H., Ardá, A., Iglesias, X. and Losada, J.L., 2021. Differences in Technical Development and Playing Space in Three UEFA Champions Leagues. *Frontiers in psychology*, 12, p.695853.
- Ang, J.C., Mirzal, A., Haron, H. and Hamed, H.N.A., 2015. Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. *IEEE/ACM transactions on computational biology and bioinformatics*, 13(5), pp.971-989.
- Anyama, O.U. and Igiri, C.P., 2015. An application of linear regression & artificial neural network model in the NFL result prediction.
- Araújo, D., Passos, P., Esteves, P., Duarte, R., Lopes, J., Hristovski, R. and Davids, K., 2015. The micro-macro link in understanding sport tactical behaviours: Integrating information and action at different levels of system analysis in sport. *Movement & Sport Sciences-Science & Motricité*, (89), pp.53-63.
- Ashford, M., Abraham, A. and Poolton, J., 2021. Understanding a player's decision-making process in team sports: a systematic review of empirical evidence. *Sports*, 9(5), p.65.
- Bartlett, M.S., 1954. A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.296-298.
- Berri, D.J. and Burke, B., 2011. Measuring productivity of NFL players. In *The economics of the National Football League: The state of the art* (pp. 137-158). New York, NY: Springer New York.
- Biro, P. and Walker, S.G., 2022. A reinforcement learning based approach to play calling in football. *Journal of Quantitative Analysis in Sports*, 18(2), pp.97-112.

Blaikie, A.D., Abud, G.J., David, J.A. and Pasteur, R.D., 2011. NFL & NCAA football prediction using artificial neural networks. In Proceedings of the Midstates Conference for Undergraduate Research in Computer Science and Mathematics, Denison University, Granville, OH.

Borkovec, M. and Madin, N., 2019. ggparty:ggplot Visualizations for the partykit Package. R package version, 1(0).

Boslaugh, S., 2012. Statistics in a nutshell: A desktop quick reference. O'Reilly Media, Inc..

Bradley, P.S., Lago-Peñas, C., Rey, E. and Gomez Diaz, A., 2013. The effect of high and low percentage ball possession on physical and technical profiles in English FA Premier League soccer matches. Journal of sports sciences, 31(12), pp.1261-1270.

Bradley, P.S., Lago-Peñas, C., Rey, E. and Sampaio, J., 2014. The influence of situational variables on ball possession in the English Premier League. Journal of Sports Sciences, 32(20), pp.1867-1873.

Budayan, C., 2008. Strategic group analysis: Strategic perspective, differentiation and performance in construction.

Byanna, N. and Klabjan, D., 2016. Evaluating the Performance of Offensive Linemen in the NFL. arXiv preprint arXiv:1603.07593.

Bø, T.H. and Jonassen, I., 2002. New feature subset selection procedures for classification of expression profiles. Genome biology, 3(4), pp.1-11.

Cafarelli, R., Rigdon, C.J. and Rigdon, S.E., 2012. Models for third down conversion in the national football league. Journal of Quantitative Analysis in Sports, 8(3).

Calvetti Jr, P.G., 2023. Optimizing the Allocation of Capital Among Offensive Positions in the NFL (Doctoral dissertation, Massachusetts Institute of Technology).

Carling, C., Williams, A.M. and Reilly, T., 2007. Handbook of soccer match analysis: A systematic approach to improving performance. Routledge.

Carroll, B., 1999. Total football II: The official encyclopedia of the National Football League. Harper Collins.

Carter, V. and Machol, R.E., 1971. Operations research on football. Operations Research, 19(2), pp.541-544.

Cattell, R.B., 1966. The scree test for the number of factors. Multivariate behavioral research, 1(2), pp.245-276.

Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A., 2014. NbClust: an R package for determining the relevant number of clusters in a data set. Journal of statistical software, 61, pp.1-36.

Cintia, P. and Pappalardo, L., 2021. Coach2vec: autoencoding the playing style of soccer coaches. arXiv preprint arXiv:2106.15444.

Clatworthy, J., Buick, D., Hankins, M., Weinman, J. and Horne, R., 2005. The use and reporting of cluster analysis in health psychology: A review. British journal of health psychology, 10(3), pp.329-358.

Clijmans, J., Van Roy, M. and Davis, J., 2022, September. Looking Beyond the Past: Analyzing the Intrinsic Playing Style of Soccer Teams. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 370-385). Cham: Springer Nature Switzerland.

Cohea, C. and Payton, M.E., 2011. Relationships between player actions and game outcomes in american football. Sports science, 15, pp.19-25.

Collet, C., 2013. The possession game? A comparative analysis of ball retention and team success in European and international football, 2007–2010. *Journal of sports sciences*, 31(2), pp.123-136.

Collins, M., Moir, G., Muresan, J. and Savage, D., 2017. *NFL Data Analytics For A New Era*.

Comrey, A.L. and Lee, H.B., 2013. *A first course in factor analysis*. Psychology press.

Correia, V., Araújo, D., Vilar, L. and Davids, K., 2013. From recording discrete actions to studying continuous goal-directed behaviours in team sports. *Journal of Sports Sciences*, 31(5), pp.546-553.

Costa, F.F., 2014. Big data in biomedicine. *Drug discovery today*, 19(4), pp.433-440.

Craig, J.D. and Winchester, N., 2021. Predicting the national football league potential of college quarterbacks. *European Journal of Operational Research*, 295(2), pp.733-743.

Craig, J.D. and Winchester, N., 2021. Predicting the national football league potential of college quarterbacks. *European Journal of Operational Research*, 295(2), pp.733-743.c

Critchfield, T.S., Meeks, E. and Stilling, S.T., 2014. Explanatory flexibility of the matching law: Situational bias interactions in football play selection. *The Psychological Record*, 64, pp.371-380.

Csataljay, G., James, N., Hughes, M.D. and Dancs, H., 2012. Performance differences between winning and losing basketball teams during close, balanced and unbalanced quarters.

da Mota, G.R., Thiengo, C.R., Gimenes, S.V. and Bradley, P.S., 2016. The effects of ball possession status on physical and technical indicators during the 2014 FIFA World Cup Finals. *Journal of sports sciences*, 34(6), pp.493-500.

David, J.A., Pasteur, R.D., Ahmad, M.S. and Janning, M.C., 2011. NFL prediction using committees of artificial neural networks. *Journal of Quantitative Analysis in Sports*, 7(2).

Davis, J.L., Fodor, A., Pfahl, M.E. and Stoner, J., 2014. Team interdependence and turnover: Evidence from the NFL. *American journal of business*, 39(3/4), pp.276-292.

De Ville, B., 2013. Decision trees. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(6), pp.448-455.

Decroos, T. and Davis, J., 2020. Player vectors: Characterizing soccer players' playing style from match event streams. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part III* (pp. 569-584). Springer International Publishing.

Deshpande, S.K. and Evans, K., 2020. Expected hypothetical completion probability. *Journal of Quantitative Analysis in Sports*, 16(2), pp.85-94.

Dežman, B., Trninić, S. and Dizdar, D., 2001. Expert model of decision-making system for efficient orientation of basketball players to positions and roles in the game—Empirical verification. *Collegium antropologicum*, 25(1), pp.141-152.

Drikos, S., Barzouka, K., Balasas, D.G. and Sotiropoulos, K., 2022. Effect of quality of opposition on game performance indicators in elite male volleyball. *International Journal of Sports Science & Coaching*, 17(1), pp.169-177.

Drikos, S., Sotiropoulos, K., Barzouka, K. and Papadopoulou, S., 2019. Multivariate analysis of the success factors in high-level male volleyball: a longitudinal study. *Trends in Sport Sciences*, 26(4).

Duda, R.O. and Hart, P.E., 1973. *Pattern classification and scene analysis* (Vol. 3, pp. 731-739). New York: Wiley.

Dutta, R., Yurko, R. and Ventura, S.L., 2020. Unsupervised methods for identifying pass coverage among defensive backs with NFL player tracking data. *Journal of Quantitative Analysis in Sports*, 16(2), pp.143-161.

Dziak, J.J., Coffman, D.L., Lanza, S.T., Li, R. and Jermiin, L.S., 2020. Sensitivity and specificity of information criteria. *Briefings in bioinformatics*, 21(2), pp.553-565.

Emara, N., Owens, D., Smith, J. and Wilmer, L., 2017. Serial correlation in National Football League play calling and its effects on outcomes. *Journal of behavioral and experimental economics*, 69, pp.125-132.

Faber, I.R., Bustin, P.M., Oosterveld, F.G., Elferink-Gemser, M.T. and Nijhuis-Van der Sanden, M.W., 2016. Assessing personal talent determinants in young racquet sport players: a systematic review. *Journal of sports sciences*, 34(5), pp.395-410.

Fabrigar, L.R. and Wegener, D.T., 2012. Exploratory factor analysis.

Federolf, P., Reid, R., Gilgien, M., Haugen, P. and Smith, G., 2014. The application of principal component analysis to quantify technique in sports. *Scandinavian journal of medicine & science in sports*, 24(3), pp.491-499.

Fernandez-Navarro, J., Fradua, L., Zubillaga, A. and McRobert, A.P., 2018. Influence of contextual variables on styles of play in soccer. *International Journal of Performance Analysis in Sport*, 18(3), pp.423-436.

Fernandez-Navarro, J., Fradua, L., Zubillaga, A., Ford, P.R. and McRobert, A.P., 2016. Attacking and defensive styles of play in soccer: analysis of Spanish and English elite teams. *Journal of sports sciences*, 34(24), pp.2195-2204.

Fernandez-Navarro, J., Fradua, L., Zubillaga, A., Ford, P.R. and McRobert, A.P., 2016. Attacking and defensive styles of play in soccer: analysis of Spanish and English elite teams. *Journal of sports sciences*, 34(24), pp.2195-2204.

Field, A., 2013. *Discovering statistics using IBM SPSS statistics*. sage.

Fokoue, E. and Foehrenbach, D., 2013. A statistical data mining approach to determining the factors that distinguish championship caliber teams in the National Football League.

Ford, J.K., MacCallum, R.C. and Tait, M., 1986. The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel psychology*, 39(2), pp.291-314.

Franks, I.M. and Miller, G., 1986. Eyewitness testimony in sport. *Journal of sport behavior*, 9(1), p.38.

García-Aliaga, A., Marquina Nieto, M., Coterón, J., Rodríguez-González, A., Gil Ares, J. and Refoyo Roman, I., 2023. A longitudinal study on the evolution of the four main football leagues using artificial intelligence: analysis of the differences in English Premier League teams. *Research Quarterly for Exercise and Sport*, 94(2), pp.529-537.

