

**NOVA**

**IMS**

Information  
Management  
School

# MDDDM

Mestrado em  
**Data-Driven Marketing**

**Data Analytics no Futebol: Avaliação do Desempenho  
Individual, Coletivo e Apoio à Decisão**

Frederico Serpa Santana Ferra

Dissertação

apresentada(o) como requisito parcial para obtenção do grau de Mestre em *Data-Driven Marketing*

**NOVA Information Management School**  
**Instituto Superior de Estatística e Gestão de Informação**  
Universidade Nova de Lisboa



**NOVA Information Management School**  
**Instituto Superior de Estatística e Gestão de Informação**  
Universidade Nova de Lisboa

**Data Analytics no Futebol: Avaliação do Desempenho Individual, Coletivo e Apoio à  
Decisão**

por

Frederico Serpa Santana Ferra

Dissertação apresentada(o) como requisito parcial para obtenção do grau de Mestre em  
*Data-Driven Marketing*, com especialização em Data Science for Marketing

**Orientada por:**

Márcia Lourenço Batista, PhD, NOVA Information Management School

Julho, 2025

## DECLARAÇÃO DE INTEGRIDADE

Declaro ter realizado o presente trabalho académico com integridade. Confirmo que não recorri à prática de plágio ou de qualquer outra forma de utilização indevida de informação ou de falsificação de resultados durante o processo de elaboração deste trabalho. Declaro ainda que tenho conhecimento das Regras de Conduta e do Código de Honra da NOVA Information Management School.

*Lisboa, 13 de Julho de 2025*

Frederico Ferra

## RESUMO

A presente dissertação investiga o papel crescente da análise de dados no futebol moderno, com foco específico na avaliação do desempenho tático e na previsão e apoio à decisão em contexto competitivo. Num cenário marcado pela digitalização e pela disponibilidade de grandes volumes de informação, clubes e equipas técnicas recorrem cada vez mais a tecnologias como *tracking*, estatísticas de eventos e algoritmos de machine learning para compreender e otimizar o comportamento individual e coletivo. Para estruturar esta investigação, foi realizada uma revisão sistemática da literatura existente, seguindo o protocolo PRISMA, que permitiu identificar e analisar quinze estudos empíricos representativos das abordagens mais utilizadas na análise tática contemporânea. O primeiro tópico de análise do trabalho centra-se na avaliação do desempenho individual, explorando modelos que quantificam ações técnico-táticas, métricas físicas e padrões posicionais, incluindo contributos como o PlayeRank, métodos de deteção automática de ações e modelos híbridos que integram fadiga e comportamento tático. O segundo, aborda a avaliação do desempenho coletivo, analisando frameworks baseadas em tracking, padrões de passe, ocupação espacial e organização ofensiva e defensiva. Por fim, são exploradas metodologias de previsão tática e apoio à decisão, com destaque para modelos que antecipam formações, estimam táticas a partir de vídeo, selecionam equipas ótimas ou recomendam substituições em tempo real. O objetivo central deste trabalho foi compreender como diferentes abordagens de Data Analytics contribuem para a leitura tática do jogo e para decisões mais informadas por parte de treinadores e analistas. Os resultados mostram que estas metodologias estão a moldar de forma crescente o planeamento desportivo, revelando potencial significativo para melhorar a avaliação de desempenho, a preparação estratégica e a capacidade de prever dinâmicas em jogo.

## PALAVRAS-CHAVE

Data Analytics; Futebol; Desempenho Tático; Machine Learning; Análise de Performance; Modelos de Previsão Tática

## **ABSTRACT**

This dissertation investigates the growing role of data analysis in modern football, with a specific focus on tactical performance assessment and tactical prediction and decision support in competitive contexts. In an era increasingly shaped by digitalization and the availability of large volumes of information, professional clubs and coaching staffs rely heavily on technologies such as tracking systems, event data and machine learning algorithms to better understand and optimize individual and collective behavior. To structure this investigation, a systematic literature review was conducted following the PRISMA protocol, which enabled the identification and examination of fifteen empirical studies representing the most relevant analytical approaches in contemporary tactical analysis. The first topic of the dissertation focuses on individual performance evaluation, exploring models that quantify technical–tactical actions, physical metrics and positional patterns, including contributions such as PlayeRank, automatic action-detection methods and hybrid models integrating fatigue and tactical behavior. The second, examines collective tactical performance, analyzing frameworks based on tracking data, passing patterns, spatial occupation and offensive and defensive organization. Finally, methodologies related to tactical prediction and decision support are explored, highlighting models capable of forecasting formations, estimating tactics from video, selecting optimal line-ups or recommending in-game substitutions in real time. The central goal of this work was to understand how different Data Analytics approaches contribute to the tactical reading of the game and to more informed decision-making by coaches and analysts. The results show that these methodologies are increasingly shaping sports planning, revealing significant potential to improve performance evaluation, strategic preparation and the ability to anticipate match dynamics.

## **KEYWORDS**

Data Analytics; Football; Tactical Performance; Machine Learning; Performance Analysis; Tactical Prediction Models

# ÍNDICE

DECLARAÇÃO DE INTEGRIDADE	ii
RESUMO	iii
ABSTRACT	iv
<b>1. INTRODUÇÃO</b>	<b>1</b>
<b>2. REVISÃO DE LITERATURA</b>	<b>3</b>
2.1 CONTEXTO E IDENTIFICAÇÃO DO PROBLEMA	3
2.2 MOTIVAÇÕES E OBJETIVOS	4
2.3. ENQUADRAMENTO SOBRE O TEMA	5
2.3.1. <i>Futebol - Uma Visão Geral</i>	5
2.3.2. <i>Conceitos e Regras</i>	5
2.3.3. <i>Football Analytics</i>	6
2.3.4. <i>Dispositivos e Softwares</i>	7
<b>3. METODOLOGIA</b>	<b>8</b>
3.1. ESTRUTURA METODOLÓGICA DA ANÁLISE	8
3.2. PESQUISA E RECOLHA DE LITERATURA	8
3.2.1. <i>Método</i>	8
3.2.2. <i>Research Questions</i>	9
3.2.3. <i>Estratégia de Pesquisa e Aplicação do Método</i>	10
3.2.4. <i>Critérios de Inclusão e Exclusão</i>	11
3.2.5. <i>Resultados Finais da pesquisa - Artigos Selecionados</i>	13
<b>4. ANÁLISE, RESULTADOS E DISCUSSÃO</b>	<b>15</b>
4.1. ANÁLISE VISUAL	15
4.2. ANÁLISE POR TÓPICO	16
4.2.1. <i>Avaliação do Desempenho Individual</i>	16
4.2.2. <i>Avaliação do Desempenho Coletivo / Tático da Equipe</i>	34
4.2.3. <i>Previsão Tática e Apoio à Decisão</i>	43
<b>5. CONCLUSÃO</b>	<b>52</b>
5.1. SÍNTESE DO TRABALHO DESENVOLVIDO	52
5.2. LIMITAÇÕES	52
5.3. TRABALHO FUTURO	52
REFERÊNCIAS BIBLIOGRÁFICAS	53

# ÍNDICE DE FIGURAS

FIGURA 1 - QUERY DE PESQUISA.....	10
FIGURA 2 - FLUXO TRIAGEM PRISMA .....	11
FIGURA 3 - MAPA BIBLIOMÉTRICO.....	15
FIGURA 4 - ARQUITETURA GERAL DO MODELO PLAYERANK (PAPPALARDO ET AL, 2018).....	17
FIGURA 5 - DISTRIBUIÇÃO DOS CENTROS DE PERFORMANCE POR PAPEL TÁTICO (PAPPALARDO ET AL, 2018).....	18
FIGURA 6 - INTERPRETAÇÃO DOS 8 CLUSTERS DETETADOS E EXEMPLOS DE JOGADORES ATRIBUÍDOS A CADA CLUSTER (PAPPALARDO ET AL, 2018) .....	18
FIGURA 7 - ANÁLISE DA VERSATILIDADE TÁTICA DOS JOGADORES (PAPPALARDO ET AL, 2018).....	18
FIGURA 8 - VALIDAÇÃO DO PLAYERANK COM SCOUTS PROFISSIONAIS (PAPPALARDO ET AL, 2018) .....	19
FIGURA 9 - EXEMPLOS DE SEQUÊNCIAS DE MOVIMENTO ROTULADAS COMO DRIBLE (BARBON ET AL, 2022) .....	20
FIGURA 10 - RESULTADOS DO AGRUPAMENTO HIERÁRQUICO DE AÇÕES SEGUNDO PERFIS DE MOVIMENTO (BARBON ET AL, 2022) ...	21
FIGURA 11 - EXEMPLOS DE AÇÕES DE DRIBLE DOS DOIS GRUPOS IDENTIFICADOS APÓS O CLUSTERING DAS AÇÕES REALIZADAS POR DEFESAS (BARBON ET AL, 2022).....	21
FIGURA 12 - DESEMPENHO DOS CLASSIFICADORES TESTADOS EM FUNÇÃO DO PARÂMETRO (S) E (N), RESPECTIVAMENTE (BARBON ET AL, 2018) .....	22
FIGURA 13 - EXEMPLO DE FALSO POSITIVO DETETADO PELO MODELO (BARBON ET AL, 2018).....	22
FIGURA 14 - DIFERENÇAS NA DISTÂNCIA TOTAL PERCORRIDA ENTRE TODAS AS POSIÇÕES (ANDRZEJEWSKI ET AL, 2015).....	23
FIGURA 15 - RESUMO DOS PASSOS REALIZADOS PARA CALCULAR OS VALORES PSINDIVIDUAL E PSTEAM (MANDORINO ET AL, 2025) 24	
FIGURA 16 - ARQUITETURA GERAL DO SISTEMA (SASEENDRAN ET AL, 2023) .....	25
FIGURA 17 - EXEMPLO DE IMAGEM PARA O TRACKING DE JOGADOR, BOLA E EQUIPA UTILIZANDO DEEPSORT E YOLOV5 (SASEENDRAN ET AL, 2023) .....	26
FIGURA 18 - ARQUITETURA DO MODELO PLS-SEM (CEFIS & CARPITA, 2025).....	27
FIGURA 19 - DISTRIBUIÇÃO NO TERRENO DE JOGO DOS 660 REMATES PROVENIENTES DOS 53 JOGOS DA ÉPOCA 19/20 DA SERIE A ITALIANA (CEFIS & CARPITA, 2025) .....	27
FIGURA 20 - ESTATÍSTICAS E VARIÁVEIS DOS 660 REMATES PROVENIENTES DOS 53 JOGOS DA ÉPOCA 19/20 DA SERIE A ITALIANA (CEFIS & CARPITA, 2025) .....	28
FIGURA 21 - DISTRIBUIÇÕES PSEUDO-R <sup>2</sup> (CEFIS & CARPITA, 2025) .....	28
FIGURA 22 - COEFICIENTES DA REGRESSÃO E ESTIMATIVAS DOS ODDS RATIOS APÓS 5.000 INTERAÇÕES, REFERENTES AOS 53 JOGOS DA ÉPOCA 19/20 DA SERIE A ITALIANA (CEFIS & CARPITA, 2025).....	29
FIGURA 23 - VISUALIZAÇÃO DA ABORDAGEM (BRANSEN & VAN HAAREN, 2019) .....	30
FIGURA 24 - TOP 10 DE JOGADORES MELHORES CLASSIFICADOS QUE JOGARAM PELO MENOS 900 MINUTOS DURANTE A ÉPOCA 17/18 NAS 5 PRINCIPAIS LIGAS EUROPEIAS (BRANSEN & VAN HAAREN, 2019).....	31
FIGURA 25 - GRÁFICOS DE DENSIDADE POR POSIÇÃO QUE MOSTRAM A CORRELAÇÃO ENTRE A PRECISÃO DE PASSE E A CONTRIBUIÇÃO DE PASSE POR 90MIN (BRANSEN & VAN HAAREN, 2019) .....	32
FIGURA 26 - OS CINCO JOGADORES MAIS SEMELHANTES A MANUEL LANZINI (BRANSEN & VAN HAAREN, 2019) .....	32
FIGURA 27 - METODOLOGIA DE SEGMENTAÇÃO TEMPORAL (ANDRIENKO ET AL, 2021) .....	34
FIGURA 28 - SEQUÊNCIAS DE 10 POSIÇÕES DOS JOGADORES E DA BOLA DURANTE A POSSE DE BOLA DO BVB, CORRESPONDENTES A DIFERENTES POSIÇÕES DO CENTRO DA EQUIPA NO TERRENO DE JOGO AO LONGO DO EIXO X (ANDRIENKO ET AL, 2021).....	35
FIGURA 29 - CAMPO DE FUTEBOL DIVIDIDO EM 30 ZONAS (TIANBIAO & ANDREAS, 2016) .....	36
FIGURA 30 - DEFINIÇÃO DOS DADOS UTILIZADOS (TIANBIAO & ANDREAS, 2016).....	37
FIGURA 31 - MODELO DE JOGADOR – 2ª PARTE (TIANBIAO & ANDREAS, 2016) .....	37
FIGURA 32 - MODELO DE JOGADOR – 2ª PARTE (TIANBIAO & ANDREAS, 2016) .....	38
FIGURA 33 - INTERFACE DO MATCH VISION STUDIO PREMIUM (BORGES ET AL, 2019) .....	39
FIGURA 34 - VARIÁVEIS AÇÕES OFENSIVAS E ABORDAGENS DE JOGO (BORGES ET AL, 2019) .....	39

FIGURA 35 - COMPARAÇÃO DOS INDICADORES TÁTICO-TÉCNICOS ENTRE DIFERENTES MÉTODOS OFENSIVOS (BORGES ET AL, 2019)..	40
FIGURA 36 - ANÁLISE DE REDE DE CORRELAÇÕES ENTRE VARIÁVEIS TÁTICAS E TÉCNICAS EM SEQUÊNCIAS OFENSIVAS QUE TERMINARAM COM SUCESSO - GOLO (BORGES ET AL, 2019) .....	40
FIGURA 37 - DADOS FÍSICOS COLETIVOS POR JOGO (MOHAMMED ET AL, 2025) .....	41
FIGURA 38 - DISTÂNCIA PERCORRIDA POR QUATRO JOGADORES-CHAVE AO LOGO DOS JOGOS – FIFA WORLD CUP 2022 (MOHAMMED ET AL, 2025) .....	42
FIGURA 39 - AGRUPAMENTO DOS JOGOS POR PERFIL DE DESEMPENHO - PCA + CLUSTERING (MOHAMMED ET AL, 2025) .....	43
FIGURA 40 - VISÃO GERAL DO PRÉ-PROCESSAMENTO E DO MÉTODO PROPOSTO (SUZUKI ET AL, 2019) .....	44
FIGURA 41 - CAMPO DE JOGO E AS SUAS 15 REGIÕES DIVIDIDAS (SUZUKI ET AL, 2019) .....	45
FIGURA 42 - ESTRUTURA DO MODELO DELM (SUZUKI ET AL, 2019) .....	45
FIGURA 43 - MÉTODO PARA CORRIGIR OS RESULTADOS DA ESTIMATIVA COM BASE NA POSSE DE BOLA (SUZUKI ET AL, 2019) .....	46
FIGURA 44 - COMPARAÇÃO DA ESTIMATIVA DAS TÁTICAS DAS EQUIPAS VS. MÉTODOS COMPARATIVOS (SUZUKI ET AL, 2019) .....	46
FIGURA 45 - CONFUSION MATRIX OBTIDO A PARTIR DOS RESULTADOS DA ESTIMATIVA ANTES DO CÁLCULO DA COMBINAÇÃO ÓTIMA (SUZUKI ET AL, 2019).....	47
FIGURA 46 - CONFUSION MATRIX OBTIDO A PARTIR DOS RESULTADOS DA ESTIMATIVA DEPOIS DO CÁLCULO DA COMBINAÇÃO ÓTIMA (SUZUKI ET AL, 2019).....	47
FIGURA 47 - BOXPLOT DOS RESULTADOS DA PONTUAÇÃO DE SIMILARIDADE (NOURAIE ET AL, 2023) .....	48
FIGURA 48 - VALORES DE MÍNIMO, 1º QUARTIL, MEDIANA, 3º QUARTIL, MÁXIMO E DESVIO-PADRÃO PARA A MÉTRICA ACCORDANCE ENTRE O TREINADOR E O ITFPS (NOURAIE ET AL, 2023) .....	48
FIGURA 49 - VALORES DE MÍNIMO, 1º QUARTIL, MEDIANA, 3º QUARTIL, MÁXIMO E DESVIO-PADRÃO PARA A MÉTRICA RATIO METRIC ENTRE O TREINADOR E O ITFPS (NOURAIE ET AL, 2023) .....	49
FIGURA 50 - REDE DE PASSES DA PRIMEIRA PARTE DO EL CLÁSSICO ENTRE O REAL MADRID (EM CIMA) E O BARCELONA (EM BAIXO) (BERMAN ET AL, 2022) .....	49
FIGURA 51 - SÉRIE TEMPORAL DE TRÊS DAS SEIS MÉTRICAS DE REDE PARA O JOGO BARCELONA VS. REAL MADRID. AS LINHAS TRACEJADAS ASSINALAM OS GOLOS DO BARCELONA (BERMAN ET AL, 2022) .....	50
FIGURA 52 - PRECISÃO DA PREVISÃO DE DESFECHO DO JOGO, DOS TRÊS CLASSIFICADORES, EM DETERMINADOS MOMENTOS DA PARTIDA (BERMAN ET AL, 2022) .....	50
FIGURA 53 - PRECISÃO DA PREVISÃO DE DESFECHO DO JOGO, UTILIZANDO ALGORITMOS BASEADOS EM CARACTERÍSTICAS (LINHAS TRACEJADAS) E OS MESMOS ALGORITMOS, MAS COM MÉTRICAS DE REDE INCLUÍDAS (LINHAS CONTINUAS) (BERMAN ET AL, 2022) .....	50
FIGURA 54 - RELAÇÃO ENTRE OS 6 CLUSTERS (BERMAN ET AL, 2022).....	51

# ÍNDICE DE TABELAS

TABELA 1 - ARTIGOS SELECIONADOS PARA ANÁLISE APÓS TRIAGEM PRISMA.....	13
---	----

# 1. INTRODUÇÃO

O futebol moderno atravessa uma transformação profunda impulsionada pela digitalização e pela crescente disponibilidade de dados observacionais e métricas de performance. Tecnologias como sistemas de tracking, sensores biométricos, plataformas estatísticas avançadas e modelos de machine learning tornaram possível analisar o jogo com um nível de detalhe que seria impensável há apenas uma década (Gudmundsson & Horton, 2017). Esta evolução tem alterado a forma como clubes, treinadores e analistas interpretam o rendimento desportivo, apoiando decisões que vão desde a preparação do treino até à gestão do plantel e estratégias competitivas (Rein & Memmert, 2016).