García-Rubio, J., Gómez, M.Á., Lago-Peñas, C. and Ibáñez, J.S., 2015. Effect of match venue, scoring first and quality of opposition on match outcome in the UEFA Champions League. *International Journal of Performance Analysis in Sport*, 15(2), pp.527-539.

Gheyas, I.A. and Smith, L.S., 2010. Feature subset selection in large dimensionality domains. *Pattern recognition*, 43(1), pp.5-13.

Gløersen, Ø., Myklebust, H., Hallén, J. and Federolf, P., 2018. Technique analysis in elite athletes using principal component analysis. *Journal of sports sciences*, 36(2), pp.229-237.

Gollan, S., Bellenger, C. and Norton, K., 2020. Contextual factors impact styles of play in the English premier league. *Journal of sports science & medicine*, 19(1), p.78.

Gollan, S., Ferrar, K. and Norton, K., 2018. Characterising game styles in the English Premier League using the moments of play framework. *International Journal of Performance Analysis in Sport*, 18(6), pp.998-1009.

González-Ródenas, J., Aranda, R. and Aranda-Malaves, R., 2020. The effect of contextual variables on the attacking style of play in professional soccer.

Goyal, U., 2020. Leveraging machine learning to predict playcalling tendencies in the NFL (Doctoral dissertation, Massachusetts Institute of Technology).

Greenham, G., Hewitt, A. and Norton, K., 2017. A pilot study to measure game style within Australian football. *International Journal of Performance Analysis in Sport*, 17(4), pp.576-585.

Guadagnoli, E. and Velicer, W.F., 1988. Relation of sample size to the stability of component patterns. *Psychological bulletin*, 103(2), p.265.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. and Pedreschi, D., 2018. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5), pp.1-42.

Gutkin, M., Shamir, R. and Dror, G., 2009. SlimPLS: a method for feature selection in gene expression-based disease classification. *PloS one*, 4(7), p.e6416.

Gómez, M.Á., Mitrotasios, M., Armatas, V. and Lago-Peñas, C., 2018. Analysis of playing styles according to team quality and match location in Greek professional soccer. *International Journal of Performance Analysis in Sport*, 18(6), pp.986-997.

Gómez, M.Á., Mitrotasios, M., Armatas, V. and Lago-Peñas, C., 2018. Analysis of playing styles according to team quality and match location in Greek professional soccer. *International Journal of Performance Analysis in Sport*, 18(6), pp.986-997.

Haar, A.V., 2019. Exploratory data analysis of passing plays using nfl tracking data. NFL Big Data Bowl: Indianapolis, IN, USA.

Hair, J., Anderson, R., Tatham, R., & Black, W. (1995). *Multivariate data analysis with readings* (3rd ed.). New York: Macmillan.

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E., 2010. *Multivariate data analysis*. In *Multivariate data analysis* (pp. 785-785).

Han, J., Kamber, M. and Pei, J., 2011. *Data mining concepts and techniques third edition*. The Morgan Kaufmann Series in Data Management Systems, 5(4), pp.83-124.

Hasan, B.M.S. and Abdulazeez, A.M., 2021. A review of principal component analysis algorithm for dimensionality reduction. *Journal of Soft Computing and Data Mining*, 2(1), pp.20-30.

He, Q., Araújo, D., Davids, K., Kee, Y.H. and Komar, J., 2023. Functional adaptability in playing style: A key determinant of competitive football performance. *Adaptive Behavior*, p.10597123-231178942.

He, Q., Kee, Y.H. and Komar, J., 2021, August. Flexibility, Stability, and Adaptability of Team Playing Style as Key Determinants of Within-Season Performance in Football. In *International Conference on Security, Privacy, and Anonymity in Computation, Communication, and Storage* (pp. 69-73). Cham: Springer International Publishing.

Henderson, F. and Olson, M., 1997. Football's West Coast Offense. *Human Kinetics*.

Henson, R.K. and Roberts, J.K., 2006. Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological measurement*, 66(3), pp.393-416.

Hewitt, A., 2007. Match analyses of Australian international women soccer players using an athlete tracking device. *Performance analysis in soccer: applications of player tracking technology*, p.65.

Ho, T. and Carl, S., 2023. nflreadr: Download 'nflverse' Data. <https://nflreadr.nflverse.com>, <https://github.com/nflverse/nflreadr>.

Hodge, K. and Petlichkoff, L., 2000. Goal profiles in sport motivation: A cluster analysis. *Journal of Sport and Exercise Psychology*, 22(3), pp.256-272.

Hothorn, T. and Zeileis, A., 2015. partykit: A modular toolkit for recursive partytioning in R. *The Journal of Machine Learning Research*, 16(1), pp.3905-3909.

Hothorn, T., Hornik, K. and Zeileis, A., 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*, 15(3), pp.651-674.

Hughes, M. and Franks, I.M. eds., 2004. *Notational analysis of sport: Systems for better coaching and performance in sport*. Psychology Press.

Hughes, M.D. and Bartlett, R.M., 2002. The use of performance indicators in performance analysis. *Journal of sports sciences*, 20(10), pp.739-754.

Humphreys, L.G. and Montanelli Jr, R.G., 1975. An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, 10(2), pp.193-205.

Ibrahim, P., 2021. *A Spatial Framework for Analyzing NFL Offensive Line Play*.

James, B., 1977. Baseball Abstract: Featuring 18 Categories of Statistical Information That You Just Can't Find Anywhere Else. Lawrence, Kansas: Self-published.

James, N., Taylor, J. and Stanley, S. (2007), 'Reliability procedures for categorical data in performance analysis', *International Journal of Performance Analysis in Sport* 7.

Jeon, G. and Park, J., 2021. Characterizing patterns of scoring and ties in competitive sports. *Physica A: Statistical Mechanics and its Applications*, 565, p.125544.

Joash Fernandes, C., Yakubov, R., Li, Y., Prasad, A.K. and Chan, T.C., 2020. Predicting plays in the National Football League. *Journal of Sports Analytics*, 6(1), pp.35-43.

Johnson, G. (2024) *Football Rules Committee proposes technology rules, NCAA.org*. Available at: <https://www.ncaa.org/news/2024/3/1/media-center-football-rules-committee-proposes-technology-rules.aspx#:~:text=Teams%20could%20have%20up%20to,capability%20or%20other%20communication%20access>. (Accessed: 06 April 2024).

Johnstone, I.M. and Titterton, D.M., 2009. Statistical challenges of high-dimensional data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1906), pp.4237-4253.

Jolliffe, I.T., 2002. Principal component analysis for special types of data (pp. 338-372). Springer New York.

Jordan, J.D., Melouk, S.H. and Perry, M.B., 2009. Optimizing football game play calling. *Journal of Quantitative Analysis in Sports*, 5(2).

Kaiser, H.F., 1974. An index of factorial simplicity. *psychometrika*, 39(1), pp.31-36.

Kim, J.O. and Mueller, C.W., 1978. Factor analysis: Statistical methods and practical issues (Vol. 14). sage.

Kong, L., Zhang, T., Zhou, C., Gomez, M.A., Hu, Y. and Zhang, S., 2022. The evaluation of playing styles integrating with contextual variables in professional soccer. *Frontiers in Psychology*, 13, p.1002566.

Kormelink, H. and Seeverens, T., 1999. Match analysis and game preparation. Reedswain Inc..

Kuhn, M., 2008. Building predictive models in R using the caret package. *Journal of statistical software*, 28, pp.1-26.

Lago, C., 2009. The influence of match location, quality of opposition, and match status on possession strategies in professional association football. *Journal of sports sciences*, 27(13), pp.1463-1469.

Lago-Ballesteros, J., Lago-Peñas, C. and Rey, E., 2012. The effect of playing tactics and situational variables on achieving score-box possessions in a professional soccer team. *Journal of sports sciences*, 30(14), pp.1455-1461.

Lago-Peñas, C. and Dellal, A., 2010. Ball possession strategies in elite soccer according to the evolution of the match-score: the influence of situational variables. *Journal of human kinetics*, 25(2010), pp.93-100.

Lago-Peñas, C. and Gómez-López, M., 2014. How important is it to score a goal? The influence of the scoreline on match performance in elite soccer. *Perceptual and motor skills*, 119(3), pp.774-784.

Lago-Peñas, C. and Lago-Ballesteros, J., 2011. Game location and team quality effects on performance profiles in professional soccer. *Journal of sports science & medicine*, 10(3), p.465.

Lago-Peñas, C., Gómez-Ruano, M. and Yang, G., 2017. Styles of play in professional soccer: an approach of the Chinese Soccer Super League. *International Journal of Performance Analysis in Sport*, 17(6), pp.1073-1084.

Lames, M. and McGarry, T., 2007. On the search for reliable performance indicators in game sports. *International Journal of Performance Analysis in Sport*, 7(1), pp.62-79.

Law, M., Stewart, D., Pollock, M., Letts, L., Bosch, J., & Westmorland, M. (1998). Guidelines for Critical Review Form: Quantitative Studies. McMaster University Occupational Therapy Evidence-Based Practice Research Group.

Lee, P., Chen, R. and Lakshman, V., 2016. Predicting offensive play types in the National Football League.

Levshina, N., 2021. Conditional inference trees and random forests. In *A practical handbook of corpus linguistics* (pp. 611-643). Cham: Springer International Publishing.

Livingstone, D.J., Manallack, D.T. and Tetko, I.V., 1997. Data modelling with neural networks: Advantages and limitations. *Journal of computer-aided molecular design*, 11, pp.135-142.

Lock, D. and Nettleton, D., 2014. Using random forests to estimate win probability before each play of an NFL game. *Journal of Quantitative Analysis in Sports*, 10(2), pp.197-205.

Lopez-Valenciano, A., Garcia-Gómez, J.A., López-Del Campo, R., Resta, R., Moreno-Perez, V., Blanco-Pita, H., Valés-Vázquez, Á. and Del Coso, J., 2022. Association between offensive and defensive playing style variables and ranking position in a national football league. *Journal of sports sciences*, 40(1), pp.50-58.

Lord, F., Pyne, D.B., Welvaert, M. and Mara, J.K., 2020. Methods of performance analysis in team invasion sports: A systematic review. *Journal of sports sciences*, 38(20), pp.2338-2349.

Louppe, G., 2014. Understanding random forests. Cornell University Library, 10.

Low, B., Coutinho, D., Gonçalves, B., Rein, R., Memmert, D. and Sampaio, J., 2020. A systematic review of collective tactical behaviours in football using positional data. *Sports Medicine*, 50, pp.343-385.

Mackenzie, R. and Cushion, C., 2012. Performance analysis in football: A critical review and implications for future research. *Journal of sports sciences*, 31(6), pp.639-676.

Mao, L., Peng, Z., Liu, H. and Gómez, M.A., 2016. Identifying keys to win in the Chinese professional soccer league. *International Journal of Performance Analysis in Sport*, 16(3), pp.935-947.

Marcelino, R., Mesquita, I. and Sampaio, J., 2011. Effects of quality of opposition and match status on technical and tactical performances in elite volleyball. *Journal of sports sciences*, 29(7), pp.733-741.

Martín-González, J.M., de Saá Guerra, Y., García-Manso, J.M., Arriaza, E. and Valverde-Estévez, T., 2016. The Poisson model limits in NBA basketball: Complexity in team sports. *Physica A: Statistical Mechanics and its Applications*, 464, pp.182-190.

May, C., Dan, M., Carl, M., Ralph, A. and John, H., 2010. Rush versus pass: Modeling the nfl. *Journal of Quantitative Analysis in Sports*.

McGarrity, J.P. and Linnen, B., 2010. Pass or run: An empirical test of the matching pennies game using data from the National Football League. *Southern Economic Journal*, 76(3), pp.791-810.

McLean, S., Salmon, P.M., Gorman, A.D., Naughton, M. and Solomon, C., 2017. Do inter-continental playing styles exist? Using social network analysis to compare goals from the 2016 EURO and COPA football tournaments knock-out stages. *Theoretical Issues in Ergonomics Science*, 18(4), pp.370-383.

Mettling, M., 2014. Quarterback Passer Rating System: Accessible for All Who Care.

Monaghan, T.F., Rahman, S.N., Agudelo, C.W., Wein, A.J., Lazar, J.M., Everaert, K. and Dmochowski, R.R., 2021. Foundational statistical principles in medical research: sensitivity, specificity, positive predictive value, and negative predictive value. *Medicina*, 57(5), p.503.

Mondal, B. and Choudhury, J.P., 2013. A comparative study on k means and pam algorithm using physical characters of different varieties of mango in India. *International Journal of Computer Applications*, 78(5).

Murray, T.J., 2018. Examining the relationship between scheduling and the outcomes of regular season games in the National Football League. *Journal of Sports Economics*, 19(5), pp.696-724.

NFL, 2008. NFL installs new coach-to-defense communications system, NFL.com. Available at: <https://www.nfl.com/news/nfl-installs-new-coach-to-defense-communications-system-09000d5d809f61c6> (Accessed: 17 March 2024).

NFL, 2024a. NFL.com. Available at: <https://www.nfl.com/scores/> (Accessed: 17 March 2024).

NFL, 2024b. NFL.com. Available at: <http://www.nfl.com/stats/team> (Accessed: 17 March 2024).

NFL Football Operations (2024). In the NFL, balancing technology with tradition. Available at: <https://operations.nfl.com/gameday/technology/technology-and-the-game/> (Accessed: 05 April 2024).

Nguyen, Q., Yurko, R. and Matthews, G.J., 2023. Here Comes the STRAIN: Analyzing Defensive Pass Rush in American Football with Player Tracking Data. arXiv preprint arXiv:2305.10262.

Oliva-Lozano, J.M., Rojas-Valverde, D., Gómez-Carmona, C.D., Fortes, V. and Pino-Ortega, J., 2021. Impact of contextual variables on the representative external load profile of Spanish professional soccer match-play: A full season study. *European Journal of Sport Science*, 21(4), pp.497-506.

O'Donoghue, P., 2008. Principal components analysis in the selection of key performance indicators in sport. *International Journal of Performance Analysis in Sport*, 8(3), pp.145-155.

O'Donoghue, P., 2009. Interacting performances theory. *International Journal of Performance Analysis in Sport*, 9(1), pp.26-46.

Osorio, G., 2011. A brief review of American football rules and statistical variables. *Sportscience*, 15, pp.25-30.

Ötting, M., 2021. Predicting play calls in the National Football League using hidden Markov models. *IMA Journal of Management Mathematics*, 32(4), pp.535-545.

Page, G.L., Fellingham, G.W. and Reese, C.S., 2007. Using box-scores to determine a position's contribution to winning basketball games. *Journal of Quantitative Analysis in Sports*, 3(4).

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E. and Chou, R., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International journal of surgery*, 88, p.105906.

Parmar, N., James, N., Hearne, G. and Jones, B., 2018. Using principal component analysis to develop performance indicators in professional rugby league. *International Journal of Performance Analysis in Sport*, 18(6), pp.938-949.

Parmar, N., James, N., Hughes, M., Jones, H. and Hearne, G., 2017. Team performance indicators that predict match outcome and points difference in professional rugby league. *International Journal of Performance Analysis in Sport*, 17(6), pp.1044-1056.

Perepolkin, D., 2024. polite: Be Nice on the Web. <https://github.com/dmi3kno/polite>.

Plakias, S., Moustakidis, S., Kokkotis, C., Tsatalas, T., Papalexi, M., Plakias, D., Giakas, G. and Tsaopoulos, D., 2023. Identifying soccer teams' styles of play: a scoping and critical review. *Journal of Functional Morphology and Kinesiology*, 8(2), p.39.

Plonsky, L., 2015. *Advancing quantitative methods in second language research*. Routledge.

Posit team, 2023. *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA. <http://www.posit.co/>.

Pro Football Focus, 2024. PFF. Available at: <https://www.pff.com/> (Accessed: 17 March 2024).

Pro-football-reference, 2023. https://www.pro-football-reference.com/about/advanced_stats.htm

Pro-football-reference, 2023. <https://www.pro-football-reference.com/years/>.

R Core Team, 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Raiche, G., Roipel, M. and Blais, J.G., 2006. Non-graphical solutions for the Catell's scree plot. In *The International Annual Meeting of the Psychometric Society*.

Raîche, G., Walls, T.A., Magis, D., Riopel, M. and Blais, J.G., 2013. Non-graphical solutions for Cattell's scree test. *Methodology*.

Reed, D.D., Critchfield, T.S. and Martens, B.K., 2006. The generalized matching law in elite sport competition: Football play calling as operant choice. *Journal of Applied Behavior Analysis*, 39(3), pp.281-297.

Refworks, <https://refworks.proquest.com> (2023) (Accessed: 17/10/2023).

Reyers, M. and Swartz, T.B., 2023. Quarterback evaluation in the national football league using tracking data. *AStA Advances in Statistical Analysis*, 107(1-2), pp.327-342.

Richman, M.B., 1986. Rotation of principal components. *Journal of climatology*, 6(3), pp.293-335.

Rico-González, M., Oliveira, R., Vieira, L.H.P., Pino-Ortega, J. and Clemente, F., 2022. Players' performance during worst-case scenarios in professional soccer matches: a systematic review. *Biology of Sport*, 39(3), pp.695-713.

Robbins, D.W., 2011. Positional physical characteristics of players drafted into the National Football League. *The Journal of Strength & Conditioning Research*, 25(10), pp.2661-2667.

Rojas-Valverde, D., Pino-Ortega, J., Gómez-Carmona, C.D. and Rico-González, M., 2020. A systematic review of methods and criteria standard proposal for the use of principal component analysis in team's sports science. *International Journal of Environmental Research and Public Health*, 17(23), p.8712.

Rojas-Valverde, D., Sánchez-Ureña, B., Pino-Ortega, J., Gómez-Carmona, C., Gutiérrez-Vargas, R., Timón, R. and Olcina, G., 2019. External workload indicators of muscle and kidney mechanical injury in endurance trail running. *International Journal of Environmental Research and Public Health*, 16(20), p.3909.

Ruan, L., Ge, H., Shen, Y., Pu, Z., Zong, S. and Cui, Y., 2022. Quantifying the effectiveness of defensive playing styles in the Chinese Football Super League. *Frontiers in Psychology*, p.2461.

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5), pp.206-215.

Rummel, R.J., 1988. *Applied factor analysis*. Northwestern University Press.

Ruscio, J. and Roche, B., 2012. Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological assessment*, 24(2), p.282.

Sampaio, J., Drinkwater, E.J. and Leite, N.M., 2010. Effects of season period, team quality, and playing time on basketball players' game-related statistics. *European Journal of Sport Science*, 10(2), pp.141-149.

Sampaio, J., Lago, C., Casais, L. and Leite, N., 2010. Effects of starting score-line, game location, and quality of opposition in basketball quarter score. *European Journal of Sport Science*, 10(6), pp.391-396.

Sankaran, S., 2014. Comparing pay versus performance of IPL Bowlers: an application of cluster analysis. *International Journal of Performance Analysis in Sport*, 14(1), pp.174-187.

Sarmiento, H., Clemente, F.M., Harper, L.D., Costa, I.T.D., Owen, A. and Figueiredo, A.J., 2018. Small sided games in soccer—a systematic review. *International journal of performance analysis in sport*, 18(5), pp.693-749.

Sarmiento, H., Marcelino, R., Anguera, M.T., Campaniço, J., Matos, N. and Leitão, J.C., 2014. Match analysis in football: a systematic review. *Journal of sports sciences*, 32(20), pp.1831-1843.

Satre-Meloy, A., Diakonova, M. and Grünewald, P., 2020. Cluster analysis and prediction of residential peak demand profiles using occupant activity data. *Applied Energy*, 260, p.114246.

Sebastian, C. and Baldwin, B., 2024. nflfastR: Functions to Efficiently Access NFL Play by Play Data. <https://github.com/nflverse/nflfastR>.

Shea, B.J., Reeves, B.C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E. and Henry, D.A., 2017. AMSTAR 2: a critical appraisal tool for

systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *bmj*, 358.

Sheehan, W.B., Tribolet, R., Novak, A.R., Fransen, J. and Watsford, M.L., 2023. A holistic analysis of collective behaviour and team performance in Australian Football via structural equation modelling. *Science and Medicine in Football*, 7(1), pp.64-73.

Sheehan, W.B., Tribolet, R., Watsford, M.L., Novak, A.R., Rennie, M. and Fransen, J., 2022. Tactical analysis of individual and team behaviour in professional Australian Football: Original investigation. *Science and Medicine in Football*, 6(2), pp.172-180.