Apesar deste avanço, a integração eficaz de Data Analytics no futebol continua a representar um desafio significativo. A quantidade massiva de informação gerada diariamente exige metodologias robustas para transformar dados brutos em conhecimento útil e acionável. Para além disso, a interpretação destes indicadores deve ser sempre contextualizada, considerando fatores como o adversário, o modelo de jogo, os momentos da partida e a própria intencionalidade tática por detrás das ações observadas (Memmert et al., 2017). Como vários estudos têm demonstrado, a análise estatística isolada não é suficiente para captar a complexidade de um jogo caracterizado pela imprevisibilidade, pela interação constante entre jogadores e pela influência do contexto competitivo (Rein & Memmert, 2016).

É neste enquadramento que surge a presente dissertação. O objetivo deste trabalho é mapear, compreender e sintetizar a forma como a análise de dados tem sido utilizada para avaliar o desempenho tático no futebol profissional, tanto numa perspetiva individual como coletiva. Através de uma revisão sistemática da literatura, seguindo o protocolo PRISMA (Page et al., 2021), irão ser analisadas metodologias que recorrem a *event data*, *tracking data*, métricas espaciais, algoritmos preditivos e modelos explicativos aplicados ao jogo. O foco principal incidirá sobre três dimensões fundamentais:

- (1) a avaliação do desempenho individual**, explorando contributos técnico-táticos, físicos e comportamentais;
- (2) a análise do desempenho coletivo**, investigando padrões de organização, ocupação de espaço e interações entre jogadores;
- (3) a previsão tática e o apoio à decisão**, onde se estudam modelos capazes de antecipar comportamentos ou recomendar ações durante o jogo.

Este trabalho procura, assim, compreender o estado atual de utilização de Data Analytics no futebol, a sua importância e identificar tendências emergentes, abordagens metodológicas relevantes e oportunidades de investigação. Pretende-se também clarificar como estas

ferramentas podem contribuir para decisões mais informadas, desde o microcontexto (a ação de um jogador), até ao macrocontexto (a estratégia global de uma equipa).

No final, espera-se que esta dissertação contribua para uma visão mais integrada e rigorosa do papel da análise de dados na compreensão do jogo, reforçando a importância da combinação entre tecnologia, conhecimento tático e interpretação humana na evolução do futebol contemporâneo.

## 2. REVISÃO DE LITERATURA

### 2.1 CONTEXTO E IDENTIFICAÇÃO DO PROBLEMA

O futebol assume hoje um papel singular no panorama desportivo global, não apenas pelo número de adeptos e praticantes, mas também pelo impacto económico e social que gera. Este crescimento tem levado a um aumento substancial do investimento por parte de clubes, acionistas e proprietários, potenciando a profissionalização do setor e elevando as exigências de desempenho desportivo. A pressão para atingir resultados, superar expectativas e competir em contextos cada vez mais exigentes tornou-se uma constante no quotidiano dos clubes profissionais (Andreff, 2011).

Neste ambiente altamente competitivo, a tomada de decisão deixou de se centrar exclusivamente na experiência do treinador ou na observação intuitiva, passando a incorporar análises sistemáticas que procuram reduzir a incerteza inerente ao jogo. A performance desportiva, enquanto principal determinante do sucesso dentro e fora de campo, influencia diretamente o acesso a competições de maior prestígio, a obtenção de prémios financeiros mais elevados e a valorização dos ativos do clube. Por isso, grande parte das receitas obtidas em competições, merchandising, bilheteira ou transferências é reinvestida em infraestruturas, equipas técnicas mais qualificadas e também profissionais e departamentos especializados em áreas emergentes, como a análise de dados.

A quantidade de informação gerada diariamente no contexto do futebol moderno é enorme: dados provenientes de treinos, análises de adversários, métricas fisiológicas, registos posicionais, ações de jogo e relatórios de equipas de formação. Transformar estes dados brutos em conhecimento acionável requer equipas técnicas competentes, capazes de interpretar padrões, contextualizar indicadores e apoiar a tomada de decisões com base em evidência (Gudmundsson & Horton, 2017). A análise detalhada de ações de jogo, complementada por técnicas de *feature engineering* e exploração de dados, permite identificar fatores críticos e melhorar a preparação competitiva (Tianbiao & Andreas, 2016).

No entanto, apesar do crescimento do interesse e investimento na área, permanece uma grande opacidade relativamente ao modo como os clubes utilizam os dados no seu processo de decisão. Não é claro de que forma a informação recolhida influencia a preparação dos treinos, a estratégia de jogo, as decisões em tempo real ou mesmo o acompanhamento psicológico dos atletas. Faltam também evidências claras sobre como estas práticas impactam diretamente a performance tática das equipas. A própria comunidade futebolística mantém, ainda hoje, uma relação ambivalente com a análise estatística: por um lado, reconhece o seu potencial, por outro, teme que a excessiva racionalização possa comprometer a autenticidade e a “romantização” tradicional do jogo (Rein & Memmert, 2016).

Esta falta de transparência levanta várias questões relevantes para a investigação e embora vários autores tenham procurado responder parcialmente a estas questões, continuam a

existir lacunas significativas, sobretudo no que diz respeito à realidade portuguesa, onde a utilização de dados para suportar decisões táticas é ainda pouco documentada. Esta ausência de referências reforça a pertinência de estudar de forma mais aprofundada o papel da análise de dados na compreensão do desempenho tático no futebol profissional, clarificando práticas, abordagens e impactos reais deste fenómeno em constante evolução.

## 2.2 MOTIVAÇÕES E OBJETIVOS

A conjugação de dispositivos de *tracking*, plataformas estatísticas e algoritmos preditivos tem-se afirmado como “chave” no futebol moderno. Contudo, subsistem dúvidas sobre como os clubes de futebol utilizam essas ferramentas e acerca do impacto concreto na performance dos clubes nas respetivas equipas de futebol profissional (Rein & Memmert, 2016).

Uma das principais aplicações da análise de dados nesta área está na identificação de padrões táticos. Algoritmos de inteligência artificial e *machine learning* são utilizados para prever comportamentos individuais e coletivos dentro de campo, fornecendo informações importantes para a adaptação de formações e dinâmicas de jogo. Ferramentas de *tracking*, como GPS e sistemas de câmaras inteligentes, permitem recolher dados de movimentação dos jogadores, avaliando não apenas a sua eficiência ao nível individual, mas também o seu impacto na organização coletiva da equipa. A avaliação do desempenho individual dos jogadores tem vindo a ser complementada por métricas avançadas, como o xG (*expected goals*) e xA (*expected assists*), que ajudam a medir a eficácia ofensiva para além das estatísticas tradicionais (Pappalardo et al, 2019).

Assim iremo-nos focar numa análise literária, com base em estudos previamente feitos com foco em análise de dados para precisamente, percebermos de que forma os dados têm sido utilizados para compreender o desempenho tático no futebol profissional. O objetivo é reunir e sintetizar contributos que permitam perceber como diferentes abordagens - desde modelos baseados em *event data* até análises suportadas por *tracking* e técnicas de *machine learning* - têm sido aplicadas para avaliar o comportamento individual dos jogadores, interpretar padrões coletivos e antecipar a dinâmica do jogo.

Ao explorar a literatura existente, pretende-se compreender como estas ferramentas têm sido aplicadas em diferentes contextos competitivos e de treino, quais as métricas mais utilizadas para descrever comportamentos táticos e que modelos têm demonstrado maior capacidade para representar o que acontece dentro de campo. Esta análise permitirá, assim, mapear o estado atual do conhecimento e identificar as abordagens que mais têm contribuído para a evolução da análise tática no futebol moderno.

## **2.3. ENQUADRAMENTO SOBRE O TEMA**

### **2.3.1. FUTEBOL - UMA VISÃO GERAL**

O futebol é, atualmente, o desporto mais popular e globalmente praticado no mundo, sendo considerado por muitos como o “desporto-rei”. A sua origem remonta a práticas antigas de jogos com bola, sendo possível encontrar referências a atividades similares em civilizações como a chinesa, japonesa, egípcia e grega. Um dos exemplos mais antigos é o jogo chinês, *Cuju*, reconhecido pela FIFA como o antecessor mais direto do futebol moderno, praticado já no século III a.C.

Apesar destas raízes remotas, o futebol como o conhecemos hoje teve origem em Inglaterra, durante o século XIX, quando as primeiras regras formais começaram a ser definidas. Em 1863, foi fundada a Football Association (FA), em Londres, responsável pela codificação das regras do jogo, marcando o início da transição do futebol de uma prática recreativa para uma modalidade desportiva organizada. A institucionalização do futebol levou à sua rápida disseminação não só pelo Reino Unido, mas também por outras partes da Europa e, posteriormente, por todo o mundo, através do comércio, da emigração e do colonialismo britânico (Robertson & Giulianotti, 2006).

A criação da FIFA, em 1904, consolidou a internacionalização da modalidade, promovendo a padronização das regras e incentivando a realização de competições internacionais. A primeira edição do Campeonato do Mundo teve lugar em 1930, no Uruguai, e desde então tem-se afirmado como o evento desportivo mais assistido a nível mundial.

Ao longo do século XX, o futebol transformou-se de um fenómeno desportivo para um fenómeno sociocultural e económico de grande escala. Os clubes deixaram de ser apenas entidades recreativas para se tornarem organizações complexas, com objetivos desportivos e comerciais (Andreff, 2011). A profissionalização, o aparecimento de ligas internacionais, os contratos de patrocínio milionários e os direitos televisivos alteraram profundamente a forma como o jogo é jogado, gerido e consumido.

Este crescimento exponencial também impulsionou a necessidade de uma abordagem mais científica ao jogo, nomeadamente no que toca à preparação física, à análise tática e, mais recentemente, à análise de dados. No futebol moderno, a recolha e utilização de grandes volumes de dados tornaram-se parte essencial das estratégias de desempenho, recrutamento e tomada de decisão, tanto dentro como fora das quatro linhas (Rein & Memmert, 2016).

### **2.3.2. CONCEITOS E REGRAS**

O principal objetivo do jogo é marcar golo na baliza adversário. O resultado pode ser vitória, empate ou derrota, conforme o número total de golos. O campo deve ter entre 90 e 120 metros de comprimento e entre 45 e 90 metros de largura, incluindo as delimitações da

grande área, linha de meio-campo, zonas de canto e marca de penálti. Cada equipa é composta por 11 jogadores (10 de campo e um guarda-redes) podendo efetuar substituições segundo as regras de cada competição. A equipa de arbitragem inclui, pelo menos, um árbitro principal e dois assistentes, e em muitos campeonatos há um quarto árbitro e equipas de VAR (Video Assistant Referee). A duração regulamentar é de 90 minutos acrescidos de compensação e em jogos a eliminar pode estender-se a 120 minutos, seguidos de penáltis se persistir o empate.

O treinador toma diversas decisões antes e durante o jogo, cujo impacto é avaliado pelos resultados obtidos. As equipas adaptam-se ao seu grau de favoritismo - clubes menos cotados optam por abordagens mais defensivas para garantir pontos. A preparação inclui decisões fundamentais como Estratégia de jogo (por exemplo, jogo direto, pressão alta), a Formação do 11 inicial (por exemplo, 4-4-2, 4-3-3) e Escolha dos titulares (seleção das 11 iniciais, alinhada com o estilo e a formação), entre muitos outros aspetos mais táticos (Memmert et al., 2017). Em situação de jogo e durante o decorrer do mesmo, o treinador pode alterar esquemas, reforçar linhas ou variar o ritmo, sempre com o objetivo de maximizar as hipóteses de vitória.

### **2.3.3. FOOTBALL ANALYTICS**

O *sport analytics*, mais precisamente o *football analytics* neste caso, consiste na combinação de recolha sistemática de dados, previsão do desenrolar do jogo e técnicas de interpretação estratégica, com o objetivo de melhorar o desempenho dos atletas, tanto a nível individual como o coletivo (Rein & Memmert, 2016). Por exemplo, no futebol, em comparação com o basquetebol, os eventos de golo são muito menos frequentes, pelo que outros momentos do jogo se tornam mais relevantes para avaliar o desempenho - nomeadamente o tempo de posse de bola, as sequências de passe e a análise espacial do território explorado, entre outros.

Através da aplicação de algoritmos de inteligência artificial e *machine learning*, bem como de tecnologias de *tracking* (GPS, câmaras inteligentes), sensores biométricos e plataformas de análise de dados, é possível registar detalhes como o trajeto da bola, velocidades atingidas, métricas fisiológicas (por exemplo, frequência cardíaca) e o posicionamento de cada jogador em qualquer instante do jogo (Gudmundsson & Horton, 2017). Estas informações são integradas em software especializado que, por um lado, apoia a análise pós-partida (avaliação de desempenhos individuais e coletivos) e, por outro, permite ajustes táticos em tempo real, conferindo ao treinador uma visão de conjunto que facilita a tomada de decisões durante o próprio jogo (Memmert et al., 2017).

#### 2.3.4. DISPOSITIVOS E SOFTWARES

A evolução do *football analytics* tem sido fortemente impulsionada pelo desenvolvimento de dispositivos de monitorização e softwares especializados que permitem recolher, processar e analisar grandes volumes de dados sobre o desempenho dos jogadores e das equipas. Atualmente, clubes profissionais em todo o mundo recorrem a uma variedade de ferramentas tecnológicas para melhorar a sua performance competitiva, apoiar decisões táticas e otimizar processos de recrutamento e prevenção de lesões.

Um dos dispositivos mais utilizados atualmente é o STATSports Apex, um sistema de rastreamento GPS usado por muitas seleções nacionais e clubes de topo como Liverpool FC, Juventus ou Manchester United. Este dispositivo vestível, em formato de colete, fornece métricas em tempo real como distância percorrida, velocidade máxima, número de sprints, carga de treino e dados fisiológicos que permitem avaliar a fadiga e o risco de lesão dos atletas (Cummins et al., 2013).

Além dos dispositivos GPS, softwares como o Wyscout, Instat, StatsBomb, Opta Sports e Second Spectrum tornaram-se fundamentais para análise tática e estatística. O Wyscout, por exemplo, é utilizado para *scouting* e análise de jogos, permitindo acesso a milhares de vídeos, estatísticas detalhadas de jogadores e equipas, e relatórios automáticos baseados em eventos de jogo. Já o StatsBomb destaca-se pelo seu detalhe nas métricas de desempenho, incluindo ações como “pressões bem-sucedidas”, “*pass breaklines*” ou “*expected threat*” (Pappalardo et al., 2019).

O Second Spectrum, por seu lado, é líder em *tracking data* ao utilizar visão computacional e inteligência artificial para gerar visualizações em tempo real, mapas de calor e reconstruções táticas 3D, sendo utilizado pela Premier League e pela MLS.

Estas plataformas não só melhoram o planeamento do treino e a análise do adversário, como também têm ganho um papel central na tomada de decisões estratégicas no mercado de transferências. A combinação de *tracking data*, *event data* e algoritmos de *machine learning* permite construir modelos preditivos robustos, contribuindo para uma abordagem mais objetiva e cientificamente fundamentada ao futebol moderno (Rein & Memmert, 2016).

## 3. METODOLOGIA

### 3.1. ESTRUTURA METODOLÓGICA DA ANÁLISE

A metodologia adotada neste trabalho segue uma abordagem sistemática que visa garantir rigor, transparência e replicabilidade no processo de recolha, seleção e análise dos estudos científicos. Tendo como objetivo compreender como a análise de dados tem sido utilizada para avaliar o desempenho tático no futebol profissional, optou-se por desenvolver uma revisão sistemática da literatura assente no protocolo PRISMA, amplamente reconhecido pela sua aplicabilidade em processos de seleção de evidência científica.

O percurso metodológico foi estruturado em duas fases principais. Na primeira fase, procedeu-se à identificação, recolha e triagem dos estudos relevantes, utilizando critérios de inclusão e exclusão previamente definidos. Esta etapa permitiu garantir que apenas artigos com contributo direto para a análise tática fossem considerados.