Sheehan, W.B., Tribolet, R., Watsford, M.L., Novak, A.R., Rennie, M. and Fransen, J., 2022. Tactical analysis of individual and team behaviour in professional Australian Football: Original investigation. *Science and Medicine in Football*, 6(2), pp.172-180.

Shelly, Z., Burch, R.F., Tian, W., Strawderman, L., Piroli, A. and Bichey, C., 2020. Using K-means clustering to create training groups for elite American football student-athletes based on game demands. *International Journal of Kinesiology and Sports Science*, 8(2), pp.47-63.

Stevens, J.P., 2012. *Applied multivariate statistics for the social sciences*. Routledge.

Stilling, S.T. and Critchfield, T.S., 2010. The matching relation and situation-specific bias modulation in professional football play selection. *Journal of the Experimental Analysis of Behavior*, 93(3), pp.435-454.

Stimel, D., 2009. A statistical analysis of NFL quarterback rating variables. *Journal of Quantitative Analysis in Sports*, 5(2).

Streiner, D.L., 1994. Figuring out factors: the use and misuse of factor analysis. *The Canadian Journal of Psychiatry*, 39(3), pp.135-140.

Strobl, C., Boulesteix, A.L., Zeileis, A. and Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, 8, pp.1-21.

Tabachnick, B.G. and Fidell, L.S., 2012. *Using Multivariate Statistics*.

Tabachnick, B.G., Fidell, L.S. and Ullman, J.B., 2013. *Using multivariate statistics* (Vol. 6, pp. 497-516). Boston, MA: pearson.

Tang, J., Alelyani, S. and Liu, H., 2014. Feature selection for classification: A review. *Data classification: Algorithms and applications*, p.37.

Taylor, J.B., Mellalieu, S.D., James, N. and Shearer, D.A., 2008. The influence of match location, quality of opposition, and match status on technical performance in professional association football. *Journal of sports sciences*, 26(9), pp.885-895.

Te Wierike, S.C.M., Van Der Sluis, A., van den Akker-Scheek, I., Elferink-Gemser, M.T. and Visscher, C., 2013. Psychosocial factors influencing the recovery of athletes with anterior cruciate ligament injury: a systematic review. *Scandinavian journal of medicine & science in sports*, 23(5), pp.527-540.

Teich, B., Lutz, R. and Kassarnig, V., 2016. NFL play prediction. *arXiv preprint arXiv:1601.00574*.

Tenga, A., Ronglan, L.T. and Bahr, R., 2010. Measuring the effectiveness of offensive match-play in professional soccer. *European journal of sport science*, 10(4), pp.269-277.

Teramoto, M., Petron, D.J., Cross, C.L. and Willick, S.E., 2015. Style of play and rate of concussions in the National Football League. *Orthopaedic journal of sports medicine*, 3(12), p.2325967115620365.

Kapil, S. and Chawla, M., 2016, July. Performance evaluation of K-means clustering algorithm with various distance metrics. In *2016 IEEE 1st international conference on power electronics, intelligent control and energy systems (ICPEICES)*(pp. 1-4). IEEE.

Timofeev, R., 2004. Classification and regression trees (CART) theory and applications. Humboldt University, Berlin, pp.1-40.

Trewin, J., Meylan, C., Varley, M.C. and Cronin, J., 2017. The influence of situational and environmental factors on match-running in soccer: a systematic review. *Science and Medicine in Football*, 1(2), pp.183-194.

Trninić, S., Dizdar, D. and Lukšić, E., 2002. Differences between winning and defeated top quality basketball teams in final tournaments of European club championship. *Collegium antropologicum*, 26(2), pp.521-531.

Tucker, L.R., Koopman, R.F. and Linn, R.L., 1969. Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, 34(4), pp.421-459.

Uzoma, A.O. and Nwachukwu, E.O., 2015. A hybrid prediction system for american NFL results. *International Journal of Computer Applications Technology and Research*, 4(01), pp.42-47.

Van Der Maaten, L., Postma, E.O. and van den Herik, H.J., 2009. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10(66-71), p.13.

Van der Maaten, L. and Hinton, G., 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Vogelbein, M., Nopp, S. and Hökelmann, A., 2014. Defensive transition in soccer—are prompt possession regains a measure of success? A quantitative analysis of German Fußball-Bundesliga 2010/2011. *Journal of sports sciences*, 32(11), pp.1076-1083.

Walsh, R., 2021. Identifying significant features for Player Evaluation in NFL comparing ANNs and Traditional Models.

Wedding, C., Woods, C.T., Sinclair, W.H., Gomez, M.A. and Leicht, A.S., 2021. Analysis of styles of play according to season and end of season rank in the National Rugby League. *Journal of science and medicine in sport*, 24(2), pp.206-210.

Wedding, C., Woods, C.T., Sinclair, W.H., Gomez, M.A. and Leicht, A.S., 2021. Exploring the effect of various match factors on team playing styles in the National Rugby League. *International Journal of Sports Science & Coaching*, 16(4), pp.976-984.

Wickham, H., 2024. rvest: Easily Harvest (Scrape) Web Pages. <https://rvest.tidyverse.org/>, <https://github.com/tidyverse/rvest>.

Williams, B., Onsman, A. and Brown, T., 2010. Exploratory factor analysis: A five-step guide for novices. *Australasian journal of paramedicine*, 8, pp.1-13.

Winter, C. and Pfeiffer, M., 2016. Tactical metrics that discriminate winning, drawing and losing teams in UEFA Euro 2012®. *Journal of sports sciences*, 34(6), pp.486-492.

Woods, C.T., McKeown, I., O'Sullivan, M., Robertson, S. and Davids, K., 2020. Theory to practice: performance preparation models in contemporary high-level sport guided by an ecological dynamics framework. *Sports medicine-open*, 6, pp.1-11.

Yadav, S. and Shukla, S., 2016, February. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In 2016 IEEE 6th International conference on advanced computing (IACC) (pp. 78-83). IEEE.

Yang, G., Leicht, A.S., Lago, C. and Gómez, M.Á., 2018. Key team physical and technical performance indicators indicative of team quality in the soccer Chinese super league. *Research in Sports Medicine*, 26(2), pp.158-167.

Yurko, R., Matano, F., Richardson, L.F., Granered, N., Pospisil, T., Pelechrinis, K. and Ventura, S.L., 2020. Going deep: models for continuous-time within-play valuation of game outcomes in American football with tracking data. *Journal of Quantitative Analysis in Sports*, 16(2), pp.163-182.

Yurko, R., Ventura, S. and Horowitz, M., 2019. nflWAR: a reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports*, 15(3), pp.163-183.

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D. and Saeed, J., 2020. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, 1(2), pp.56-70.

Zuccolotto, P., Manisera, M. and Sandri, M., 2018. Big data analytics for modeling scoring probability in basketball: The effect of shooting under high-pressure conditions. *International journal of sports science & coaching*, 13(4), pp.569-589.

Zurloni, V., Cavalera, C., Diana, B., Elia, M. and Jonsson, G., 2014. Detecting regularities in soccer dynamics: A T-pattern approach. *Revista de Psicología del Deporte*, 23(1), pp.157-164.

Zwick, W.R. and Velicer, W.F., 1982. Factors influencing four rules for determining the number of components to retain. *Multivariate behavioral research*, 17(2), pp.253-269.

Appendices

A Keyword searches and combination of Boolean operators

IEEE (222): ("Publication Title":playing style) OR ("Publication Title":style of play) OR ("Publication Title":collective system) OR ("Publication Title":team behaviour) OR ("All Metadata":strategy) AND ("All Metadata":sport) AND ("All Metadata":performance) AND ("All Metadata":analysis)

ScienceDirect (231): "playing style" OR "style of play" OR "collective system" OR "team behaviour" OR strategy" AND "sport" AND "performance AND "analysis"

ACM (323):[Title: "playing style"] OR [Title: "style of play"] OR [Title: "collective systems"] OR [Title: "strategy"] OR [[Title: "team behaviour"] AND [All: "sport"] AND [All: "analysis"] AND [All: "performance"]]

Google Scholar (769): performance AND sport AND analysis AND tactic AND strategy AND "playing style" OR performance AND sport AND analysis AND tactic AND "style of play" OR performance AND sport AND analysis AND tactic AND "collective systems"

PubMed (262): (team behaviour) AND (performance) AND (sport) AND (analysis) OR (playing style) AND (performance) AND (sport) AND (analysis) OR (style of play) AND (sport) AND (performance) AND (analysis) OR (collective systems) AND (sport) AND (performance) AND (analysis) OR (strategy) AND (tactic) AND (sport) AND (performance) AND (analysis)

B Automated duplicate removal and filtering script

```
library(dplyr)
raw_text <- readLines("/pathname/export.ris")
converted <- iconv(raw_text, to = 'ASCII//TRANSLIT')
writeLines(converted, con = "converted_export.bib")
keywords <- c("team behaviour", "playing style", "style of play", "collective systems", "game style", "sport", "tactic", "team", "tactical", "strategy", "strategic", "performance", "league", "match", "indicators", "interaction", "player", "play", "game", "technical", "performance", "sports")
filtered <- revtools::read_bibliography("/Users/bencarter/converted_export.bib", return_df = TRUE) %>% dplyr::mutate(title = tolower(title), keywords = tolower(keywords)) %>%
  filter(!duplicated(title)) %>%
  filter(grepl(keywords, keywords) |
         grepl(keywords, title))
revtools::write_bibliography(filtered, filename = "title_review.ris", format = "ris")
```

C Summary of literary findings

Study	Year	Sport	Aim	Sample	Data types	Method	Reliability	Results	limitations	Comments/Summary	MQ
Amatria, M., Maneiro, R., Casal, C.A., Papadopoulou, S., Sarmento, H., Ardá, A., Iglesias, X. and Losada, J.L., 2021. Differences in Technical Development and Playing Space in Three UEFA Champions Leagues. <i>Frontiers in Psychology</i> , 12, p.695853.	2021	Association Football	To identify the differences between the styles of play that led to the conquest of the UEFA Champions League in the 3-year pre-COVID stage.	UEFA Champions League champions: 2016–2019 resulting in 3000 multi-events (387 complex offensive actions or plays).	Sequence	Pearson’s Chi-square Polar Coordinate analysis	Cohen’s Kappa k > 0.95	Styles of play for all teams analysed (UCL Champions) were identified and described both quantitatively and qualitatively.	The small sample of teams (1 team per season for 3 seasons) reduces the generalisability of the results. Moreover, the sample resulted in only 2 different teams. The nature of the competition analysed further inhibits the applicability of the findings as they may not be an indication of a team’s typical behaviour or style.	The notably small sample size reduces the generalisability and applicability of the study to very few teams. No information was obtained that didn’t pertain to 2 individual champions league-winning teams. Additionally, the relationship between outcomes and plays wasn’t evaluated. Moreover, defensive data was excluded from the analyses, preventing any further insights.	78%
Cintia, P. and Pappalardo, L., 2021. Coach2vec: autoencoding the playing style of soccer coaches. arXiv preprint arXiv:2106.15444.	2021	Association Football	To capture the playing style of professional coaches through a five-step process based on soccer-logs.	All matches from the previous 4 seasons of the Italian first division.	Event & Contextual	k-means cluster analysis Scree test Unsupervised artificial neural network	None reported.	8 distinct styles of play were established.	The study was limited to the variables considered appropriate by the authors, with little consideration to the value of additional variables. The methods employed rendered inhibited potential for the interpretation of the playing styles resulting in few practical applications for the results.	Only 5 variables and permutations (location-based) therein were considered in the analysis. Due to the absence of any attempt to identify or make use of previously employed key performance indicators (variables selected based on author choice) the resultant clusters or playing styles may have been negatively impacted regarding their validity.	56%
Clijmans, J., Van Roy, M. and Davis, J., 2022, June. Looking Beyond the Past: Analyzing the Intrinsic Playing Style of Soccer Teams. In ECML PKDD 2022, Presented at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases.	2022	Association Football	To propose a novel approach for playing style analysis based on a learned model that captures a team’s intrinsic offensive behaviour from historical event stream data.	2019/20 English Premier League - 380 matches.	Event	Discrete-time Markov chain	None reported.	Playing styles identified in the analysis are less “influenced by the rarity of shots and goals”. The method captures a team’s positional and sequential style.	The study pre-determined “high-level indicators of playing style” from which features were derived resulting in potential bias in the findings. Greater or a wider consideration of variables combined with less supervision concerning the procedure would allow for the underlying structure of the game to emerge.	The methodology bypasses extant objective attempts to identify playing styles through the use of performance and key performance indicators, rather characterising styles of play and related variables prior to the analyses - potentially incurring biases.	67%
Decroos, T. and Davis, J., 2020. Player vectors: Characterizing soccer players’ playing style from match event streams. In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part III (pp. 569–584). Springer International Publishing.	2019	Association Football	To characterise a player’s playing style in an objective and data-driven manner based on event stream or play-by-play data.	9155 matches (Premier League, Bundesliga, Primera Division, Serie A and Ligue Un - Almost all matches between the 2012/2013 and 2016/2017 seasons.	Event	Principal components analysis: k ≥ 70% of variance explained non-negative matrix factorisation Manhattan distance	None reported.	The component criterion revealed 4 shooting and crossing components, as well as 5 dribbling and passing components - totalling 18 player playing style vectors.	Determining an inherently team-based characteristic in style of play through the sole consideration of individual actions may produce obscure findings. Resulting in a recommendation to “expand our player vectors with features that capture the tactics a player is involved”. The methods proved a valid method for deriving positional profiles rather than individual playing styles.	In an inherently complex and dynamic sport, the quantity of variables used in the analysis embodies a limitation to the research (19 actions or events). This was accentuated, but the generalisability of the methodology which retains the potential to provide a comprehensive breakdown of player performance and profile.	50%
Fernandez-Navarro, J., Fradua, L., Zubillaga, A., Ford, P.R. and McRobert, A.P., 2016. Attacking and defensive styles of play in soccer: analysis of Spanish and English elite teams. <i>Journal of sports sciences</i> , 34(24), pp.2195–2204.	2016	Association Football	To define styles of play in elite soccer and identify associated performance indicators. To classify teams based on the styles to create playing style profiles.	La Liga and Premier League, (37 teams) - a selection of 97 matches from the 2006/2007 and 2010/2011 seasons.	Match	Principal components analysis: Varimax rotation k1 criterion Factor loadings > 0.7 were included. Kaiser-Meyer-Olkin (KMO) sampling adequacy. Bartlett’s Test of Sphericity Commonalities >0.7	None reported.	6 playing styles were identified accounting for 87.54% of the variance within the data. Furthermore, playing style interactions and teams’ respective efficiency and efficacy may be important for future research. More variables and matches (the sample was particularly arbitrary: “4 matches for 15 teams, 3 matches for 8 teams and 2 matches for 14 teams”) should be considered to supply conclusive definitions for playing styles and generalisability of the data	Self-reported limitations in consideration of contextual variables (namely location and team quality to account for “missing variance”) which may retain the potential to affect styles of play used by teams. This preliminary study established the importance of styles of play in association with team performance - recommending the inclusion of playing style as an important factor in characterising variation between teams. Further emphasising its potential in and applicability in future research. Nonetheless, the sample negatively impacted the generalisability of the findings in regard of its selection of variables and gap between analysed seasons.	89%	
García-Aliaga, A., Marquina Nieto, M., Coterón, J., Rodríguez-González, A., Gil Ares, J. and Refoyo Román, I., 2022. A Longitudinal Study on the Evolution of the Four Main Football Leagues Using Artificial Intelligence: Analysis of the Differences in English Premier League Teams. <i>Research Quarterly for Exercise and Sport</i> , pp.1–9.	2022	Association Football	To analyse the evolution of 4 leagues and identify differences between the EPL and other European leagues.	More than 600 EPL, LaLiga, Bundesliga and Serie A matches from the 2014/2015 to 2018/2019 seasons - around 5000 different players.	Event & Match	t-SNE (t-Distributed Stochastic Neighbor Embedding) RIPPER (Repeated Incremental Pruning Produce Error Reduction)	Inter-operator Kappa > 0.9	The tactical behaviours related to playing styles in top European leagues are becoming more homogenised. Nonetheless, differences were apparent between the styles of play observed in the EPL relative to top European analogues. Moreover, variation of high quality from lower quality teams in the EPL was evident.	In addition to the RIPPER algorithm, the application of logistic regression was recommended. However, the importance of “prior feature selection” or dimensionality reduction was emphasised due to the 52 variables employed generating increasingly complex influences in any attempt to determine styles of play.	Two groups of teams disparate playing styles discussed as statistically significantly different; however, the level of significance was not reported. Nonetheless, the techniques applied produced a valid representation of teams’ styles of play.	90%

Study	Year	Sport	Aim	Sample	Data types	Method	Reliability	Results	limitations	Comments/Summary	MQ
Gollan, S., Ferrar, K. and Norton, K., 2018. Characterising game styles in the English Premier League using the "moments of play" framework. <i>International Journal of Performance Analysis in Sport</i> , 18(6), pp.998-1009.	2018	Association Football	To identify different playing styles on the 2015/16 Premier League season.	380 Premier League games - 760 unique team datasets.	Match	k-means cluster analysis Chi-square	OptaPro Intra-class correlation coefficients (ICC) 0.88–1.00	3 clusters of teams differentiated by game styles were observed. High quality teams embodied a transitional style (dominance in offence and defensive transition). Lesser quality teams performed a majority of games under a possession style. Low quality teams (bottom half of table) played predominantly in established defence.	The predominant findings from the article regressed to future studies expanding the range of the data "beyond the commonly reported game-related variables" in order to develop a greater understanding of game styles in addition to the roles of spatiotemporal data.	The clustering methodology provided an indicator of team quality or proficiencies within the pre-defined playing styles. However, the supervised nature (or "framework") of the factors used in the study incurs potential bias and limitations with regard to its applicability. Furthermore, the season for which the analysis was conducted represented an outlier in 'successful' playing style - with the league champions employing a novel style of play, having only been promoted to the premier league in the previous season.	89%
Gómez, M.Á., Mitrotrasos, M., Armatas, V. and Lago-Peñas, C., 2018. Analysis of playing styles according to team quality and match location in Greek professional soccer. <i>International Journal of Performance Analysis in Sport</i> , 18(6), pp.986-997.	2018	Association Football	To identify the normative styles of play according to team quality and match location using performance indicators related to space, time and actions.	301 matches from the Greek Superleague from the 2013/2014 regular season.	Match	Principal components analysis: Varimax rotation k1 criterion Kaiser-Meyer-Olkin (KMO) sampling adequacy. Commonalities >0.5 Factor loadings > 0.6 were included. k-means cluster analysis	Kappa correlation coefficients Attacking: k = 0.94 Defensive: k = 0.91 Set plays: k = 0.90	8 playing styles were identified. This dimensionality reduction accounted for 64.29% of the variance within the data.	Contextual variables were not considered within the factor analysis. Furthermore, a greater sample regarding matches, variables and leagues was recommended to provide conclusive descriptions and measures for playing styles and generalisability of the data. The combined effects of offensive and defensive tactics and opponent interactions should be included in the future.	Playing styles were found to be affected by match location. Yet, the impact of venue on teams' styles of play is subject to team and opposition quality. The use of factor analysis to score a team's reliance on specific playing styles with this methodology can render an objective assessment of performance.	100%
He, Q., Araújo, D., Davids, K., Kee, Y.H. and Komar, J., 2023. Functional adaptability in playing style: A key determinant of competitive football performance. <i>Adaptive Behavior</i> , p.10597123231178942.	2023	Association Football	To examine playing style adaptability of teams over the duration of a season (i.e., functional variability between matches) and its relationship with match performance indicators.	All matches played in the Premier League, Ligue 1, Bundesliga, Serie A and La Liga from the 2014/15 to 2019/20 seasons.	Event & Match	Unsupervised cluster analysis (Gaussian Mixture Model): Akaike information criterion (AIC) Bayesian information criterion (BIC)	woscored.com via OptaPro Intra-class correlation coefficients (ICC): 0.88–1.00	The model found to be the best fit for the data clustered the "match action profiles" into 12 playing styles.	Greater detail regarding within-match and within-season variation in playing styles or adaptability and their relationship to match performance indicators was recommended for future research (possession or sequential information).	Leading-on from previous research by the authors (He, Lee and Komar, 2021), teams with the capacity to exercise and respond to a wider range of playing styles in the CSL are able to achieve greater offensive and defensive success - leading to a greater win rate.	89%
He, Q., Kee, Y.H. and Komar, J., 2021, August. Flexibility, Stability, and Adaptability of Team Playing Style as Key Determinants of Within-Season Performance in Football. In <i>International Conference on Security, Privacy, and Anonymity in Computation, Communication, and Storage</i> (pp. 69-73). Cham: Springer International Publishing.	2021	Association Football	To investigate the relationship between collective tactical behaviour (i.e., playing style) variability and match performance outcomes throughout the seasons.	All matches played in the top 5 European leagues from 2014/15 to 2019/2020 - 21,708 unique team datasets.	Match	Unsupervised cluster analysis (Gaussian Mixture Model): Akaike information criterion (AIC) Bayesian information criterion (BIC)	None reported.	The best fit for "match action profiles" was 8 playing style clusters for a best fit model (i.e., most representative of the 'optimal' number of playing styles).	Though the methodology allows for identifying distinct styles of play, only assessing teams' playing style on a per-match basis may produce "biased indicator of team performance." Thus, recommendations were made to address the inter-match or within season variability game styles to gain greater practical insight.	The conclusions draw from the exercised methods and resultant findings lead to the authors hypothesising that teams retaining the ability to effectively adapt their styles of play throughout the season may be associated with successful performance - particularly in factors related to offensive team outcomes.	67%
Kong, L., Zhang, T., Zhou, C., Gomez, M.A., Hu, Y. and Zhang, S., 2022. The evaluation of playing styles integrating with contextual variables in professional soccer. <i>Frontiers in Psychology</i> , 13, p.1002566.	2022	Association Football	To explore styles of play between match location and team quality in the CSL.	480 performance records from 240 matches during the 2019 CSL season.	Match, Physical & Contextual	Principal components analysis: k1 criterion k ≥ 60% of variance explained Factor loadings > 0.6 were included. Bartlett's Test of Sphericity Kaiser-Meyer-Olkin (KMO) sampling adequacy.	Amisco Pro Correlation & Reliability coefficients r 0.999 total error 0.05 limits of agreement 0.12	2 principal components (PCs) were discovered. The first PC was positively correlated with physical variables, while the second PC was positively associated with passing variables. These 2 styles of play were interpreted as derivative of "intense" or high intensity and possession approaches respectively.	Self-reported limited generalisability of the findings due to potential for playing styles to produce variation "subject to the policy and rules of the competition" in addition to cross-league tactical or strategies variation. The identified playing styles may be impacted by contextual variables (e.g., fixture congestion, and opposition quality). Inter-performance shift in styles of play and efficacy therein was suggested as an avenue for future CSL research.	High-intensity and possession playing styles were found to be positively associated with team quality of end of season ranking in the Chinese Super League. However, these different styles of play were found to be prevalent for "medium and weak teams" (possession-play regardless of playing location) and "strong" teams (high-intensity).	100%
Lago-Peñas, C., Gómez-Ruano, M. and Yang, G., 2017. Styles of play in professional soccer: an approach of the Chinese Soccer Super League. <i>International Journal of Performance Analysis in Sport</i> , 17(6), pp.1073-1084.	201	Association Football	To describe and measure playing styles in professional soccer.	240 matches from the 2016 Chinese Super League season.	Event	Principal components analysis: Varimax rotation k1 criterion Factor loadings > 0.6 were included. Commonalities >0.5	OptaPro Intra-class correlation coefficients (ICC): 0.88–1.00	5 factors were extracted. The principal components accounted for 79.6% of the variance present within the data.	Situational variables such as location, team quality and the game state) were not considered and may affect styles of play adaptation. The consideration of additional games, action variables and competition variety were recommended to improve the conclusively of interpretation and quantification of styles - further enhancing the generalisability of the results. The interactions of offensive and defensive strategy should be included in the future research.	Various playing styles (Possession, Set Piece, Counterattack and Transitional Play) identified in the current study through the use of performance indicators and factor analysis (principal components analysis). The retained factors explained 79.6% of the variance within the dataset.	89%