Na segunda fase, os estudos selecionados foram analisados em profundidade, permitindo organizar os contributos da literatura em três tópicos, já apresentados anteriormente, que estruturam o capítulo de resultados: (1) avaliação de desempenho individual, (2) avaliação do desempenho coletivo e tático da equipa, (3) previsão de táticas e estilos de jogo e apoio à decisão tática. Esta organização possibilita uma leitura coerente da evolução da análise tática no futebol, articulando diferentes metodologias e modelos utilizados na investigação científica.

### 3.2. PESQUISA E RECOLHA DE LITERATURA

#### 3.2.1. MÉTODO

Para identificar os estudos com maior relevância para a presente investigação, recorreu-se ao método PRISMA, composto pelas etapas fundamentais de identificação, triagem, elegibilidade e inclusão. Este processo permitiu garantir uma seleção rigorosa e transparente, assegurando que apenas os artigos com contributo direto para a análise do desempenho tático fossem considerados.

A primeira etapa consistiu na **(1) identificação** de artigos em bases de dados científicas, a partir de um conjunto alargado de palavras-chave relacionadas com *football analytics*, *tactical performance*, *tracking data* e outras expressões associadas à avaliação tática no futebol. Seguiu-se a fase de **(2) triagem**, onde títulos e resumos foram analisados de forma preliminar, permitindo excluir estudos sem relação com o foco da investigação ou que não apresentassem dados empíricos relevantes. Na etapa de **(3) elegibilidade**, procedeu-se à leitura integral dos artigos que cumpriam os critérios iniciais, sendo aplicados filtros adicionais relacionados com qualidade metodológica, pertinência conceptual e clareza da ligação entre

as metodologias analíticas e a avaliação do desempenho tático. Por fim, a fase de **(4) inclusão** resultou na definição do conjunto final de estudos que sustentam a análise desenvolvida neste trabalho.

Para reforçar a consistência do processo, este método foi complementado com uma análise bibliométrica, que permitiu identificar os termos mais recorrentes, a frequência das abordagens analíticas e a convergência temática entre os estudos selecionados. Esta etapa ajudou a validar a coerência da amostra final com o objetivo da investigação.

De referir que, mais adiante serão apresentados de forma detalhada, o fluxo de triagem dos estudos segundo cada fase do PRISMA, os critérios de inclusão e exclusão aplicados, e o mapa bibliométrico construído a partir dos termos predominantes na literatura analisada. Estes elementos complementares permitem assegurar transparência no processo metodológico e reforçar a robustez da seleção final dos artigos.

### **3.2.2. RESEARCH QUESTIONS**

A definição das *Research Questions* assume um papel determinante na orientação metodológica deste trabalho, funcionando como eixo central para toda a estratégia de pesquisa, triagem e seleção dos estudos incluídos nesta análise. No âmbito do protocolo PRISMA, estas questões permitem delimitar com precisão o objeto de estudo, clarificar o tipo de evidência procurada e assegurar que cada fase do processo - identificação, triagem, elegibilidade e inclusão - se mantém alinhada com os objetivos do presente trabalho.

Assim, as *Research Questions* definidas foram fundamentais em termos de guia de critérios para inclusão e exclusão, servindo como guia para determinar que estudos seriam considerados relevantes para a investigação. Ao longo da triagem PRISMA, cada artigo foi avaliado tendo em conta a sua capacidade de contribuir para a resposta a pelo menos uma das questões de investigação, garantindo assim a coerência temática da amostra final.

As *Research Questions* definidas são as seguintes:

**RQ1:** Como tem sido utilizada a análise de dados para avaliar o desempenho tático de jogadores e equipas no futebol profissional?

**RQ2:** Que modelos, frameworks e metodologias têm sido desenvolvidos para interpretar padrões táticos, antecipar comportamentos de jogo e apoiar decisões estratégicas?

**RQ3:** De que forma os dados de tracking - como GPS, event data e positional data - contribuem para a melhoria da tomada de decisão tática em contexto competitivo?

### 3.2.3. ESTRATÉGIA DE PESQUISA E APLICAÇÃO DO MÉTODO

A pesquisa foi realizada exclusivamente na base de dados Scopus, por ser reconhecida como a plataforma científica mais completa e abrangente, reunindo publicações de referência nas áreas da ciência do esporte, ciência de dados, engenharia, ciências sociais e tecnologias aplicadas ao futebol. Esta opção permitiu assegurar que a pesquisa inicial cobria de forma consistente os estudos mais relevantes no domínio da análise de desempenho tático.

A pesquisa bibliográfica foi conduzida através de uma *query* construída para capturar estudos relacionados com futebol, desempenho tático, análise posicional e técnicas de data analytics aplicadas ao jogo.

```
( "football" OR "soccer" ) AND ( "tactical performance" OR "tactical analysis"  
OR tactic OR "team tactics" OR "team performance" OR "player performance" )  
AND ( "tracking data" OR "positional data" OR "spatiotemporal data" OR  
"event data" OR "match data" OR "performance analysis" OR "match analysis"  
OR "video analysis" OR "machine learning" OR "deep learning" OR "data  
analytics" )
```

Figura 1 - Query de pesquisa

Após a submissão da *query* de pesquisa, foram inicialmente identificados 1.068 artigos. Aplicando o filtro temporal (artigos publicados entre 2015 e 2025), de modo a garantir atualidade científica e tecnológica, o total foi reduzido para 969 estudos. De seguida, ao restringir para os tipos de documento mais adequados à investigação académica — artigos científicos e *conference papers* — o número diminuiu para 873. A aplicação do filtro de idioma (inglês ou português) que garantiu a possibilidade de leitura crítica e rigorosa dos trabalhos, resultou em 860 artigos.

Numa fase seguinte, foi aplicado um filtro por áreas científicas relevantes (*Computer Science, Engineering, Social Sciences e Decision Sciences*), eliminando publicações sem ligação direta às metodologias ou ao enquadramento pretendido. Este filtro reduziu o total para 435 artigos, número que se manteve após a remoção de duplicados. A triagem por título constituiu o primeiro momento de exclusão temática, permitindo identificar estudos claramente fora do âmbito da análise tática, nesta fase, permaneceram 130. A análise dos resumos representou a etapa mais importante da triagem, orientada pelos critérios de inclusão e exclusão definidos previamente, deste processo resultaram 76 artigos com potencial para leitura integral.

Finalmente, após a leitura completa dos mesmos, ficamos com 15 artigos finais que servirão de base para a análise de resultados e discussão desta tese.

Este fluxo de triagem PRISMA, detalhado no diagrama apresentado, permitiu assegurar que a seleção dos artigos foi conduzida com rigor, garantindo que apenas os estudos diretamente relevantes para a compreensão do desempenho tático e das metodologias analíticas aplicadas ao futebol profissional fossem incluídos na revisão.

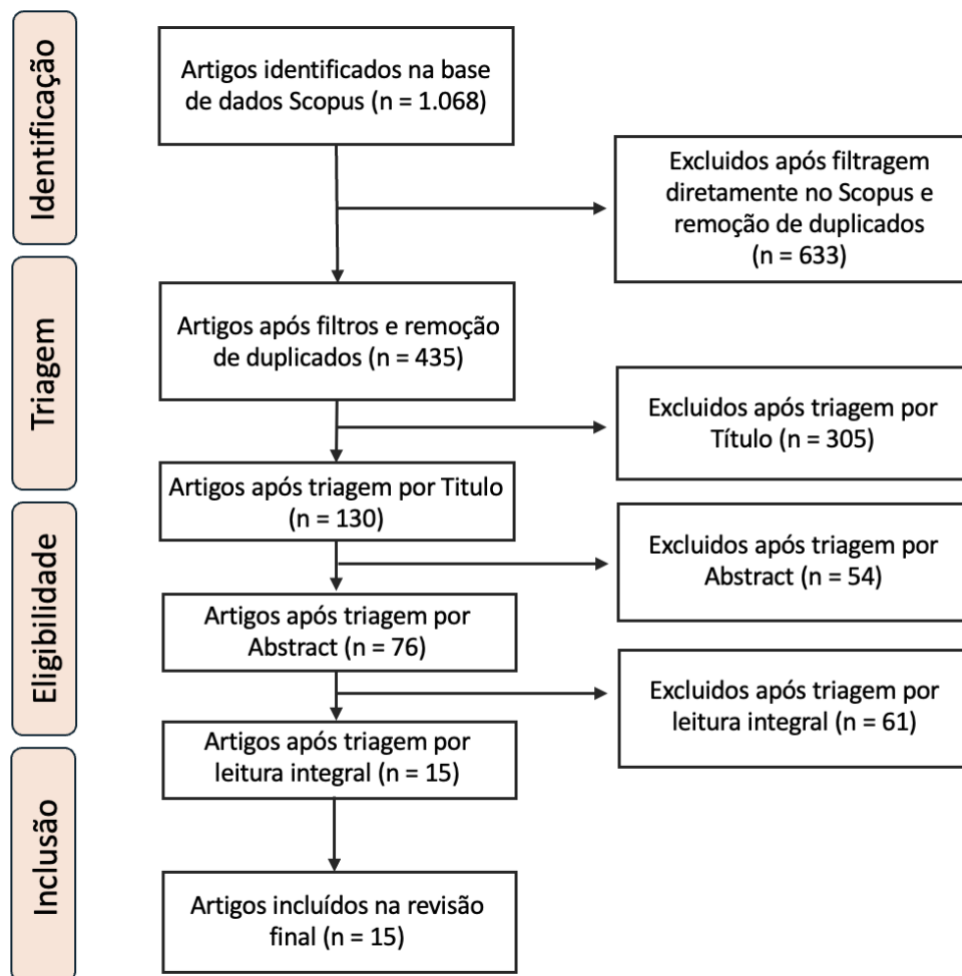


Figura 2 - Fluxo Triagem PRISMA

### 3.2.4. CRITÉRIOS DE INCLUSÃO E EXCLUSÃO

Complementando, a definição dos critérios de inclusão e exclusão constituiu um passo fundamental para garantir a consistência e a relevância científica da revisão sistemática. Estes critérios foram estabelecidos previamente, em alinhamento com as Research Questions e com as recomendações PRISMA para processos de triagem rigorosos e reprodutíveis, permitindo assegurar que apenas os estudos diretamente relacionados com o desempenho tático no futebol fossem considerados ao longo da análise.

No que diz respeito aos **critérios de inclusão**, privilegiaram-se estudos centrados explicitamente na modalidade futebol - considerando como equivalentes os termos *football*

e *soccer* - de modo a eliminar ambiguidades e evitar a incorporação de trabalhos oriundos de outras modalidades. Para além desta condição fundamental, foram incluídos todos os estudos cujo foco temático estivesse claramente relacionado com a performance individual ou coletiva, ou que explorassem dimensões táticas relevantes, como sistemas de jogo, relações posicionais, padrões espaciais, comportamento coletivo em momentos específicos do jogo ou métricas derivadas de *tracking* e *event data*. Este enquadramento permitiu integrar abordagens variadas, desde métricas físicas e técnicas até análises espaciais e comportamentais, desde que inseridas no contexto da dinâmica tática de uma equipa ou do desempenho individual do jogador.

Foi igualmente considerado como critério de inclusão o recurso a dados empíricos - provenientes de tracking, positional data, event data, GPS ou vídeo - ou a metodologias quantitativas aplicadas à análise do jogo, incluindo modelação estatística, machine learning ou técnicas de análise computacional. Em situações em que o resumo apresentava relevância potencial, ainda que não explicitasse de forma totalmente clara a componente tática, optou-se por incluir o estudo para leitura integral. Esta decisão metodológica pretendeu minimizar o risco de exclusão prematura de artigos que, embora ambíguos numa fase inicial, poderiam revelar contributos significativos após análise mais aprofundada.

Em contrapartida, os **critérios de exclusão** foram aplicados para eliminar estudos que, apesar de surgirem nos resultados da pesquisa, não apresentavam ligação direta com o objetivo desta investigação. Foram excluídos todos os trabalhos centrados noutras modalidades desportivas - como basquetebol, voleibol, rugby, cricket, NFL ou e-sports. Também foram excluídos estudos cujo tema central se afastava da dimensão da performance ou da tática, nomeadamente investigações sobre mercado de transferências, valoração económica, estratégia empresarial, previsões de resultados sem ligação ao comportamento tático, análises financeiras ou modelos teóricos sem aplicação concreta ao jogo. Foram igualmente excluídos trabalhos de carácter tecnológico orientados apenas para o desenvolvimento de sensores, hardware ou algoritmos de deteção visual, sempre que estes não resultavam na construção de métricas ou na análise efetiva do rendimento tático.

A adoção destes critérios seguiu as orientações PRISMA, que recomendam a utilização de processos de triagem objetivos, replicáveis e alinhados com o propósito da investigação. A aplicação consistente destes parâmetros permitiu reduzir significativamente o universo inicial de publicações, assegurando que a amostra final fosse constituída apenas por estudos com contributo efetivo para a compreensão da performance tática no futebol profissional, dentro dos três tópicos de estudo definidos.

### 3.2.5. RESULTADOS FINAIS DA PESQUISA - ARTIGOS SELECIONADOS

A tabela seguinte reúne os quinze artigos que foram selecionados para análise, organizando para cada um a referência, o título, a aplicação prática, os métodos utilizados e o tópico em que se enquadra. Esta estrutura permite ter uma visão clara do tipo de contributos que cada estudo traz e facilita a comparação entre abordagens.

Os artigos foram distribuídos por quatro tópicos - **(2.1) avaliação de desempenho individual, (2.2) avaliação do desempenho coletivo/tático da equipa e (2.3) previsão tática e apoio à decisão** - o que ajuda a perceber desde já as diferentes perspetivas presentes na literatura. Esta classificação será útil mais à frente, na análise dos resultados, onde cada grupo será explorado de forma mais detalhada.

Tabela 1 - Artigos selecionados para análise após triagem PRISMA

Artigo	Referência	Título	Aplicação	Métodos / Técnicas	Tópico
A1	Andrzejewski et al. (2015)	Sprinting Activities and Distance Covered by Top-Level Europa League Soccer Players	Avaliação da performance física individual	Estatística descritiva	2.1
A2	Bransen & van Haaren (2019)	Measuring Football Players	Quantificação do contributo individual em jogo	Modelação quantitativa	2.1
A3	Cefis & Carpita (2025)	A New xG Model for Football Analytics	Avaliação da qualidade dos remates individuais	Modelo xG probabilístico	2.1
A4	Saseendran et al. (2023)	Analysis of Player Tracking Data Extracted from Match Feed	Análise do impacto individual através de métricas espaciais	Pitch Control, Expected Threat	2.1
A5	Pappalardo et al. (2019)	PlayeRank: Data-driven performance evaluation and player ranking in soccer	Classificação e avaliação de desempenho individual	ML supervisionado; Feature weighting	2.1
A6	Barbon Junior et al. (2022)	Sport action mining: Automatic dribbling recognition	Deteção automática de ações técnicas individuais	Pattern Mining; Clustering	2.1

Artigo	Referência	Título	Aplicação	Métodos / Técnicas	Tópico
A7	Mandorino et al. (2025)	The Interaction of Fitness and Fatigue on Physical and Tactical Performance in Football	Monitorização do estado físico individual e impacto no jogo	Random Forest; Fitness–fatigue model	2.1
A8	Tianbiao & Andreas (2016)	Apriori-Based Diagnostical Analysis of Passings in the Football Game	Identificação de padrões coletivos de passe	Apriori; Data Mining	2.2
A9	Andrienko et al. (2021)	Constructing Spaces and Times for Tactical Analysis in Football	Análise tática coletiva baseada em ocupação de espaço	Space–Time Analysis; Clustering	2.2
A10	Mohammed et al. (2025)	Tactical and Physical Profiling of the Moroccan National Team at FIFA WC 2022	Caracterização tático-física de uma equipa	Análise tático-física	2.2
A11	Borges et al. (2019)	Tactical Efficacy and Offensive Game Processes in Italian and Brazilian Youth Players	Avaliação da eficácia dos processos ofensivos coletivos	FUT-SAT; Análise tática ofensiva	2.2
A12	Lee & Jung (2022)	DNN-Based Multi-Output Model for Predicting Soccer Team Tactics	Previsão automatizada de táticas de equipa	Deep Neural Network	2.3
A13	Suzuki et al. (2018)	Team Tactics Estimation in Soccer Videos Based on a DELM	Estimação automática de padrões táticos	DELM (Deep Extreme Learning Machine)	2.3
A14	Nouraie et al. (2023)	Intelligent Team Formation and Player Selection Using Deep Learning	Seleção otimizada de jogadores e formação da equipa	DNN + Hungarian Algorithm	2.3
A15	Berman et al. (2022)	Temporal Match Analysis & Recommending Substitutions in Live Soccer Games (NMELSA)	Recomendação de substituições em tempo real	Network Metrics; Clustering; XGBoost; FDNN	2.3



processos táticos, mostrando que estes temas continuam a ser o núcleo conceptual onde grande parte da literatura se concentra.

O segundo polo ganha forma em torno de termos como *“machine learning”*, *“deep learning”*, *“data mining”* e *“artificial intelligence”*. Aqui percebe-se claramente a influência crescente das abordagens computacionais e dos modelos preditivos. Este grupo está fortemente ligado a conceitos como *“forecasting”*, *“learning algorithms”* e *“computer vision”*, sugerindo que a investigação mais recente tem procurado integrar técnicas avançadas para interpretar padrões de jogo, prever comportamentos e apoiar a tomada de decisão.