Study	Year	Sport	Aim	Sample	Data types	Method	Reliability	Results	limitations	Comments/Summary	MQ
Plakias, S., Kokkotis, C., Moustakidis, S., Tsatalas, T., Papalexi, M., Kasioura, C., Giakas, G. and Tsaopoulos, D., 2023. Identifying playing styles of European soccer teams during the key moments of the game. <i>Journal of Physical Education and Sport</i> , 23(4), pp.878-890.	2023	Association Football	To identify playing styles adopted in multiple European Leagues. To detect playing styles for all phases and sub-phases of the game to develop a structured framework.	2996 matches from the 2021-2022 seasons in 1st league divisions of 11 European countries - 5992 unique team datasets.	Event	Principal components analysis: k1 criterion Varimax rotation Factor loadings >0.6] were included. Kaiser-Meyer-Olkin (KMO) sampling adequacy.	Cohen's Kappa (Instatscout) inter-reliability: k = 0.93	19 components explained 84.81% of the total variance ranging from 18.23% to approximately 1%.	The inclusion of physical performance variables was suggested for inclusion if future analyses. Additional recommendations to account for variation in competition, formations and strategy and their relation to styles of play were reported.	Rather than refining the quantity of factors to retain using alternative methods to the k1 criterion, the study derived 19 factors from the principal components analysis (16 playing styles and 3 match dynamics). A score-based system was used to evaluate a team's reliance on specific styles concerned with phases and sub-phases of the game.	100 %
Ruan, L., Ge, H., Shen, Y., Pu, Z., Zong, S. and Cui, Y., 2022. Quantifying the effectiveness of defensive playing styles in the Chinese Football Super League. <i>Frontiers in Psychology</i> , p.2461.	2022	Association Football	To describe defensive playing styles in professional football by considering a wider range of defensive and spatial variables. To rate the effectiveness of different styles - accounting for opponent's scoring opportunities (xG).	1120 matches from the 2016-2020 CSL seasons.	Event	Principal components analysis: Varimax rotation Kaiser-Meyer-Olkin (KMO) sampling adequacy Bartlett's Test of Sphericity Commonalities >0.6] were included. Factor loadings >0.6] were included.	OptaPro Intra-class correlation coefficients (ICC): 0.88-1.00 SE = 0.0.37 Inter-reliability Cohen's Kappa > 0.9	8 principal components accounted for 85.17% of the total variance present within the data.	It is recommended that off-ball variables related to technical, physical and spatial data be considered in future research. Interactive effects of situational variables (locations, match status, game period) must be addressed to with regard to their influence defensive data inclusion and efficacy of playing styles. The generalisability of findings in the CSL necessitate the analysis of numerous competitions to be validated.	This research identified 8 factors as a result of principal components analysis. The authors were only able to interpret 7 factors as styles of play potentially due to variable selection. This may be a common occurrence among attempts to quantify defensive style - inherent to the scarcity of off-ball variable recording. Additionally, only 15 defensive performance indicators were included in the sample. Further analyses associated 5 of 7 playing styles with defensive proficiency/efficacy in the CSL. This was achieved through the use of defensive styles of play in a multiple regression model - assessing their influence on expected goals conceded.	89%
Sheehan, W.B., Tribolet, R., Novak, A.R., Fransen, J. and Watsford, M.L., 2022. A holistic analysis of collective behaviour and team performance in Australian Football via structural equation modelling. <i>Science and Medicine in Football</i> , pp.1-10.	2022	Australian Rules Football	To utilise aggregate metrics derived by previous research and examine their relation to performance according to match quarter margin in AF.	68 matches providing a sample of 272 quarters. 56 male professionals (age: 24.8 ± 4.5 years; professional playing experience: 5.8 ± 4.3 years) from 1 AFL over 4 seasons (2016-2019).	Network, Match, Physical & Spatiotemporal	Principal components analysis: Scree plot (elbow) Varimax rotation Kaiser-Meyer-Olkin (KMO) sampling adequacy Bartlett's Test of Sphericity Commonalities >0.4] were excluded. Factor loadings <0.4] were excluded.	Intra-class correlation coefficients (ICC): 0.947-1 Root mean squared error (RMSE): 0-4.5	5 principal components interpreted as physical behaviour, scoring opportunity, ball movement, unpredictability, and uncontested behaviour explained 70% of the variance within the data.	The interpretation of principal components requires expert guidance, and they may not account for all of the variation within the data.	The principal components characterised 5 distinct styles of play comprising of scoring ability, attacking ball movement and longitudinal behaviours, grouped by the authors interpretation as "Scoring opportunity", as well as connectedness and low-pressure success - suggested as an indicator of "Ball Movement" proficiency. These factors were found to directly influenced quarter margin. Yet, despite the authors contending that the sample size was sufficient though only analysed one team.	89%
Sheehan, W.B., Tribolet, R., Watsford, M.L., Novak, A.R., Rennie, M. and Fransen, J., 2022. Tactical analysis of individual and team behaviour in professional Australian Football: Original investigation. <i>Science and Medicine in Football</i> , 6(2), pp.172-180.	2022	Australian Rules Football	To reduce the dimensionality of commonly reported spatiotemporal variables and facilitate their practical use and interpretability in AF.	50 matches (1099 individual datasets (22.9 ± 17.6 per player). 48 male professionals (age: 25.0 ± 3.8 years; playing experience: 5.2 ± 3.4 years) from 1 AFL team over 3 seasons (2016-2018).	Spatiotemporal	Principal components analysis: Scree plot (elbow) Varimax rotation Kaiser-Meyer-Olkin (KMO) sampling adequacy Bartlett's Test of Sphericity Commonalities >0.4] were excluded. Factor loadings <0.4] were excluded.	Intra-class correlation coefficients (ICC) low to high-speed running: 0.8-4.8% very high-speed running: 11.5-11.7%	5 team and 4 individual components accounted for 69% and 64% of the variance within the dataset, respectively.	The sample inhibits the applicability and generalisability of the results due to the observation of only 1 professional team over the duration of 3 seasons.	The methodology undertaken in the study provided a somewhat novel approach to establishing playing styles in Australian football through the use of spatiotemporal behaviours. However, as the project provided insights at a case study level (only considering one team's performance), the generalisability of the results inhibits empirical and applicable conclusions.	78%
Sheehan, W.B., Tribolet, R., Watsford, M.L., Novak, A.R., Rennie, M. and Fransen, J., 2020. Using cooperative networks to analyse behaviour in professional Australian Football. <i>Journal of science and medicine in sport</i> , 23(3), pp.291-296.	2020	Australian Rules Football	To reduce the dimensionality of commonly reported complex network characteristics obtained from AFL games to facilitate their practical use and interpretability.	73 matches with 1605 individual observations (33.4 ± 25.6 per player). 48 male professionals (age: 24.97 ± 3.78 years; playing experience: 5.22 ± 3.44 years) from 1 AFL team, playing at least 1 game in 3 seasons (2016-2018).	Network & Match	Principal components analysis: Varimax rotation Kaiser-Meyer-Olkin (KMO) sampling adequacy Bartlett's Test of Sphericity Commonalities >0.4] were excluded. Factor loadings <0.4] were excluded.	Intra-class correlation coefficients (ICC): 0.947-1 Root mean squared error (RMSE): 0-4.5	2 individual and 3 team factors were derived from the principal components analysis accounted for 87.3% and 80.99% respectively within the dataset.	The findings necessitate the validation of the principal components or "new metrics" via their association with team and individual performance outcomes.	The research attempts to simply and elaborate a novel approach to the analysis of complex network structures in Australian Football. However, in sampling and combining both team and individual data, the research falls short of quantifying team playing styles. Contrary to the authors claims, the sole employ of one team in the sample renders the findings limited in generalisability and applicability.	89%
Wedding, C., Woods, C.T., Sinclair, W.H., Gomez, M.A. and Leicht, A.S., 2021. Analysis of styles of play according to season and end of season rank in the National Rugby League. <i>Journal of science and medicine in sport</i> , 24(2), pp.206-210.	2021	Rugby League	To identify styles of play within the NRL relative to season and end of season rank across the 2015-2019 seasons.	16 teams and 1005 matches from the 2015-2019 seasons.	Event	Principal components analysis: Eigenvalues > 1.2 Varimax rotation Kaiser-Meyer-Olkin (KMO) sampling adequacy Bartlett's Test of Sphericity Factor loadings >0.6] were included.	Proprietor-reported reliability (Stats Perform): coefficient of variation <1%	9 factors explained approximately 51% of total performance variance within the NRL.	Contextual variables were not accounted for. Future research should consider the influence of data related to score differential and location. The data was also suggested to be limited in its ability to reflect long-term evolution of playing styles in the NRL.	In relation to end of season rankings or team quality, winning teams' style of play appeared to reflect an "attacking focus". Though the authors determined that the 9 factors resulting from the principal components analysis explained ~51% of the data (highlighted as contrasting extant rugby league research), the sample was restricted by the use of variable availability of a third-party resource.	89%
Wedding, C., Woods, C.T., Sinclair, W.H., Gomez, M.A. and Leicht, A.S., 2021. Exploring the effect of various match factors on team playing styles in the National Rugby League. <i>International Journal of Sports Science & Coaching</i> , 16(4), pp.976-984.	2021	Rugby League	To examine the effect of various match-related factors on team playing styles in the NRL.	16 teams and 1005 matches from the 2015-2019 seasons - 2010 unique team datasets.	Event	Principal components analysis: Eigenvalues > 1.2 Varimax rotation Kaiser-Meyer-Olkin (KMO) sampling adequacy Bartlett's Test of Sphericity Factor loadings >0.6] were included.	Proprietor-reported reliability (Stats Perform): coefficient of variation <1%	9 factors or playing styles accounted for 54.2% of the total technical performance variance (attacking play, line-breaks, errors, defensive play, penalties, play the ball won and lost, conceded linebreak, supports and kick breaks).	Only 21 of the 38 (55%) analysed team performance indicators were included post-dimensionality reduction. Furthermore, just 6 of 9 styles of play were associated with match outcome. Nonetheless, significant differences stated, yet not reported.	In conjunction with previously published research (Wedding et al., 2021), the low proportion of variance explained by the retained factors - post-principal components analysis (51-54.2% using available variables from the same third-party resource) was highlighted as a limitation. In this case with respect to venue and team quality. Nonetheless, winning and losing teams were successfully classified distinct using 4 identified playing styles.	89%