Entre estes dois polos existe uma zona de grande densidade, onde aparecem termos como *“sports analytics”*, *“team performance”* e *“decision making”*, funcionando quase como uma ponte entre a análise futebolística mais clássica e as metodologias computacionais. A presença desta zona intermédia mostra que a área se tem tornado cada vez mais multidisciplinar, cruzando o futebol com ciência de dados, estatística e inteligência artificial.

De um modo geral, o mapa confirma a evolução natural do campo: do foco inicial na análise descritiva e tática, para abordagens mais robustas apoiadas em modelos de machine learning e sistemas de previsão. Ao mesmo tempo, evidencia que a investigação atual não se limita à recolha de dados, mas procura integrar estes métodos numa compreensão mais completa do jogo, seja na avaliação do desempenho, na identificação de padrões ou na antecipação de decisões.

## **4.2. ANÁLISE POR TÓPICO**

### **4.2.1. AVALIAÇÃO DO DESEMPENHO INDIVIDUAL**

A literatura contemporânea dedicada à avaliação de desempenho individual no futebol tem evoluído de forma significativa, impulsionada pelo acesso crescente a dados de eventos, dados posicionais, métricas fisiológicas e, mais recentemente, abordagens híbridas que combinam várias fontes num mesmo modelo analítico. Os estudos analisados refletem uma tendência clara: a performance de um jogador deve ser compreendida como um fenómeno multidimensional, onde ações técnicas, interações táticas, capacidades físicas e padrões comportamentais se articulam para produzir o rendimento observado em competição.

Entre os contributos mais relevantes da literatura recente sobre avaliação de desempenho individual destaca-se o PlayeRank, desenvolvido por Pappalardo et al. (2019), que introduz uma estrutura metodológica particularmente robusta para analisar a performance de um jogador de forma multidimensional e dependente do seu papel em campo. O modelo parte da constatação de que as métricas tradicionalmente utilizadas — como golos, passes concluídos ou desarmes — são insuficientes para captar a riqueza e a variedade de ações que caracterizam o futebol contemporâneo. Em vez disso, o PlayeRank constrói a avaliação do

jogador como um vetor composto por dezenas de variáveis técnico-táticas, às quais são atribuídos pesos aprendidos automaticamente a partir de dados de grandes competições.

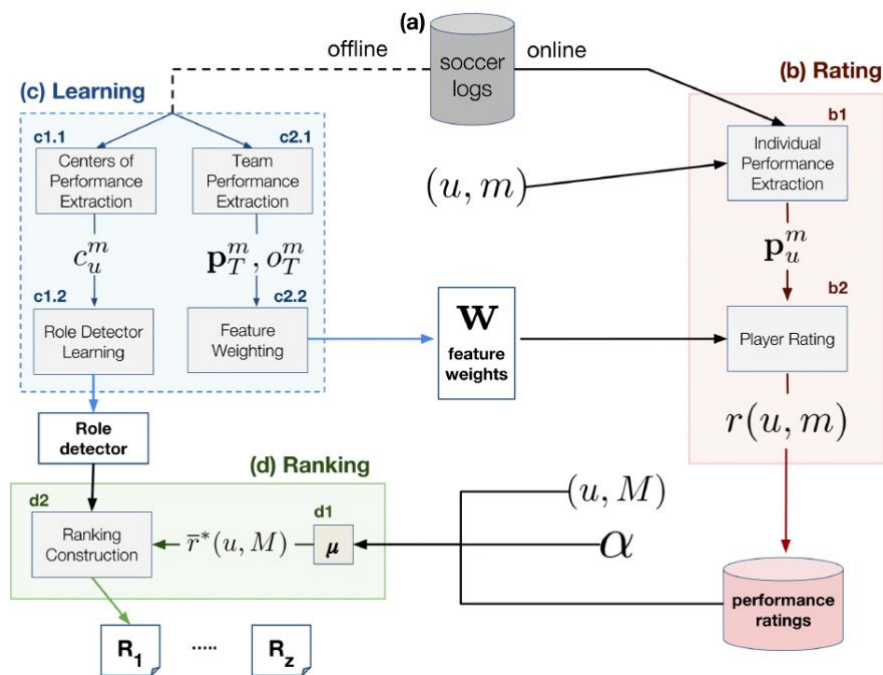


Figura 4 - Arquitetura Geral do Modelo PlayeRank (Pappalardo et al, 2018)

A arquitetura do modelo assenta em três componentes centrais. Em primeiro lugar, os autores recorrem a um processo de aprendizagem supervisionada para determinar a importância relativa de cada ação registada nos *soccer-logs*. Esta fase, mostra que ações muitas vezes subvalorizadas — como recuperações em zonas altas, passes progressivos ou intervenções defensivas no último terço — assumem um peso estatístico mais elevado do que indicadores tradicionais amplamente difundidos na análise futebolística. Esta redistribuição da importância das ações revela a vantagem de utilizar abordagens centradas em dados, capazes de identificar padrões de rendimento que nem sempre são evidentes à observação.

Numa segunda fase, o PlayeRank identifica automaticamente os papéis táticos desempenhados pelos jogadores ao longo dos jogos, através de algoritmos de clustering aplicados às suas posições médias em campo. O gráfico que ilustra esta etapa revela oito funções bem definidas, desde laterais e centrais até médios ofensivos e extremos, permitindo que a avaliação final seja contextualizada em função da responsabilidade tática de cada atleta.

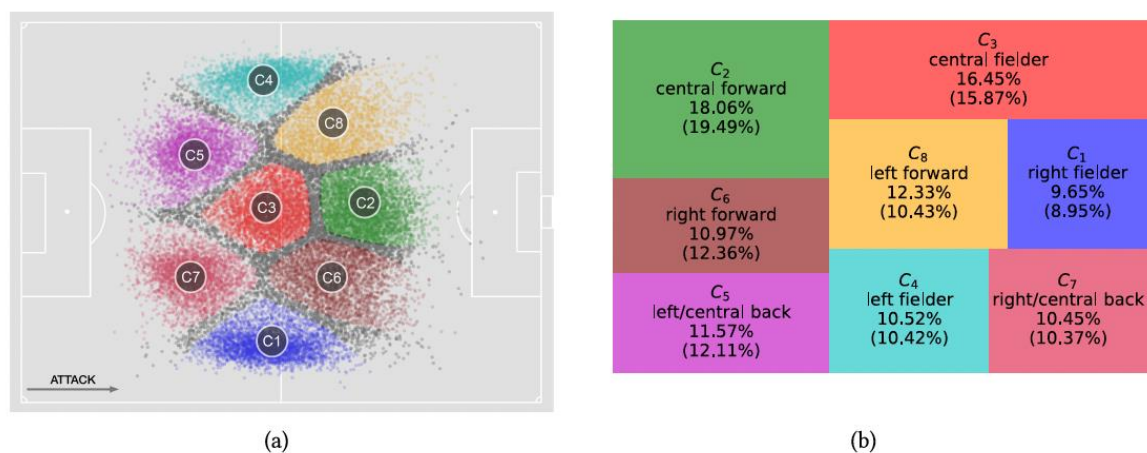


Figura 5 - Distribuição dos centros de performance por papel tático (Pappalardo et al, 2018)

Esta dimensão é particularmente relevante porque evita comparações inadequadas entre jogadores que ocupam zonas distintas do terreno e que, naturalmente, apresentam perfis estatísticos incomparáveis. Avaliar um médio-defensivo e um extremo com as mesmas métricas seria metodologicamente incorreto, e é precisamente esse erro que o PlayeRank procura eliminar.

cluster	description	examples	
C1	right fielder	plays on the right as a wing, back, or both	S. Roberto, Danilo
C2	central forward	plays in the center, close to the opponent's goal	Messi, Suárez
C3	central fielder	plays in the center	Kroos, Pjanić
C4	left fielder	plays on the left as a wing, back, or both	Nolito, J. Alba
C5	left central back	plays close to his own goal, on the left	Bartra, Maguire
C6	right forward	plays on the right, close to the opponent's area	Robben, Dembélé
C7	right central back	plays close to his own goal, on the right	J. Martínez, Matip
C8	left forward	plays on the left, close to the opponent's area	Neymar, Insigne

Figura 6 - Interpretação dos 8 clusters detetados e exemplos de jogadores atribuídos a cada cluster (Pappalardo et al, 2018)

A terceira componente do modelo consiste na combinação dos pesos aprendidos com o vetor de ações executadas pelo jogador em cada jogo, produzindo uma pontuação final de desempenho. Esta abordagem transforma cada participação individual numa representação numérica multifacetada, permitindo comparar jogadores dentro do mesmo papel com rigor e consistência. Um dos resultados mais interessantes desta etapa é a análise da versatilidade, calculada com base na entropia dos papéis desempenhados por cada jogador ao longo da época. Os autores ilustram claramente estes padrões com exemplos como Sergi Roberto e Neymar, cujos centros de performance revelam graus de variabilidade bastante distintos

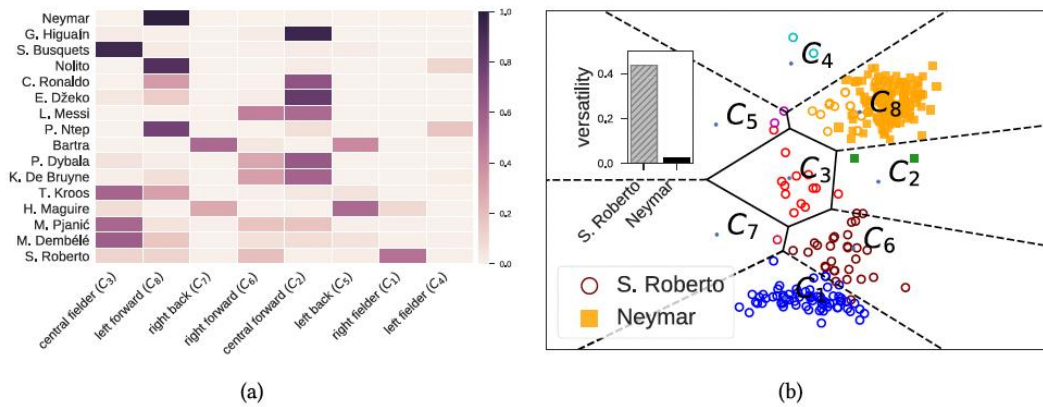


Figura 7 - Análise da versatilidade tática dos jogadores (Pappalardo et al, 2018)

Outro aspeto que reforça a credibilidade do PlayeRank é a validação realizada com um painel de *scouts* profissionais. Num gráfico particularmente elucidativo, observa-se que o modelo atinge níveis elevados de concordância com a avaliação humana, sobretudo quando a diferença entre dois jogadores comparados é significativa. Esta validação externa é crucial para demonstrar que o sistema não apenas replica dados, mas produz classificações coerentes com o conhecimento especializado utilizado na prática profissional.

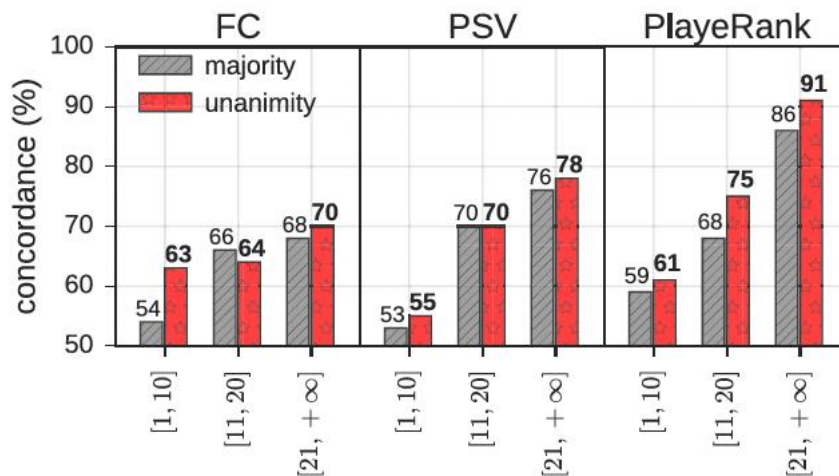


Figura 8 - Validação do PlayeRank com scouts profissionais (Pappalardo et al, 2018)

Apesar da sua robustez conceptual e metodológica, o PlayeRank apresenta algumas limitações que importa referir. Por depender exclusivamente de dados de eventos, o modelo não integra o contexto espacial em que as ações ocorrem, ignorando movimentos sem bola, linhas de passe criadas, posicionamento relativo ou pressão adversária - elementos fundamentais para interpretar ações individuais num enquadramento verdadeiramente tático. Além disso, a ausência de dados fisiológicos ou de tracking reduz a capacidade do

modelo para captar intenções, comportamentos preventivos ou dinâmicas defensivas que não resultam necessariamente num evento registado. Estas limitações não diminuem a utilidade do PlayeRank, mas evidenciam a necessidade de complementar este tipo de abordagem com modelos híbridos que integrem informação posicional, fisiológica e contextual.

Ainda assim, o PlayeRank permanece uma das propostas mais completas e replicáveis disponíveis na literatura existente, oferecendo uma estrutura sólida para comparar jogadores de forma sistemática e contextualizada. Ao combinar escalabilidade — com milhões de eventos analisados — coerência tática e validação externa, o modelo constitui uma referência incontornável na avaliação de desempenho individual e um ponto de partida fundamental para metodologias mais avançadas que pretendam integrar múltiplas dimensões do jogo.

Numa perspetiva complementar, o estudo de Barbon et al. (2022) propõe uma abordagem inovadora ao focar-se exclusivamente na deteção automática de dribles a partir de dados posicionais. Esta perspetiva contrasta com modelos baseados em eventos ou métricas agregadas, ao assumir que o drible pode ser descrito como um padrão de movimento codificado por probabilidades de regras de associação. O método — Sports Action Mining (SAM) — é operacionalizado através de uma pipeline que transforma deslocamentos em *traces*, gera regras frequentes e constrói vetores de confiança que alimentam algoritmos de classificação

A criação do *dataset*, anotado por especialistas e composto por 1 966 eventos de drible e não-drible, é acompanhada por exemplos visuais que ilustram o comportamento típico da ação, com mudanças de direção, aproximação ao adversário e situações de um para um, permitindo observar a complexidade da tarefa.

Match	Championship	Year	Number of Players					Dribble	Non-Dribble
			DF	FN	DF MD	MD	FW		
Team A × Team B	Brazilian Series A	2008	7	4	5	4	6	144	141
Team A × Team C	Brazilian Series A	2008	7	5	4	5	5	73	129
Team A × Team D	Brazilian Series A	2008	6	4	5	5	5	32	140
Team E × Team F	Brazilian Series A	2016	4	5	3	8	6	57	133
Team E × Team G	Brazilian Series A	2016	4	4	4	10	4	86	148
Team E × Team H	Brazilian Series A	2016	4	4	4	9	5	72	126
Team E × Team I	Brazilian Series A	2016	5	5	3	10	3	96	142
Team J × Team K	São Paulo State 2nd division	2014	4	5	4	6	7	113	128
Team J × Team L	São Paulo State 2nd division	2014	4	5	4	6	6	62	144
Total								735	1,231

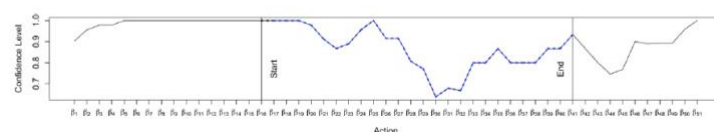


Figura 9 - Exemplos de sequências de movimento rotuladas como drible (Barbon et al, 2022)

Estes exemplos evidenciam uma das dificuldades centrais do problema: distinguir o drible de ações bio-mecanicamente similares, como a condução ou corrida em progressão.

A análise não supervisionada revela que diferentes estilos de drible podem ser identificados através de *clustering* hierárquico, tanto em defesas como em avançados.

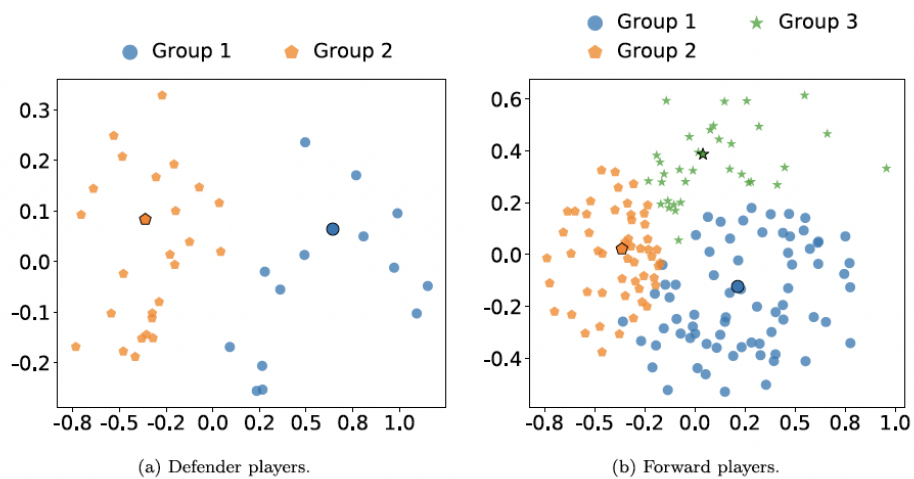


Figura 10 - Resultados do agrupamento hierárquico de ações segundo perfis de movimento (Barbon et al, 2022)

Exemplos adicionais mostram pequenas variações técnicas - mudanças suaves versus abruptas de direção, confrontos diretos ou tentativas falhadas de interceção - demonstrando que o modelo capta nuances relevantes da ação.

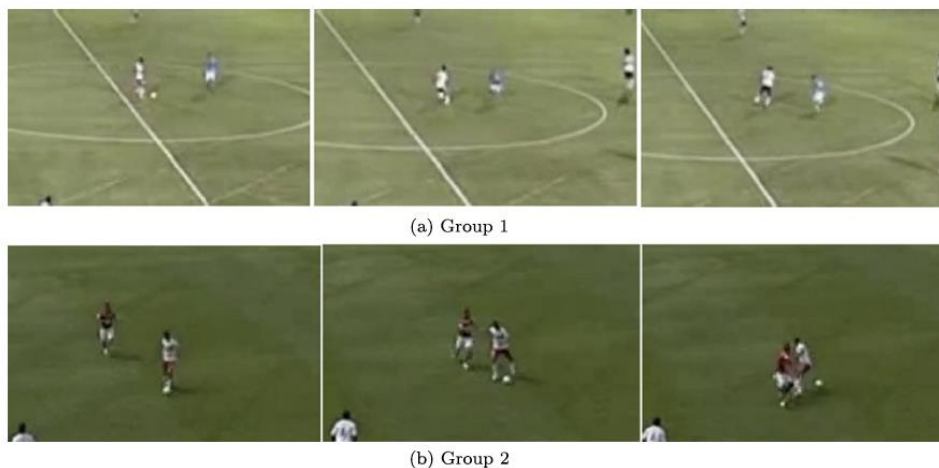


Figura 11 - Exemplos de ações de drible dos dois grupos identificados após o clustering das ações realizadas por defesas (Barbon et al, 2022)

Na vertente supervisionada, os autores testam quatro classificadores e mostram que o desempenho depende fortemente da parametrização temporal ( $s$  e  $n$ ), evidenciado nos

gráficos de *balanced accuracy*. O Random Forest atinge uma precisão de 93.3%, destacando-se como o modelo mais eficaz para a detecção automática de dribles.

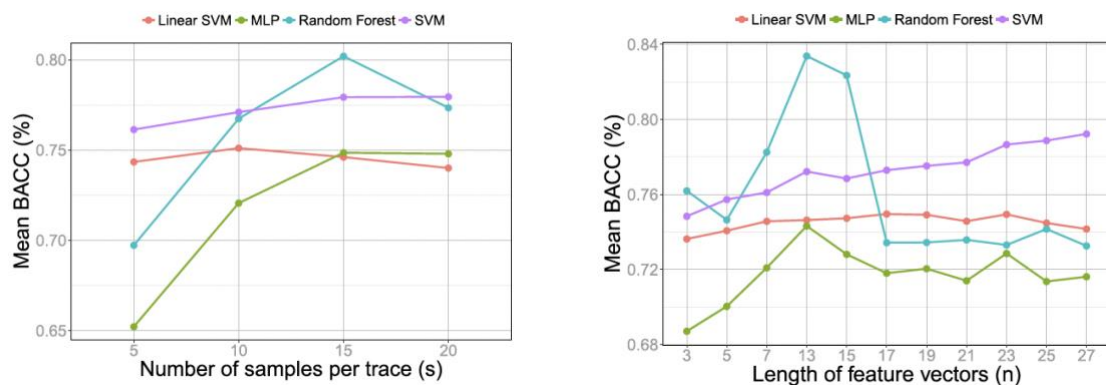


Figura 12 - Desempenho dos classificadores testados em função do parâmetro (s) e (n), respectivamente (Barbon et al, 2018)

O artigo inclui ainda exemplos de classificações incorretas, como falsos positivos em ações semelhantes à condução, sublinhando limitações inerentes ao método.



Figura 13 - Exemplo de falso positivo detetado pelo modelo (Barbon et al, 2018)

Do ponto de vista crítico, o SAM representa um avanço claro ao demonstrar que dados posicionais minimalistas são suficientes para caracterizar ações complexas, oferecendo potencial para scouting automático, análise de perfis técnicos e detecção de talento. No entanto, o modelo continua limitado pela ausência de contexto tático - como pressão, posição da bola ou dinâmica coletiva - o que impede uma avaliação mais completa da eficácia do drible em termos de impacto no jogo. Esta limitação sugere que futuras abordagens deverão integrar dados multimodais para enriquecer a interpretação.

Já a dimensão física surge com grande relevância nos estudos de Andrzejewski et al. (2015), que analisam atividades de sprint em jogos da Europa League recorrendo ao sistema Amisco Pro®. Segundo os autores, a análise de 147 jogadores permite perceber diferenças claras

entre posições: médios-centro percorrem maiores distâncias totais, enquanto extremos e avançados acumulam a maioria dos sprints de alta intensidade.

Uma das conclusões mais relevantes identificadas por Andrzejewski et al. (2015) envolve a distância média dos sprints, que raramente excede os 20 metros, contrariando práticas de treino ainda frequentes. Esta conclusão, ilustrada no estudo através de um gráfico específico, leva os autores a defender que os métodos de treino devem alinhar-se mais com o comportamento real observado em competição.

		ED (n=35)	CD (n=39)	CM (n=35)	EM (n=20)	F (n=18)	Total (n=147)
Covered distance, m	I half	5 463.5	5 082.7	5 819.4	5 920.4	5 508.0	5 514.8
		±445.39	±271.25	±502.45	±413.92	±336.23	±504.86
	II half	5 599.8	5 253.2	5 941.0	5 825.1	5 431.8	5 599.2
		±407.66	±290.83	±384.39	±368.59	±359.33	±445.26
	Total	11 063.2	10 335.9	11 760.4	11 745.7	10 939.8	11 114.0
		±790.71	±470.77	±797.27	±689.98	±648.18	±884.91
Sprinting distance, m	I half	124.0	91.8	82.8	163.1	180.3	117.9
		±68.83	±38.10	±44.07	±82.57	±86.81	±70.32
	II half	140.8	94.1	84.5	150.8	165.7	119.4
		±78.28	±58.33	±51.94	±68.90	±77.35	±72.31
	Total	264.8	185.8	167.4	313.9	345.5	237.2
		±121.14	±82.33	±87.24	±122.57	±129.50	±123.16
Percentage of the total sprinting distance,%	I half	2.25	1.80	1.43	2.77	3.25	2.13
		±1.22	±0.71	±0.76	±1.42	±1.46	±1.22
	II half	2.49	1.77	1.43	2.61	3.01	2.13
		±1.37	±1.05	±0.87	±1.22	±1.36	±1.27
	Total	2.38	1.78	1.42	2.68	3.13	2.13
		±1.04	±0.74	±0.74	±1.07	±1.09	±1.07

Figura 14 - Diferenças na distância total percorrida entre todas as posições (Andrzejewski et al, 2015)

Apesar do rigor dos dados físicos apresentados, os autores reconhecem que o estudo não avalia o impacto tático das ações, ou seja, não distingue sprints que resultam em vantagem territorial, cobertura defensiva ou criação de espaço. Esta limitação, mostra que a análise física isolada é insuficiente para captar plenamente o contributo individual no jogo.

Assim, a análise posiciona-se como um contributo claro para a individualização do treino, mas evidencia também uma lacuna clássica: a dissociação entre esforço físico e impacto tático. Embora seja possível caracterizar o perfil físico do jogador, estes dados, isoladamente, dizem pouco sobre como esse perfil influencia decisões e comportamentos ao longo do jogo.

Neste sentido, o estudo de Mandorino et al. (2025) oferece um avanço conceptual relevante, ao integrar métricas fisiológicas. Apresenta um contributo particularmente relevante ao explorar a interação entre *fitness*, fadiga e desempenho tático em competição. Segundo os autores, compreender como o estado fisiológico influencia o comportamento do jogador em jogo real é essencial para uma avaliação verdadeiramente holística da performance. O estudo operacionaliza estes conceitos através de índices compostos — FI (*fitness*) e FA (*fadiga*) -

derivados de métricas como PlayerLoad, respostas cardíacas e dados tático-técnicos. A estrutura metodológica é sintetizada no próprio artigo num modelo conceptual claro.

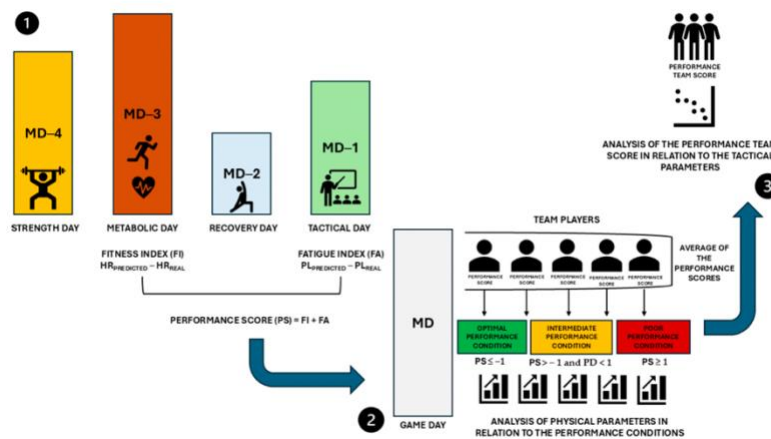


Figura 15 - Resumo dos passos realizados para calcular os valores PSindividual e PSteam (Mandorino et al, 2025)

**PlayerLoad** - Métrica derivada de acelerometria tri-axial que quantifica a carga física externa de um jogador através da soma das mudanças de aceleração nos três eixos (vertical, lateral e frontal). Representa, de forma agregada, o “stress mecânico” suportado pelo atleta durante esforços intermitentes, sendo amplamente utilizada para monitorizar intensidade e exigência física em treino e competição.

**PSindividual** - Índice que expressa a *performance score* individual, combinando indicadores tático-técnicos e métricas de intensidade física para quantificar o contributo de cada jogador ao longo do jogo. Resume o impacto individual numa escala contínua, permitindo comparar jogadores em diferentes momentos ou estados fisiológicos.

**PSteam** - Índice que expressa a *performance score* coletivo, agregando as contribuições individuais para avaliar o comportamento global da equipa em termos de domínio territorial, progressão, ações ofensivas e organização defensiva. Funciona como uma métrica de síntese que reflete o rendimento tático da equipa como um todo.

Apesar de metodologicamente sólido, Mandorino et al. (2025) reconhecem limitações importantes neste estudo: a amostra reduzida - um único clube e 23 jogadores - e a dependência de tecnologia de monitorização dispendiosa limitam a generalização das conclusões. Além disso, os autores admitem que variáveis psicológicas, como stress competitivo ou tomada de decisão sob fadiga, não foram consideradas, sugerindo a necessidade de modelos mais ricos e multimodais.

A avaliação de desempenho individual também se estende a domínios mais prescritivos, nomeadamente a análise da qualidade das decisões, da criação de espaço e da ameaça ofensiva. O estudo de Saseendran et al. (2023) mostra como, através de dados de tracking extraídos automaticamente a partir de vídeo de transmissão, utilizando técnicas de computer vision como YOLOv5 e DeepSORT, é possível calcular métricas como *pitch control* e *expected threat*, permitindo avaliar a capacidade de um jogador tomar decisões ótimas em contexto dinâmico.

Segundo os autores, esta solução visa democratizar o acesso a métricas avançadas - tradicionalmente dependentes de sistemas de tracking dispendiosos - oferecendo uma alternativa de baixo custo com precisão aceitável. O pipeline completo proposto no estudo é sintetizado de forma clara através de um diagrama operacional.

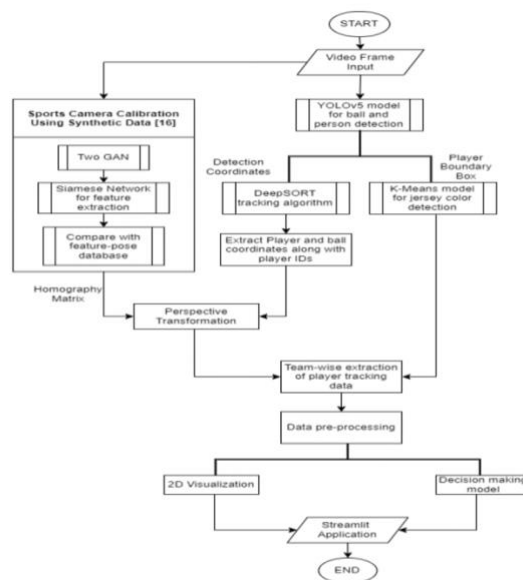


Figura 16 - Arquitetura Geral do Sistema (Saseendran et al, 2023)

De acordo com Saseendran et al. (2023), a metodologia desenvolvida consegue identificar e seguir jogadores e bola em tempo real, permitindo calcular métricas espaciais fundamentais como *pitch control* e *expected threat*. Os autores demonstram estes cálculos recorrendo a casos de estudo que ilustram a ocupação zonal, a pressão defensiva e o controlo do espaço em diferentes momentos do jogo. Estes exemplos são usados para mostrar que a informação gerada pelo sistema pode apoiar a avaliação da qualidade das decisões individuais.

Um dos contributos mais relevantes apresentados no estudo é a análise comparativa entre decisões “boas” e “menos eficazes”, utilizando métricas como *expected threat*. O estudo mostra-nos que jogadores que optam por passes para zonas de maior ameaça apresentam maior valor esperado da ação, reforçando a utilidade do tracking extraído do vídeo na análise do processo de decisão.



Figura 17 - Exemplo de imagem para o tracking de jogador, bola e equipa utilizando DeepSORT e YOLOv5 (Saseendran et al, 2023)

Apesar da utilidade da abordagem, Saseendran et al. (2023) reconhecem algumas limitações importantes. A precisão do sistema depende fortemente da qualidade da transmissão, da estabilidade da câmara e das condições de iluminação, podendo introduzir erros na deteção e identificação dos jogadores. Além disso, a metodologia não incorpora, de forma explícita, informação contextual como pressão adversária medida diretamente, ou dados fisiológicos, o que limita a profundidade da avaliação individual.

Ainda assim, concluem que a sua proposta permite aproximar clubes com recursos limitados de análises avançadas de tomada de decisão e de ocupação espacial, representando um avanço significativo na democratização das ferramentas de tracking e na avaliação comportamento tático individual.

Outro contributo importante surge do domínio das métricas de finalização, Cefis & Carpita (2025) apresentam um contributo relevante ao proporem um novo modelo de *expected goals* (xG) que integra informação de tracking, indicadores compostos e técnicas de regressão logística ajustadas. Segundo os autores, os modelos xG tradicionais - baseados quase exclusivamente em coordenadas de remate - não captam adequadamente fatores críticos como o posicionamento do guarda-redes, a pressão defensiva ou o histórico técnico do jogador. Assim, esta abordagem supera vários modelos anteriores ao melhorar a capacidade de previsão de golo e ao incorporar novas dimensões como o posicionamento do guarda-redes e a presença de defesas próximos.

O modelo desenvolvido incorpora variáveis espaciais (ângulo de remate, distância, posição relativa do guarda-redes), métricas de pressão defensiva derivadas de tracking e dois indicadores compostos: um associado ao comportamento do guarda-redes e outro ao perfil do finalizador. Estes indicadores são construídos através de *Partial Least Squares Structural Equation Modeling* (PLS-SEM), cuja arquitetura é apresentada pelos autores no seguinte esquema.

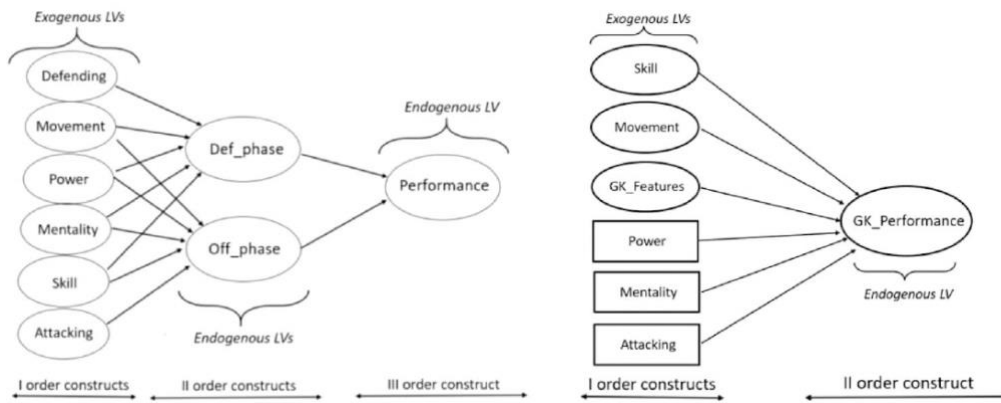


Figura 18 - Arquitetura do Modelo PLS-SEM (Cefis & Carpita, 2025)

O dataset é composto por 660 remates da Serie A 2019/20, combinando Understat, Math&Sport e KPIs da plataforma Sofifa. A distribuição espacial dos remates aparece representada na Figura 19, enquanto que a figura 20, mais à frente, sintetiza as variáveis utilizadas, desde ângulo e distância até métricas de pressão, posicionamento do guarda-redes e indicadores compostos. Esta diversidade de informação permite aos autores avançar para além dos modelos geométricos clássicos e incorporar fatores críticos que influenciam a probabilidade de golo.