Study	Year	Sport	Aim	Sample & Resources	Method & Search terms	Results	limitations	Comments/Summary	MQ
Lord, F., Pyne, D.B., Welvaert, M. and Mara, J.K., 2020. Methods of performance analysis in team invasion sports: A systematic review. <i>Journal of sports sciences</i> , 38(20), pp.2338-2349.	2020	Team Invasion Sports	To describe traditional and contemporary data capture and methods of performance analysis research in team invasion sports, evaluate the practicality of methods, and formulate practical recommendations for analysing tactics and strategies.	537 articles (completed on the 13 March 2019). SPORTDiscus, Web of Science, Scopus, MEDLINE, and PubMed. Publications from December 1997 to March 2019.	PRISMA (Preferred Reporting Items for Systematic reviews and Meta-analyses) Keywords: "performance analysis", "notational analysis", "match analysis", "game analysis", "tactical analysis", "patterns of play", "game styles", "performance indicators", "dynamic systems", "systems analysis", "sports analytics", "performance analytics", and "team behaviour". Context: "team sport", "invasion games", "field hockey", "soccer", "football", "rugby", "basketball", "netball", "handball", and "water polo". Excluded: "injury", "injury prevention", "recovery", "physical demands", "match demands", "loading", "match load", "technique", "biomechanics", "running", "conditioning", "fitness", "physiology", and "psychology". Searches of keywords, contexts and exclusions used the Boolean operators OR, AND and NOT respectively.	Initial search returned 2153 articles, after which the screening procedure filtered the sample to 537 articles for systematic review. Findings identified 7 articles investigating game styles, in which styles of play were analysed using descriptive statistics (n = 3), factor analysis (n = 3), inferential statistics (n = 2), neural network (n = 2) and prediction (n = 1). Match venue and team quality were found to influence playing style (n = 1). Furthermore, game styles were used to predict "team identity" (n = 1).	Action variables provide the structure for the analysis of test invasion sports - embodying a basis for performance analysis. However, in favour of describing what happened in a performance, outcome-focused PIs do not reveal why events occur. While the analysis of KPIs allows practitioners to differentiate teams through match outcomes, an increasingly holistic approach should be adopted to account for dynamic actions and collective team behaviours through the view of playing styles. The insights that may originate from the analysis of game styles present a greater potential for understanding strategy and produce practical findings.	Approaches capturing and navigating dynamic match events through data aggregation methodologies including factor analysis and neural networks can serve to encompass complex data in the creation of simplified playing styles.	64%
Plakias, S., Moustakidis, S., Kokkotis, C., Tsatalas, T., Papalex, M., Plakias, D., Giakas, G. and Tsaopoulos, D., 2023. Identifying soccer teams' styles of play: a scoping and critical review. <i>Journal of Functional Morphology and Kinesiology</i> , 8(2), p.39.	2023	Association Football	To critically examine how the term is used in the international literature, how research is conducted on styles of play and identify/analyse knowledge gaps.	40 articles (completed on the 25 September 2022). PubMed, Web of Science, and Google Scholar.	22-item Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Keywords: "soccer", and "style" using the Boolean operator AND. Eligibility: Presence of both keywords in the title or in the abstract.	Initial search returned 1417 titles, with 40 articles remaining for the scoping review, post-screening procedure. The findings revealed that 3 study themes were observed: Recognition of playing styles, influence of contextual variables on the adoption of game styles, and style efficacy	Through bias in variable selection, the recognition/identification of playing styles were formulated according to authors' subjective perceptions. Many contextual variables retaining the potential to influence style of play received little to no attention. Moreover, research with a focus regarding on-field practical applications for game styles was presented as a requirement for future research. Sample size and diversity was a notable limitation within the reviewed studies - highlighting 1 competition as insufficient to generalise results.	Researchers should apply a greater emphasis on applicable solutions for teams and coaches. Factor analysis is an excellent method for grouping of variables to describe playing styles, however the interpretation and naming of these structures requires careful attention. Enhancing findings with visualisations improve the interpretability of conclusions. Moreover, the description of novel games styles allows analysts to develop new variables with more tactical significance - offering alternatives methods to evaluate performance.	73%

D Performance variable operational definitions

Variable	Operational definition
Complete pass	Successful pass attempts.
Deep passes	Total pass attempts achieving 15 or more air yards.
Defensive blitzes	Frequency of defensive plays where 5+ players or a defensive back attempted to reach the quarterback.
Defensive combined tackles	The combination of solo and assisted tackles completed by the defence.
Defensive completed air yards allowed	Total completed passing air yards conceded by the defence.
Defensive hurries	Frequency of passing attempted earlier than intended or the quarterback was chased out of the pocket.
Defensive missed tackles	The combination of solo and assisted tackles failed by the defence.
Defensive quarterback knockdowns	Frequency that defenders make contact with the quarterback - forcing them to the ground after a pass attempt.
Defensive sacks	Frequency of tackles made on the passer or quarterback whilst in possession of the ball.
Defensive yards after catch allowed	Total yards after the catch conceded by the defence.
Inaccurate passes	Total pass attempts that were not catchable by the receiver with 'normal' effort.
Incomplete passes	Total attempts to pass the ball forwards that did not result in a catch.
Intended passing air yards	Total distance achieved in the air on passing attempts prior to an attempted catch or pass incompleteness.
Pass attempts	Total attempts to pass the ball forwards.
Passing air yards completed	Total distance achieved in the air on successful catches.
Passing first downs	Frequency of first downs achieved on passing attempts.
Passing yards	The combination of air yards achieved and yards after the catch achieved on completed passing attempts.
Passing yards after catch	Total yards achieved by a receiver after catching a pass attempt.

Quarterback drop	Frequency of plays where the ball was snapped to the quarterback directly behind (in close proximity to) the centre and dropping back with the intent to attempt a pass.
Quarterback hits	Total contacts made defensive players on the quarterback at any time during the play.
Quarterback scramble	Intended passing play resulting in an improvised run by the quarterback reaching beyond the line of scrimmage.
Receiving dropped passes	Total opportunities to catch the ball from passing attempts dropped by the receiver.
Receiving tackles broken	Total tackles broken during pass receiver ball carries.
Rush attempts	Designed running play where no forward pass was attempted.
Rushing first downs	Frequency of first downs achieved on designed run attempts.
Rushing tackles broken	Total tackles broken during rush attempts by ball carriers.
Rushing yards	Total yards achieved during rushing attempts.
Rushing yards after contact	Total yards achieved during rushing attempts after any defensive player makes contact with the ball carrier.
Rushing yards before contact	Total yards achieved during rushing attempts before any defensive player makes contact with the ball carrier.
Scramble yards	Total yards achieved from scramble attempts
Scrambling first downs	Frequency of first downs achieved on scrambling attempts.
Short passes	Total pass attempts achieving 15 or fewer air yards.
Solo tackles	Total individual tackles made on ball carriers.
Successful passing fourth downs	Total fourth down opportunities converted from passing attempts.
Successful passing third downs	Total third down opportunities converted from passing attempts.
Successful rushing fourth downs	Total fourth down opportunities converted from designed run attempts.
Successful rushing third downs	Total third down opportunities converted from designed run attempts.
Successful scrambling fourth downs	Total fourth down opportunities converted from scrambling attempts.
Successful scrambling third downs	Frequency of third downs converted by scrambling attempts.