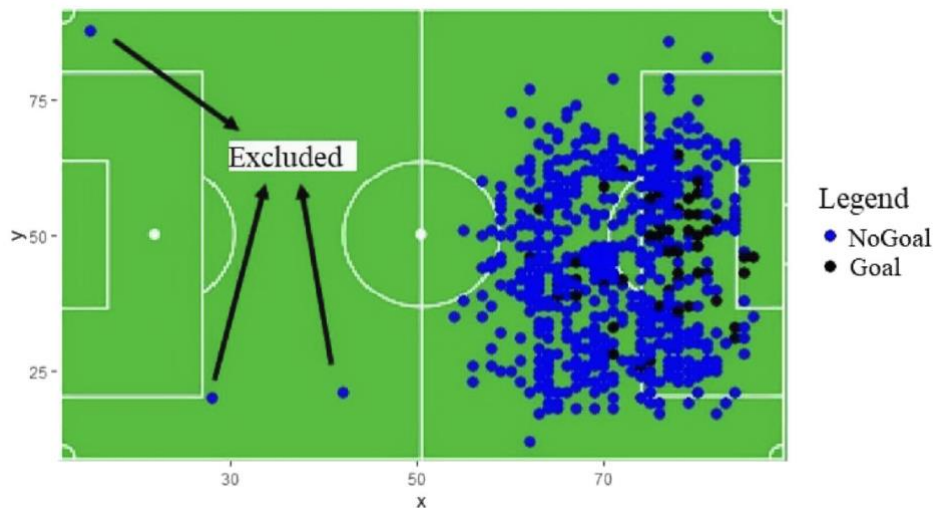


Figura 19 - Distribuição no terreno de jogo dos 660 remates provenientes dos 53 jogos da época 19/20 da Serie A Italiana (Cefis & Carpita, 2025)

Variable description	Source Dataset	Mean	Std.	Q1	Q2	Q3
GOAL (Yes/No)	Understat	0.10	0.30	-	-	-
x (shooter coordinate on the pitch, %)	Understat	83.22	7.64	77.00	83.00	89.00
y (shooter coordinate on the pitch, %)	Understat	49.98	15.06	37.00	50.00	63.00
Favourite foot? (Yes/No)	Understat	0.77	0.42	-	-	-
Previous dribbling to shot? (Yes/No)	Understat	0.32	0.47	-	-	-
Angle of shot (degree)	Understat	37.39	20.81	20.22	39.42	52.80
Previous ball distance (% the ball distance to goal before arriving to the shooter)	Tracking	14.43	8.41	8.60	14.33	17.09
Possession duration (sec.)	Tracking	6.49	4.81	5.68	6.49	6.49
D1.OpponentsPlayer (Dummy, if 1 = at least one player between shooter and goalkeeper)	Tracking	0.13	0.33	-	-	-
D2.OpponentsPlayer (Dummy, if 1 = at least two players between shooter and goalkeeper)	Tracking	0.05	0.21	-	-	-
Goalkeeper x coordinate on the pitch (%)	Tracking	96.74	5.25	96.55	97.67	98.41
Goalkeeper y coordinate on the pitch (%)	Tracking	50.08	5.06	47.37	50.04	53.07
Defending shooter ability	Sofifa	0.06	0.99	-0.62	0.14	0.61
Mentality shooter ability	Sofifa	0.38	1.00	-0.28	0.41	1.17
Movement shooter ability	Sofifa	0.41	1.03	-0.27	0.32	1.00
Power shooter ability	Sofifa	0.49	1.06	-0.16	0.53	1.10
Skill shooter ability	Sofifa	0.47	0.91	-0.15	0.49	1.12
Goalkeeper Attacking/Offensive ability	Sofifa	-0.03	1.01	-0.78	0.2	0.77
Goalkeeper Features ability	Sofifa	0.68	0.78	0.32	0.82	1.12
Goalkeeper Mentality ability	Sofifa	0.26	0.98	0.08	0.43	0.82
Goalkeeper Movement ability	Sofifa	0.55	0.73	0.12	0.53	1.02
Goalkeeper Power ability	Sofifa	0.45	0.92	0.10	0.45	1.19
Goalkeeper Skill ability	Sofifa	0.06	0.79	-0.51	-0.19	0.52

Figura 20 - Estatísticas e variáveis dos 660 remates provenientes dos 53 jogos da época 19/20 da Serie A Italiana (Cefis & Carpita, 2025)

Segundo Cefis & Carpita (2025), a inclusão destas novas dimensões melhora a capacidade explicativa do modelo, algo evidente quando comparam a versão completa com um modelo reduzido baseado apenas em ângulo e distância. Essa diferença está claramente demonstrada na Figura 21, onde o modelo completo apresenta valores de Pseudo- $R^2$  superiores, reforçando que o contexto - e não apenas a geometria do remate - desempenha um papel central na probabilidade de gol.

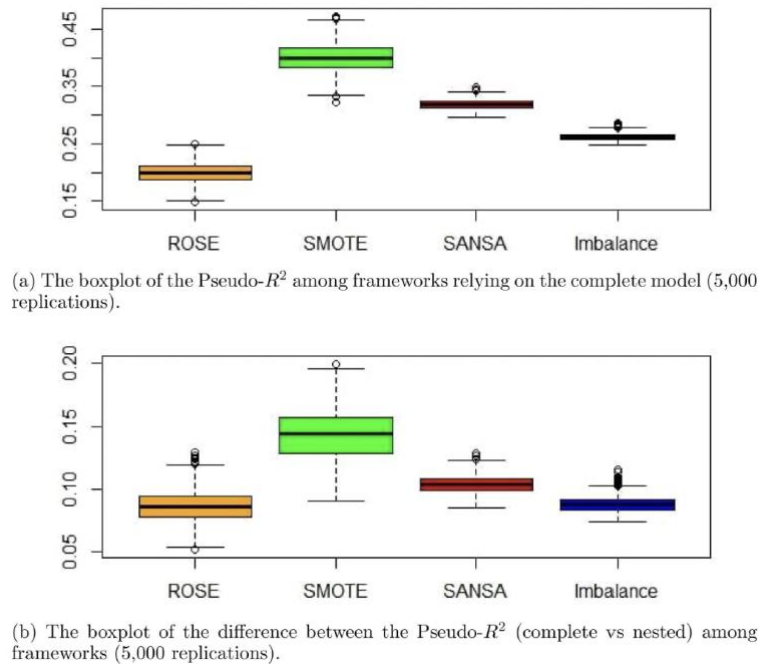


Figura 21 - Distribuições Pseudo- $R^2$  (Cefis & Carpita, 2025)

Para contexto:

**ROSE (Random Over-Sampling Examples)** - método de balanceamento que cria exemplos sintéticos através da estimativa da distribuição dos dados. Em vez de duplicar exemplos minoritários, gera novas observações “realistas” para equilibrar classes raras, reduzindo o viés de modelos treinados com dados não balanceados.

**SMOTE (Synthetic Minority Over-sampling Technique)** - um dos métodos mais usados para lidar com classes minoritárias. Funciona criando exemplos sintéticos ao interpolar observações da classe rara com os seus vizinhos mais próximos. Ajuda o modelo a reconhecer padrões associados a eventos pouco frequentes, como golos.

**SANSA (Self-Adjusted Nearest Synthetic Algorithm)** - uma variante mais recente do SMOTE, ajustando dinamicamente a criação de exemplos sintéticos consoante a densidade local dos dados. Isto permite gerar exemplos mais consistentes com a estrutura real da amostra, reduzindo a probabilidade de criar casos artificiais pouco representativos.

**Imbalance (sem resampling)** - refere-se simplesmente a treinar o modelo sem qualquer técnica de balanceamento, utilizando diretamente a proporção original dos dados (onde a classe golo é rara). Neste cenário, o modelo tende a favorecer a classe maioritária, mas pode melhorar métricas como precisão ou especificidade.

Já a análise dos coeficientes, apresentada na Figura 22 mostra que variáveis como o posicionamento do guarda-redes, o número de defensores entre o rematador e a baliza e indicadores como *Movement* e *GK Skill* têm impacto significativo. Isto confirma a ideia central dos autores: a finalização deve ser entendida como uma ação contextual e multifatorial.

Regressor	Coefficients				Odds			
	ROSE	SMOTE	SANSA	Imbl.	ROSE	SMOTE	SANSA	Imbl.
x	0.94***	2.09***	1.50***	1.74***	2.57	8.10	4.50	5.70
y	0.02	-0.05	0.07**	-0.09***	1.02	0.95	1.07	0.91
Favourite foot	-0.16	0.10	0.03	0.17***	0.85	1.10	1.03	1.19
Previous dribbling	0.21*	0.33*	0.30***	0.46***	1.23	1.39	1.34	1.59
Angle of shot	-0.53***	-1.40***	-1.16***	-1.09***	0.59	0.25	0.31	0.34
Previous ball distance	-0.09*	-0.22*	-0.24***	-0.06**	0.92	0.80	0.79	0.94
Possession duration	-0.18	-0.25*	-0.21***	-0.22***	0.84	0.78	0.81	0.80
D1.OpponentsPlayer	-0.74	-1.02*	-1.03	-0.71	0.48	0.36	0.36	0.49
D2. OpponentsPlayer	-5.69***	-6.13***	-6.08*	-5.68***	0.00	0.00	0.00	0.00
GK x coordinate	-0.33***	-0.59**	-0.30***	-0.33***	0.72	0.56	0.75	0.72
GK y coordinate	0.27***	0.48***	0.42***	0.46***	1.32	1.62	1.52	1.59
Defending	0.14*	0.13	0.14***	0.16***	1.15	1.14	1.15	1.17
Mentality	0.03	-0.23	-0.10*	-0.40***	1.03	0.79	0.91	0.67
Movement	0.12*	0.40*	0.25***	0.27***	1.13	1.49	1.29	1.31
Power	-0.01	-0.07	-0.08*	0.14***	0.99	0.93	0.92	1.16
Skill	0.08	0.23*	0.20***	0.21***	1.09	1.26	1.22	1.23
GK_Attacking	0.02	0.05	0.06*	0.03*	1.02	1.05	1.06	1.03
GK_features	0.03	0.32*	0.28***	0.22***	1.03	1.38	1.33	1.24
GK_Mentality	-0.15*	-0.44**	-0.38***	-0.33***	0.86	0.64	0.68	0.72
GK_Movement	0.10*	0.15	0.15**	0.07**	1.10	1.17	1.16	1.08
GK_Power	-0.17*	-0.20	-0.29***	-0.17**	0.84	0.82	0.75	0.84
GK_Skill	0.10*	0.27**	0.20***	0.20***	1.11	1.31	1.22	1.22

Figura 22 - Coeficientes da regressão e Estimativas dos Odds Ratios após 5.000 iterações, referentes aos 53 jogos da época 19/20 da Serie A Italiana (Cefis & Carpita, 2025)

Apesar dos avanços, Cefis & Carpita (2025) reconhecem limitações importantes: a amostra é relativamente pequena, restrita a uma única época; e a regressão logística, embora interpretável, pode não captar relações não-lineares mais complexas entre variáveis de tracking. Mesmo assim, o estudo representa um passo significativo ao aproximar os modelos de xG da realidade do jogo, oferecendo uma leitura mais rica e contextual da qualidade das finalizações.

Por último, o estudo de Bransen & Van Haaren (2019) representa um contributo importante na avaliação do desempenho individual ao focar-se num aspeto frequentemente negligenciado na literatura: o valor das ações com bola que não são remates, em particular os passes. Embora a maioria dos modelos avance sobretudo na quantificação da qualidade dos remates, os autores lembram que apenas 2% das ações com bola são finalizações, enquanto 75% são passes, o que torna evidente a necessidade de métricas que capturem o contributo destes momentos para o rendimento individual e coletivo.

Para resolver esta lacuna, os autores desenvolvem uma abordagem que atribui a cada passe um valor baseado na sua capacidade de aumentar a probabilidade de a sequência de posse resultar num golo. A lógica deste processo está ilustrada de forma clara na Figura 23, onde se observa como o valor de um passe corresponde à diferença entre o valor esperado da posse antes e depois da ação - uma interpretação intuitiva que aproxima estes modelos das leituras reais que analistas e treinadores procuram fazer durante os jogos.

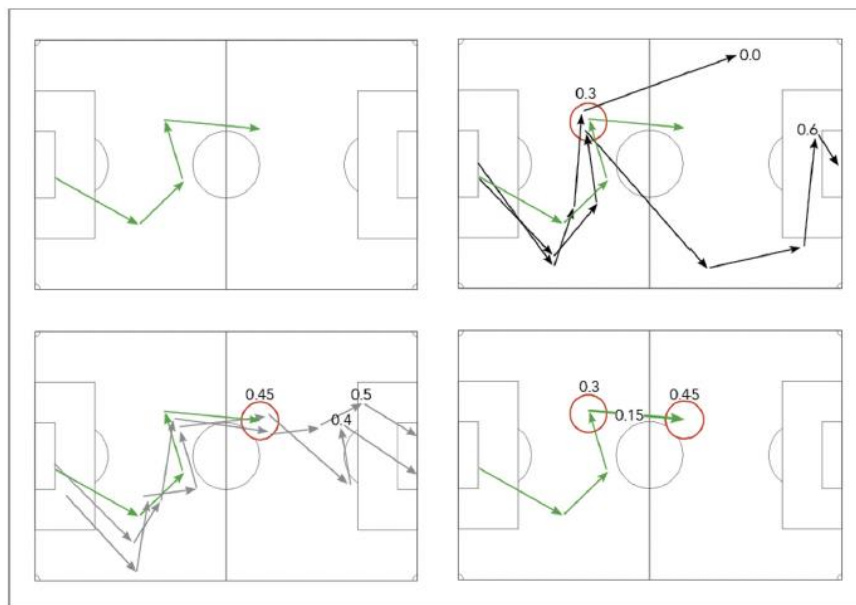


Figura 23 - Visualização da abordagem (Bransen & Van Haaren, 2019)

A metodologia combina quatro elementos essenciais: (1) construção de sequências de posse, (2) rotulagem dessas sequências com o seu expected goals, (3) decomposição das sequências

em subsequências e (4) estimativa do valor de cada passe através de k-nearest neighbors com DTW (Dynamic Time Warping).

Este uso do DTW permite comparar sequências de ações com base nas suas trajetórias espaço-temporais, aproximando o modelo do comportamento real em jogo. Para aumentar a eficiência computacional, os autores recorrem a um sistema de clusterização do campo, reduzindo milhões de possíveis comparações para conjuntos mais pequenos e coerentes.

A robustez do método é avaliada em contextos reais, utilizando mais de 12 milhões de ações provenientes de sete grandes ligas europeias e os autores validam ainda o modelo aplicando previsões de resultados esperados com base nos contributos dos passes por jogador. Os resultados mostram que a abordagem supera métricas tradicionais como *pass accuracy*, que por si só, dizem pouco sobre a verdadeira influência das ações de um jogador.

Os rankings finais apresentados na Figura 24 ilustram bem esta diferença: jogadores como Ragnar Klavan, Mesut Özil e Toni Kroos surgem destacados não por volume de ações, mas pela sua capacidade de aumentar consistentemente o valor das posses da sua equipa.

Rank	Player	Team	Contribution P90
1	Ragnar Klavan	Liverpool FC	0.1133
2	Mesut Özil	Arsenal	0.1034
3	Toni Kroos	Real Madrid	0.0943
4	Manuel Lanzini	West Ham United	0.0892
5	Joan Jordán	SD Eibar	0.0830
6	Esteban Granero	Espanyol	0.0797
7	Nuri Sahin	Borussia Dortmund	0.0796
8	Mahmoud Dahoud	Borussia Dortmund	0.0775
9	Granit Xhaka	Arsenal	0.0774
10	Faouzi Ghoulam	SSC Napoli	0.0765

Figura 24 - Top 10 de jogadores melhores classificados que jogaram pelo menos 900 minutos durante a época 17/18 nas 5 principais ligas europeias (Bransen & Van Haaren, 2019)

A leitura dos resultados ilustra ainda a relação entre valor médio por passe e número de passes por 90 minutos, evidencia também perfis contrastantes - entre jogadores que influenciam muito por poucas ações e jogadores que acumulam contribuição pelo volume.

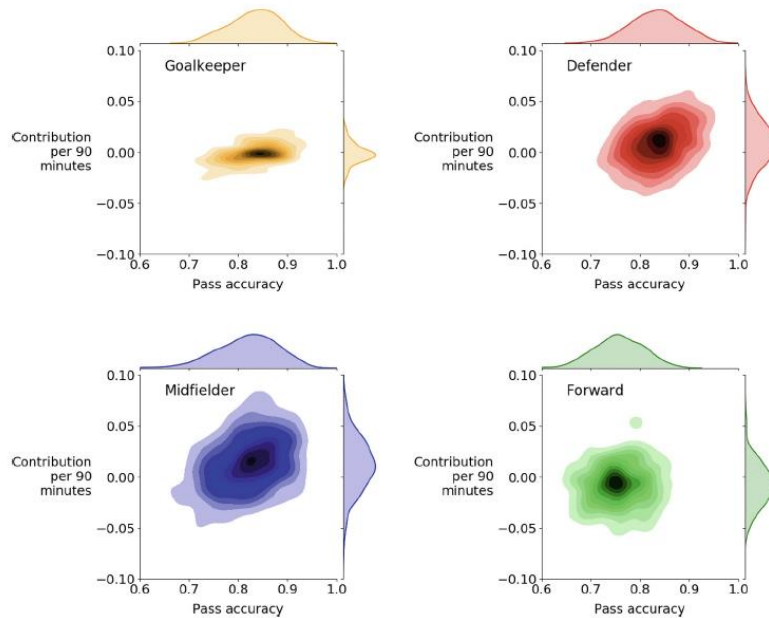


Figura 25 - Gráficos de densidade por posição que mostram a correlação entre a precisão de passe e a contribuição de passe por 90min (Bransen & Van Haaren, 2019)

Bransen & Van Haaren (2019) mostram uma aplicação prática e muito interessante do seu modelo ao utilizá-lo para encontrar um substituto para Manuel Lanzini, jogador do West Ham United, depois da sua lesão grave. A ideia é simples, mas eficaz: criar uma função de similaridade baseada em três dimensões - contributo por passe, número de passes por 90 min e precisão de passe - e procurar jogadores cujo perfil seja comparável ao médio argentino.

A análise revela que Lanzini combina elevado impacto por passe com precisão relativamente baixa, o que sugere um jogador que arrisca mais e procura constantemente progredir no terreno.

Com base na função de similaridade, a Figura 26 apresenta os jogadores mais próximos do seu perfil, destacando-se Mahmoud Dahoud, Joan Jordán e Naby Keïta, todos médios que assumem risco e têm capacidade para acrescentar valor através do passe. Esta identificação automática demonstra bem o potencial do modelo para apoiar não só potenciais substituições ao nível de jogo e equipa, mas também em processos de scouting, ajudando a encontrar jogadores com impacto semelhante, mesmo em ligas e contextos distintos.

Rank	Player	Team	Similarity score
1	Mahmoud Dahoud	Borussia Dortmund	0.9955
2	Joan Jordán	SD Eibar	0.9881
3	Naby Keïta	RB Leipzig	0.9794
4	Dominik Kohr	Bayer 04 Leverkusen	0.9717
5	Medrán	Deportivo Alavés	0.9591

Figura 26 - Os cinco jogadores mais semelhantes a Manuel Lanzini (Bransen & Van Haaren, 2019)

O estudo tem méritos claros, sobretudo ao criar uma métrica que atribui significado às ações mais frequentes no jogo e ao aproximar a avaliação do passe de uma lógica contextual e probabilística. No entanto, algumas limitações permanecem. O modelo depende exclusivamente de dados de eventos, não incorporando tracking, o que significa que perde nuances como pressão adversária, linhas de passe possíveis ou ângulos disponíveis no momento da ação. Adicionalmente, a abordagem assenta fortemente na qualidade do modelo de xG inicial, e qualquer viés na sua construção propaga-se para toda a avaliação subsequente. Por último, como os autores reconhecem, o DTW, apesar de poderoso, é computacionalmente dispendioso e baseado em similaridade geométrica, o que pode não captar integralmente a intenção tática da jogada.

Ainda assim, Bransen & Van Haaren (2019) demonstram que é possível quantificar de forma objetiva o valor dos passes e identificar jogadores que influenciam o jogo através da progressão e criação de ações promissoras, algo que não transparece nas métricas tradicionais. O estudo reforça a ideia, presente ao longo desta literatura, de que a avaliação individual no futebol exige métricas contextuais e orientadas para impacto, e não apenas volume ou precisão.

Em síntese, os estudos convergem numa ideia essencial: a avaliação de desempenho individual no futebol deve ser multidimensional, integrando fatores técnicos, táticos, físicos e fisiológicos. Os contributos analisados diferem nos métodos — uns mais centrados em eventos, outros em tracking, outros ainda em métricas fisiológicas — mas todos reforçam a necessidade de contextualização tática e de modelos híbridos que capturem as interações complexas que definem o rendimento de um jogador. Para além disso, os estudos mais recentes evidenciam a relevância da explicabilidade: modelos demasiado opacos, sustentados por redes neuronais profundas ou métodos de amostragem complexos, apresentam desafios de adoção prática por parte de treinadores e departamentos de análise, que procuram não apenas previsões corretas, mas também razões claras que sustentem decisões.

Persistem, contudo, lacunas significativas. Em primeiro lugar, muitos modelos continuam dependentes de um único tipo de dado, o que dificulta a construção de uma visão verdadeiramente integrada da performance. Em segundo lugar, a contextualização situacional permanece insuficiente: o comportamento do adversário, o espaço disponível, o momento estratégico do jogo e as dinâmicas coletivas continuam subexplorados. Em terceiro lugar, vários estudos recorrem a amostras reduzidas ou específicas, o que limita a capacidade de generalização.

Por fim, a explicabilidade dos modelos permanece uma barreira crítica à sua adoção por equipas técnicas. A investigação futura deverá, por isso, orientar-se para frameworks multimodais, contextualmente ricos e interpretáveis, que permitam transformar dados complexos em conhecimento acionável e útil para o treino, para o scouting e para a gestão de talento.

#### 4.2.2. AVALIAÇÃO DO DESEMPENHO COLETIVO / TÁTICO DA EQUIPA

A avaliação do desempenho coletivo tem ganho destaque na investigação recente em futebol, impulsionada pelo acesso a dados posicionais de alta frequência e pelo desenvolvimento de métodos capazes de captar a complexidade das interações entre jogadores. Os estudos incluídos nesta categoria convergem na tentativa de compreender o comportamento da equipa enquanto unidade dinâmica, analisando formações, padrões de movimento, sincronização entre setores e princípios táticos como compactação, pressão, organização defensiva e processos ofensivos.

O trabalho de Andrienko et al. (2021) constitui um dos contributos mais relevantes na análise tática coletiva, ao propor uma framework conceptual e computacional que integra espaço, tempo e interação para compreender o comportamento de equipas em contexto competitivo. Os autores partem da premissa de que o futebol não pode ser analisado apenas como uma sucessão de eventos discretos, mas sim como um sistema dinâmico, onde a organização espacial e as relações temporais entre jogadores determinam grande parte da eficácia tática das equipas. Esta abordagem contrasta com modelos tradicionais centrados no evento e aproxima-se de perspetivas sistémicas, sustentadas em métricas derivadas de tracking.

A metodologia de segmentação temporal utilizada pelos autores é exemplificada na Figura 27, que apresenta uma sequência de operações de consulta temporal: as faixas amarelas marcam os intervalos selecionados, as áreas verdes mostram os intervalos filtrados após remoção dos trechos curtos, e as extensões em azul correspondem às ampliações por modificadores — ou seja, a figura ilustra como os “momentos táticos” são extraídos e estabilizados a partir do fluxo contínuo de *tracking*. Com base nesta segmentação, Andrienko et al. (2021) analisam os momentos em várias escalas temporais (1 a 2 segundos vs. 10 a 20 segundos), permitindo detetar reorganizações rápidas e padrões mais lentos de basculação e compactação. Esta abordagem é crucial, pois muitas organizações defensivas só se tornam evidentes quando analisadas num contexto temporal alargado.

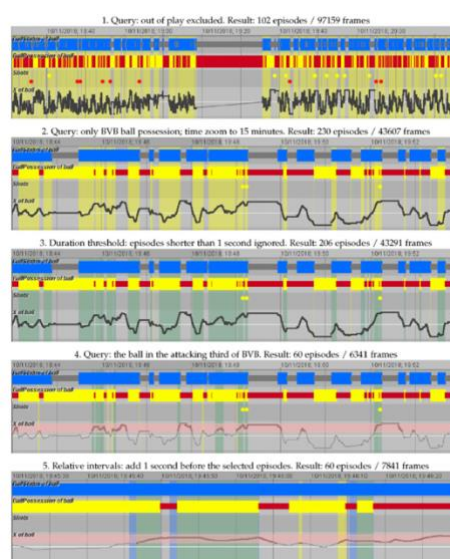


Figura 27 - Metodologia de Segmentação Temporal (Andrienko et al, 2021)

Com base nesta segmentação, Andrienko et al. (2021) exploram métricas espaciais derivadas de tracking, tais como ocupação do espaço, forma da equipa, compactação, zonas de acesso, linhas de jogo e padrões de cobertura. A força da abordagem está na forma como estas métricas são combinadas para representar comportamentos coletivos, permitindo observar dinâmicas ofensivas e defensivas em diferentes escalas temporais. Nas escalas curtas identificam-se microajustamentos, como por exemplo, pequenos deslocamentos coordenados para controlar zonas específicas, enquanto escalas mais longas revelam padrões de organização coletiva, como basculações estruturadas ou reorganizações após perda da bola.

A aplicação prática da framework é evidenciada na Figura 28, onde os autores apresentam exemplos completos de visualizações espaço-temporais baseadas em tracking. Estas visualizações incluem reconstruções das posições dos jogadores, zonas de influência e padrões de movimento ao longo do tempo, permitindo observar momentos de transição, reajustamentos defensivos e alinhamentos estruturais. A clareza destas representações torna a abordagem particularmente útil para analistas táticos, uma vez que facilita a interpretação visual de comportamentos que, de outra forma, seriam difíceis de descrever através de estatísticas tradicionais.

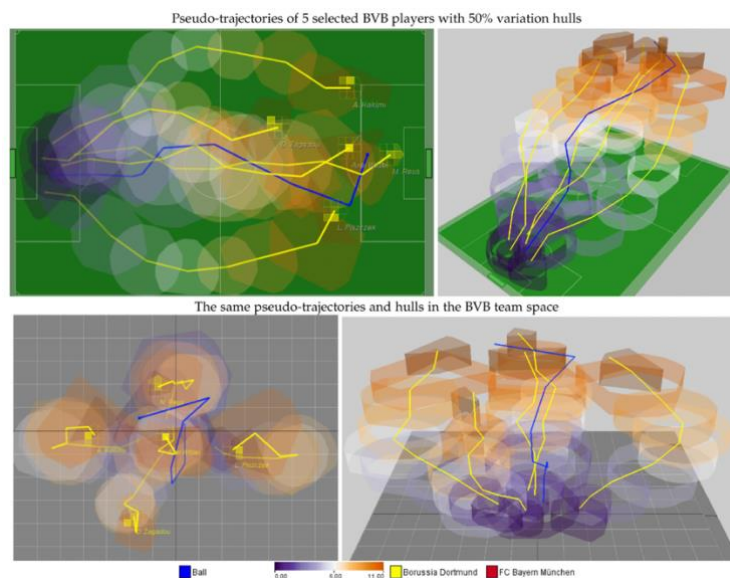


Figura 28 - Sequências de 10 posições dos jogadores e da bola durante a posse de bola do BVB, correspondentes a diferentes posições do centro da equipa no terreno de jogo ao longo do eixo X (Andrienko et al, 2021)

Segundo Andrienko et al. (2021), a principal vantagem da framework é a sua capacidade de integrar informação contínua, espacial e temporal, numa representação coerente do comportamento coletivo. Esta integração permite compreender não só *onde* os jogadores estão, mas *como* e *quando* se movimentam enquanto unidade, captando fenómenos como velocidade de reorganização, sincronização entre linhas, ou padrões de cobertura durante momentos defensivos. Trata-se de um avanço importante relativamente a abordagens puramente *event-based*, que não captam estas dinâmicas estruturais.

No entanto, a framework também apresenta limitações significativas. Em primeiro lugar, exige dados de tracking de alta qualidade e elevada frequência, o que limita a sua aplicabilidade em contextos com menor capacidade tecnológica. Em segundo lugar, embora descreva de forma detalhada os comportamentos coletivos, o modelo não integra métricas de ameaça ou ligação direta a criação de oportunidades - como *xThreat* ou *pitch value* - o que impede que os padrões identificados sejam relacionados com impacto ofensivo real. Em terceiro lugar, a complexidade computacional e a natureza exploratória da ferramenta tornam-na mais adequada à investigação ou a clubes com departamentos analíticos avançados do que a contextos mais tradicionais.

No estudo seguinte, de Tianbiao & Andreas (2016), os autores aplicam técnicas de *data mining* - em particular o algoritmo Apriori - para identificar padrões recorrentes de passe que caracterizam o comportamento coletivo das equipas. A ideia central é simples: as sequências de passe não são apenas ações isoladas, mas expressões consistentes da organização tática, e a sua recorrência pode revelar dinâmicas ofensivas e padrões estruturais do jogo.

Tianbiao e Andreas (2016) procederam à análise da final da Algarve Cup de 2012 através de métodos de observação estruturada. A abordagem metodológica adotada consistiu na segmentação do campo de jogo em 30 zonas distintas, conforme ilustrado na Figura 29.

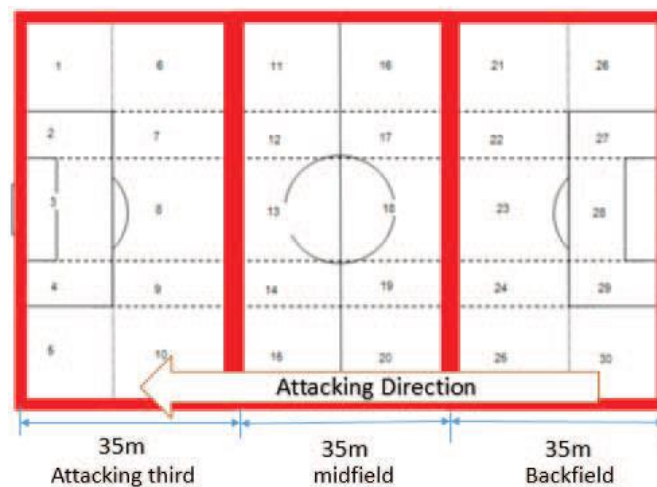


Figura 29 - Campo de futebol dividido em 30 zonas (Tianbiao & Andreas, 2016)

O modelo seguido compreendeu várias etapas, nomeadamente a definição, o registo, a limpeza e o subsequente processamento dos dados. Todos os eventos relevantes do jogo foram devidamente registados, conforme se apresenta na Figura 30. No que diz respeito à fase de limpeza dos dados, optou-se por considerar apenas os últimos cinco eventos da cadeia de controlo.

Para o tratamento analítico dos dados, foi utilizado o suplemento de mineração de dados da Microsoft SQL, recorrendo a uma versão modificada do algoritmo “Apriori”.

index	Start of Sequence	index	End of Sequence
100	kickoff	111	Goal
101	Free kick	112	Shoot but caught by goalkeeper
102	Throw in	113	Shoot bounced (by GK or goalframe)
103	Conner kick	114	Shoot out of goalframe
104	Goal kick	116	Be fouled
105	Attacking starts	117	be kicked out of touchline by opponent
106	Goalkeeper starts the ball	118	be kicked out of goalline by opponent
107	Referee whistles to restart the game	119	passing/shooting but intercepted
122	offside	120	out of touchline/goalline
123	Referee whistles to stop the game	121	dribbling, receiving but lost control

Figura 30 - Definição dos dados utilizados (Tianbiao & Andreas, 2016)

A aplicação do algoritmo implicava a separação de duas variáveis fundamentais para a estimativa da probabilidade, também denominada confiança. Esta probabilidade é calculada com base no conceito de support, que, no contexto da mineração de dados, corresponde à frequência mínima com que determinado padrão deve ocorrer nos registos para ser considerado relevante (Tianbiao & Andreas, 2016).

Através desta metodologia, foram extraídas Regras de Associação, permitindo identificar combinações de jogadores eficazes e recorrentes, tanto na primeira como na segunda parte da partida. Conforme ilustrado na Figura 31, na primeira metade do jogo, observaram-se interações mais frequentes entre os jogadores com os números 4 e 14, bem como entre os números 7 e 13. Já na segunda parte, destacaram-se as ligações entre os jogadores 2 e 17, e 2 e 9. Estas interações estão evidenciadas por setas pretas mais espessas, simbolizando a sua influência positiva na geração de oportunidades de golo (Tianbiao & Andreas, 2016).

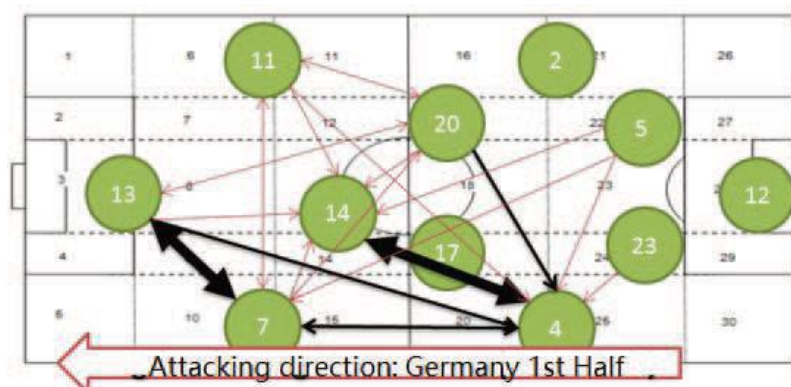


Figura 31 - Modelo de Jogador – 1ª Parte (Tianbiao & Andreas, 2016)

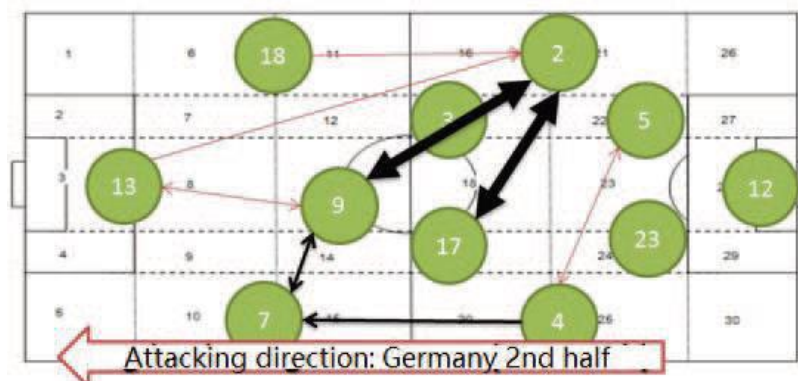


Figura 32 - Modelo de Jogador – 2ª Parte (Tianbiao & Andreas, 2016)

Os resultados obtidos, em conjugação com a utilização da versão adaptada do algoritmo Apriori, permitiram concluir que a equipa analisada demonstrou um desempenho superior na primeira parte do jogo, com especial destaque para o jogador número 4, identificado como o elemento mais influente. Na segunda parte, apesar de uma descida na qualidade exibicional, o jogador número 2 revelou-se o mais impactante.