Tackled for loss	Frequency of ball carriers tackled behind the line of scrimmage on pass, scramble or rush attempts.
Unsuccessful passing fourth downs	Total fourth down opportunities failed from passing attempts.
Unsuccessful passing third downs	Total third down opportunities failed from passing attempts.
Unsuccessful rushing fourth downs	Total fourth down opportunities failed from designed run attempts.
Unsuccessful rushing third downs	Total third down opportunities failed from designed run attempts.
Unsuccessful scrambling fourth downs	Total fourth down opportunities failed from scrambling attempts.
Unsuccessful scrambling third downs	Total third down opportunities failed from scrambling attempts.

E Styles of play and composite weightings (factor loadings)

Style of play	Composite variables (weighting)
passing efficacy	completed passes (.9), passing first downs (.83), passing yards after catch (.82), short passes (.77), passing yards (.76), pass attempts (.61), quarterback drops (.61), successful passing third downs (.48), receiving tackles broken (.46), defensive combined tackles (-.37).
deep passing	deep passes (.97), intended passing air yards (.92), passing air yards completed (.92), inaccurate passes (.58), incomplete passes (.54).
scrambling	scramble yards (.92), scramble attempts (.92), scrambling first downs (.91), successful scrambling third downs (.72), unsuccessful scrambling third downs (.38).
rushing	rushing yards (.94), rushing first downs (.89), rushing attempts (.79), rushing yards before contact (.74), rushing yards after contact (.71), successful rushing third downs (.61), unsuccessful rushing third downs (.5), rushing tackles broken (.38), defensive combined tackles (-.41).
targeting and separation	unsuccessful passing third downs (.81), incomplete passes (.58), quarterback hits (.49), unsuccessful passing fourth downs (.39), inaccurate passes (.42), short passes (.39), quarterback drops (.38), pass attempts (.38), receiving dropped passes (.35).
quarterback suppression	defensive completed air yards allowed (.66), defensive yards after catch allowed (.64), defensive quarterback knockdowns (.49), defensive combined tackles (.47), defensive blitzes (.39).
defensive perturbations	defensive sacks (.61), defensive tackles for loss (.51), defensive interceptions (.47), defensive quarterback knockdowns (.39), defensive blitzes (.36), defensive hurries (.36).
tackling	rushing tackles broken (.6), defensive missed tackles (.58), defensive hurries (.42), receiving tackles broken (.41).

F Example of cluster validation index used for majority rule

Example of the indices consulted in the majority rule using NbClust: an R package for determining the relevant number of clusters in a data set (Charrad et al., 2014):

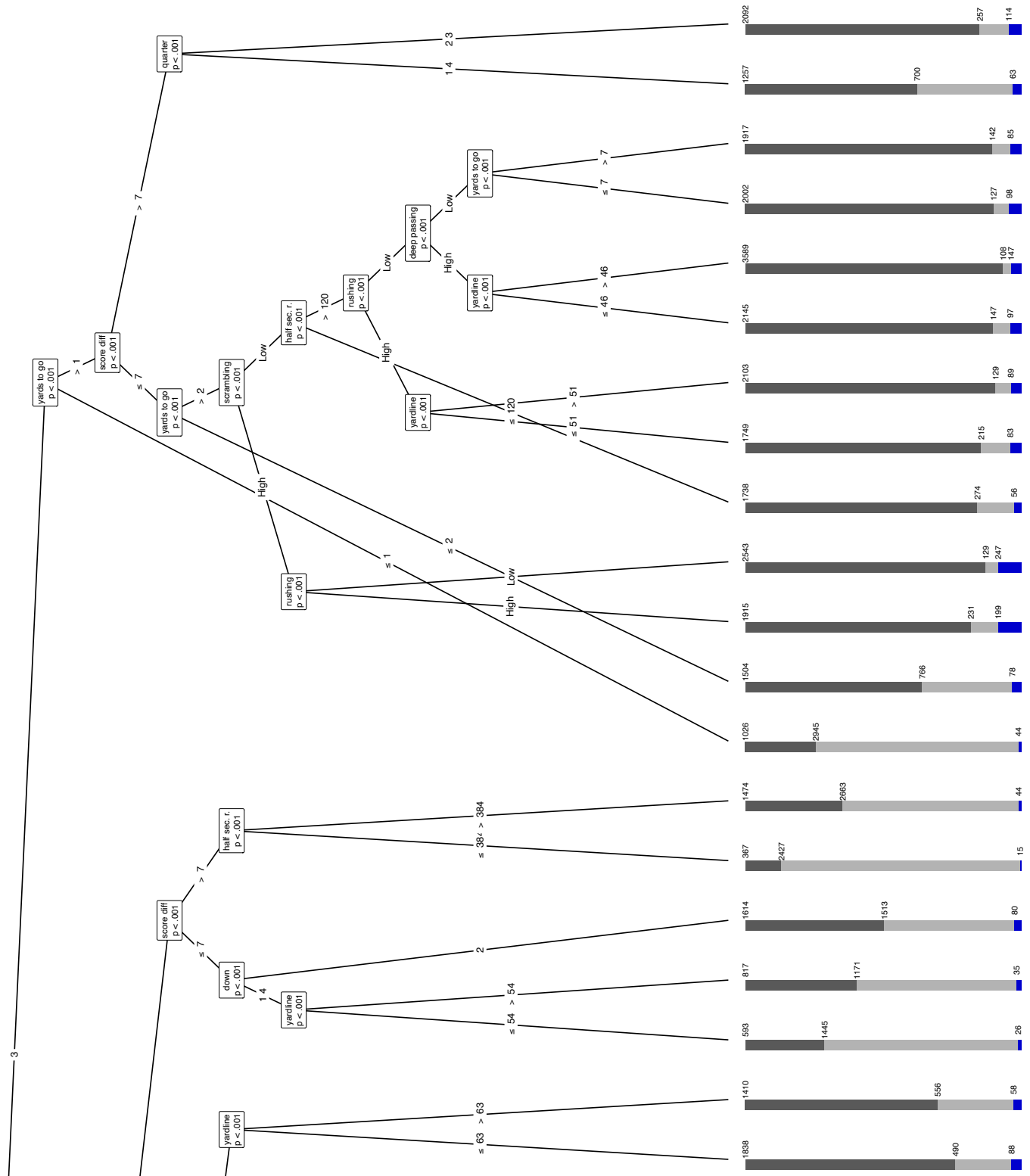
$$\frac{Je(2)}{Je(1)} = \frac{W_k + W_l}{W_m}$$

where $Je(2)$ is the sum of squared errors within clusters when the data are partitioned into two clusters, and $Je(1)$ gives the squared errors when only one cluster is present.

G Situational variable operational definitions

Variable	Operational definition
Deep passing	Classification of playing style proficiency (High, Low)
Defensive perturbations	Classification of playing style proficiency (High, Low)
Divisional game	Categorical indicator of whether the game was a divisional matchup (Yes, No)
Down	The down for any given play (1-4)
Half seconds remaining	The number of seconds remaining in the half to play
No huddle	Categorical indicator of whether the play occurred with or without a pre-snap huddle (Yes, No)
Passing efficacy	Classification of playing style proficiency (High, Mid, Low)
Play clock	The number of seconds remaining when the ball was snapped
Play type	Categorical indicator of whether the executed play type resulted in a pass, run or scramble attempt
Posteam type	Categorical indicator for the possession team location (Home, away)
Quarter	The quarter for any given play (1-5)
Quarterback suppression	Classification of playing style proficiency (High, Mid, Low)
Rushing	Classification of playing style proficiency (High, Low)
Score differential	The score differential for the possession team
Scrambling	Classification of playing style proficiency (High, Low)
Tackling	Classification of playing style proficiency (High, Low)
Targeting and separation	Classification of playing style proficiency (High, Low)
Timeouts remaining	The number of timeouts remaining in the possession team
weather	Categorical indicator of the weather during the game (Dry, Adverse)
Yard line	The number of yards from the opposition endzone
Yards to go	The number of yards from the line to gain

H Conditional inference tree (CTREE)





I Ethical approval letter



London Sport Institute REC

The Burroughs
Hendon
London NW4 4BT

Main Switchboard: 0208 411 5000

24/11/2021

APPLICATION NUMBER: 18704

Dear Ben Adam Carter and all collaborators/co-investigators

Re your application title: Play style proficiency on play calling tendencies (NFL).

Supervisor:

Co-investigators/collaborators:

Thank you for submitting your application. I can confirm that your application has been given APPROVAL from the date of this letter by the London Sport Institute REC.

The following documents have been reviewed and approved as part of this research ethics application:

Document Type	File Name	Date	Version
Aims, objectives and hypotheses	Ben Carter - The influence of play style proficiency on play calling tendencies in the National Football League (NFL). - Project Outline Template	14/09/2021	1
Methods and data	Ben Carter - The influence of play style proficiency on play calling tendencies in the National Football League (NFL). - Research Procedure Template	14/09/2021	1

Although your application has been approved, the reviewers of your application may have made some useful comments on your application. Please look at your online application again to check whether the reviewers have added any comments for you to look at.

Also, please note the following:

1. Please ensure that you contact your supervisor/research ethics committee (REC) if any changes are made to the research project which could affect your ethics approval. There is an Amendment sub-form on MORE that can be completed and submitted to your REC for further review.
2. You must notify your supervisor/REC if there is a breach in data protection management or any issues that arise that may lead to a health and safety concern or conflict of interests.
3. If you require more time to complete your research, i.e., beyond the date specified in your application, please complete the Extension sub-form on MORE and submit it your REC for review.
4. Please quote the application number in any correspondence.
5. It is important that you retain this document as evidence of research ethics approval, as it may be required for submission to external bodies (e.g., NHS, grant awarding bodies) or as part of your research report, dissemination (e.g., journal articles) and data management plan.
6. Also, please forward any other information that would be helpful in enhancing our application form and procedures - please contact MOREsupport@mdx.ac.uk to provide feedback.

Good luck with your research.

Yours sincerely,

Chairs Dr Rhonda Cohen/ Dr Anne Elliott

London Sport Institute REC