Apesar da utilidade prática do método, Tianbiao & Andreas (2016) reconhecem limitações claras. A abordagem depende exclusivamente de dados de eventos, não incorporando informação de tracking nem variáveis temporais contínuas, o que impede a identificação de ajustes dinâmicos ou padrões de movimento sem bola. Além disso, a extração de padrões frequentes não distingue sequências eficazes de sequências neutras, já que o método não avalia diretamente o impacto ofensivo ou a relação com criação de perigo.

Mesmo assim, o estudo oferece uma ferramenta simples e eficaz para avaliar a consistência tática coletiva, ajudando a identificar tendências estruturais no jogo da equipa e contribuindo para uma análise mais informada das dinâmicas ofensivas.

Numa outra investigação, mais focada no futebol de formação, Borges et al. (2019), analisa como diferentes abordagens ofensivas - counter-attack, quick attack e positional attack - influenciam o sucesso das sequências ofensivas em jogadores U-15 e U-17. A ferramenta utilizada para suporte à codificação e tratamento da informação foi o software *Match Vision Studio Premium*, que permitiu aos autores elaborar uma matriz categorial estruturada segundo as variáveis definidas para estudo.

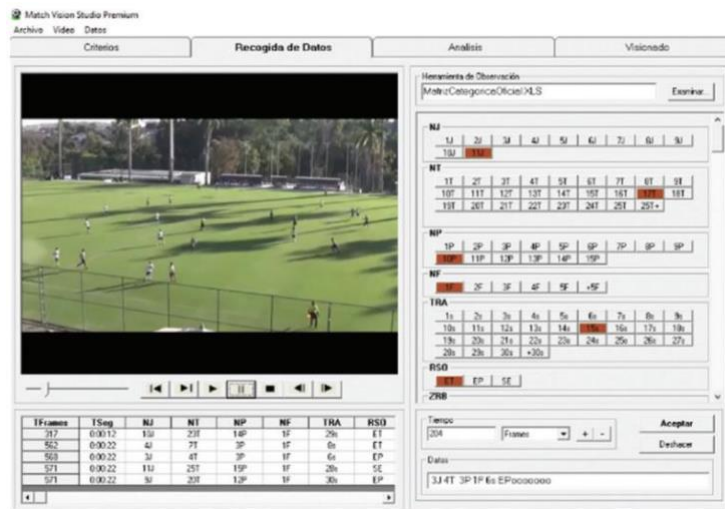


Figura 33 - Interface do match Vision Studio Premium (Borges et al, 2019)

A abordagem é totalmente observacional, centrada em 218 sequências de jogos reais, e recorre a uma matriz categorial construída especificamente para captar indicadores tático-técnicos relevantes, como número de passes, duração da ação, jogadores envolvidos e mudanças de corredor.

Variables	Description	Counter attack (CA)	Quick attack (QA)	Positional attack (PA)
<b>Players Involved</b>	Number of players that were involved, touching the ball during the offensive action.	Ball recovered in any area of the playing field	Ball recovered in any area of the playing field	Ball recovered in any area of the playing field
<b>Ball Touches</b>	Total number of ball touches performed by players during the offensive action.	Performs equal or less than 5 passes	Performs a maximum of 7 passes	Performs more than 7 passes
<b>Passing</b>	Total number of passes made with any part of the body that was received by the attacking partner and continued the offensive phase of the team.	Offensive sequence duration equal to or less than 12 seconds	Offensive sequence duration equal to or less than 18 seconds	Offensive sequence duration exceeding 18 seconds
<b>Duration</b>	The duration of the offensive phase, from the interception of the ball, to the end of the offensive action (seconds).	Opponent team advanced on the pitch and defensively unbalanced	Opponent team balanced defensively	Opponent team balanced defensively
<b>Corridor Changes*</b>	Number of times that the ball changed field corridors during the offensive action, taking into account the division of the field into 3 corridors (left, central, and right).	Ball circulation more in depth than width	Ball circulation in width and depth	Ball circulation more in width than depth
<b>Results of Offensive Actions</b>	<b>Description</b>	High play intensity	High play intensity	Cadenced play intensity
<b>Success</b>	Sequence finished in goal.			
<b>Failure</b>	Sequence finished with kick out or goalkeeper's defense.			

Figura 34 - Variáveis Ações Ofensivas e Abordagens de Jogo (Borges et al, 2019)

Um dos resultados mais evidentes surge na Figura 35, onde Borges et al. (2019) mostram que o *positional attack* envolve mais jogadores, mais passes, mais toques e dura significativamente mais tempo do que os ataques rápidos ou contra-ataques. Esta diferença confirma a natureza mais elaborada deste método ofensivo, caracterizado por circulação de bola mais rápida e exploração mais ampla do espaço.

	Counter-Attack Md (Q1-Q3)	Quick Attack Md (Q1-Q3)	Positional Attack Md (Q1-Q3)	p	$\eta^2$
Players Involved (no)	3.00 (2.00-3.00)	3.00 (2.00-4.00)	8.00 (7.00-9.00)*	<0.01	0.23
Passing (no)	2.00 (2.00-3.00)	2.00 (1.00-3.00)	9.50 (8.00-11.25)*	<0.01	0.24
Duration (sec.)	11.00 (9.00-12.00)	8.00 (4.50-14.00)	30.00 (26.00-36.25)*	<0.01	0.25
Ball Touches (no)	7.00 (5.00-8.00)	6.00 (3.00-9.00)	24.00 (20.00-25.50)*	<0.01	0.24
Corridor Changes (no)	1.00 (0.00-2.00)	1.00 (1.00-2.00)	5.00 (3.00-7.00)*	<0.01	0.17

\*Different from Counter-Attack and Quick Attack

Figura 35 - Comparação dos indicadores tático-técnicos entre diferentes métodos ofensivos (Borges et al, 2019)

Noutra análise interessante, Borges et al. (2019) comparam sequências bem-sucedidas versus falhadas. As ações que terminam em gol têm, de forma consistente, mais passes, mais toques e maior duração, sugerindo que, em idades jovens, manter a bola por mais tempo e envolver mais jogadores tende a aumentar a probabilidade de sucesso. Este padrão é reforçado pela análise de correlação da Figura 36 onde variáveis como passes, toques e duração aparecem como elementos centrais do sucesso ofensivo.

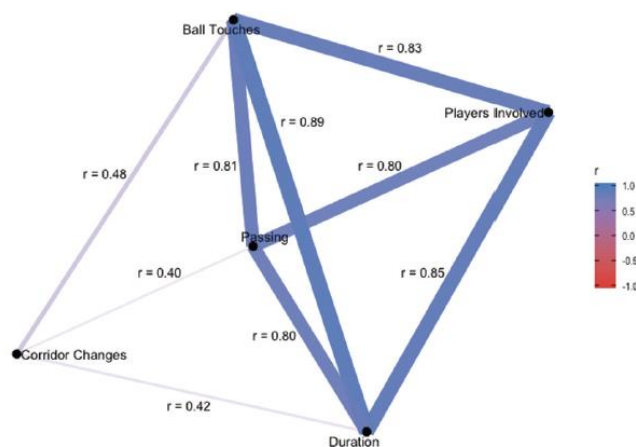


Figura 36 - Análise de rede de correlações entre variáveis táticas e técnicas em sequências ofensivas que terminaram com sucesso - gol (Borges et al, 2019)

Em termos interpretativos, os autores argumentam que o futebol juvenil parece beneficiar de comportamentos ofensivos mais elaborados e organizados, ao contrário de abordagens excessivamente diretas. No entanto, reconhecem que a eficácia também depende da capacidade da equipa adversária se reorganizar defensivamente, algo que métodos rápidos, como o *counter-attack*, conseguem explorar quando o adversário está desorganizado.

O estudo é sólido, mas limitado pela amostra restrita a duas equipas específicas, o que impede generalizações mais amplas. Ainda assim, Borges et al. (2019) oferecem uma leitura clara sobre como jovens jogadores constroem ações ofensivas eficazes, mostrando que a qualidade

da execução - e não a abordagem ofensiva em si - é o elemento diferenciador em contextos de formação.

Por fim, no estudo de Mohammed et al. (2025), os autores analisam de forma integrada o rendimento tático e físico da seleção marroquina ao longo dos sete jogos no Mundial 2022, recorrendo exclusivamente aos relatórios oficiais da FIFA e a uma framework composta por três níveis de análise: descritiva, visual e exploratória. Esta abordagem permite perceber como Marrocos combinou organização defensiva, esforço físico contínuo e transições rápidas para alcançar resultados excecionais face a adversários teoricamente superiores.

A análise descritiva mostra logo um padrão marcante: Marrocos jogou com posse reduzida na maioria dos jogos, não ultrapassando os 40% em cinco partidas. Apesar da baixa posse, os indicadores de carga física foram muito elevados, com picos de distância total frente a Espanha (120,2 km) e França (118,9 km). Esta combinação - pouca posse, elevada capacidade de esforço - ajuda a explicar a eficácia do bloco baixo marroquino.

Match	Possession (%)	Shots (on Target)	Passes Completed (%)	Total Distance (km)	High Intensity (km)	Pressure Applied	Total Shots
Croatia (1)	35.2	2	81	106.3	14.5	250	5
Belgium	33.5	2	80	108.4	13.9	261	7
Canada	41.5	2	83	109.1	15.2	245	6
Spain	23.0	1	70	120.2	16.1	288	3
Portugal	27.5	1	75	112.6	14.8	277	5
France	49.1	3	89	118.9	16.0	299	13
Croatia (2)	45.1	2	87	111.5	15.5	277	7

Figura 37 - Dados físicos coletivos por jogo (Mohammed et al, 2025)

Os perfis individuais reforçam esta leitura. A Figura 38 mostra que jogadores como Amrabat, Hakimi, Ounahi e Ziyech apresentaram padrões de deslocamento e intensidade consistentes, mas diferenciados: Amrabat pela estabilidade defensiva, Hakimi pela explosão em sprints, Ounahi pela capacidade de projeção entre zonas e Ziyech pelo desequilíbrio ofensivo aliado a grande volume físico. Estes perfis complementares foram determinantes para sustentar um modelo baseado em solidariedade coletiva e ocupação racional do espaço.

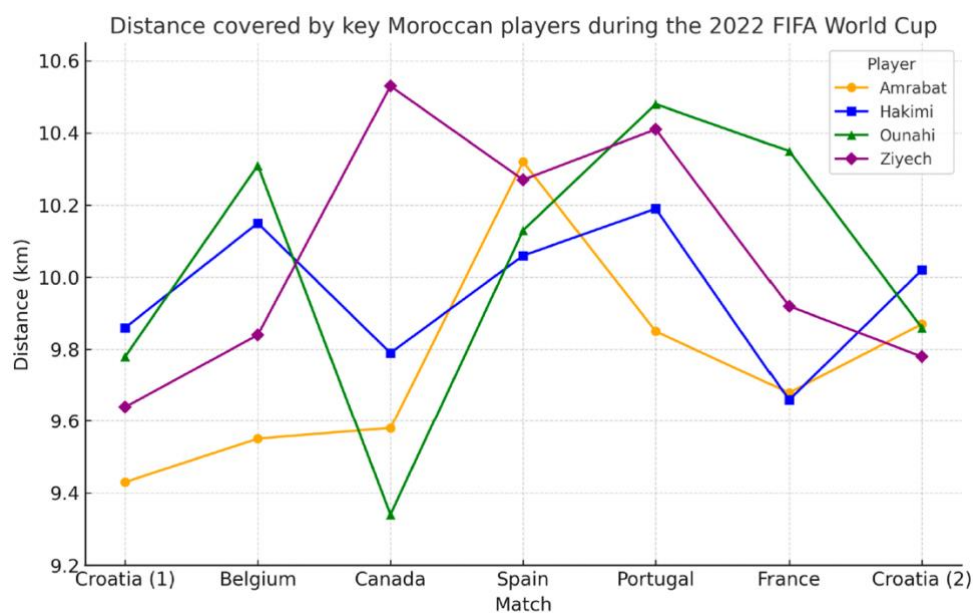


Figura 38 - Distância percorrida por quatro jogadores-chave ao longo dos jogos – FIFA World Cup 2022 (Mohammed et al, 2025)

As visualizações táticas, revelaram também uma forte dependência do corredor direito (Hakimi e Ziyech), com Amrabat a funcionar como pivô estrutural. Contra equipas como Espanha ou França, esta rede torna-se mais curta e concentrada, refletindo a pressão adversária e a necessidade de defender mais baixo, já diante de Portugal, o padrão alarga, com maior exploração dos corredores laterais e transições verticais, coerente com o plano de jogo. A análise exploratória através de PCA e clustering, distingue três tipos de jogo ao longo do torneio:

1. Jogos de bloco baixo: marcados por forte compactação e grande volume físico (ex.: Espanha);
2. Jogos de transição rápida: com recuperações altas e grande verticalidade (ex.: Bélgica, Portugal);
3. Jogos mais abertos: onde Marrocos tentou assumir mais posse, embora com menor eficácia (ex.: França).

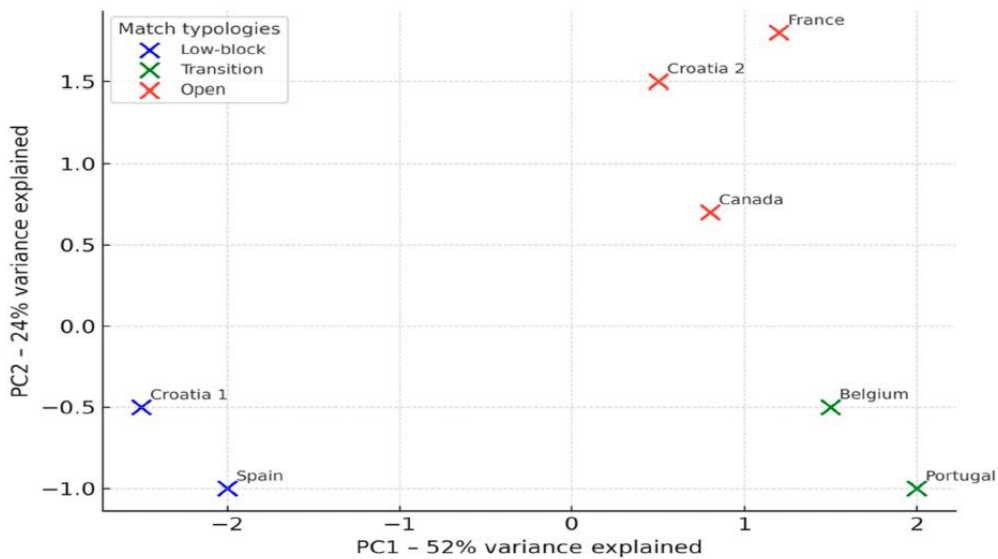


Figura 39 - Agrupamento dos jogos por perfil de desempenho - PCA + Clustering (Mohammed et al, 2025)

Esta classificação mostra que a equipa exibiu elevada adaptabilidade, ajustando o modelo de jogo ao contexto competitivo e ao perfil do adversário, uma das conclusões mais fortes de Mohammed et al. (2025).

No conjunto, o estudo demonstra que o sucesso de Marrocos no Mundial assentou num equilíbrio entre estrutura defensiva, intensidade física e inteligência contextual. Apesar de depender apenas de dados oficiais da FIFA, o trabalho oferece uma leitura clara e fundamentada da coerência tático-física da equipa, destacando um modelo altamente pragmático e eficiente em torneios de curta duração.

Em síntese, os estudos incluídos nesta categoria evidenciam um caminho comum: a necessidade de compreender o desempenho coletivo através de padrões relacionais, estruturas dinâmicas e interações constantes entre jogadores, superando a análise puramente posicional ou estatística. O desempenho da equipa é entendido como uma entidade emergente, resultante de comportamentos coordenados e de adaptações estratégicas que se desenvolvem em resposta ao adversário e ao contexto competitivo.

#### 4.2.3. PREVISÃO TÁTICA E APOIO À DECISÃO

Este tópico constitui uma área emergente na investigação sobre análise de dados em futebol, impulsionada pelo desenvolvimento de técnicas de aprendizagem automática capazes de modelar padrões complexos e não lineares presentes nos comportamentos coletivos. Os estudos incluídos nesta categoria procuram antecipar configurações táticas, formações, padrões estratégicos e estilos de jogo antes ou durante o próprio jogo, oferecendo novas

possibilidades de apoio à decisão para treinadores, analistas e departamentos de performance.

A aplicação de modelos analíticos ao apoio direto à tomada de decisão representa uma das evoluções mais significativas do uso de dados no futebol moderno. Ao contrário das categorias anteriores - centradas na descrição, avaliação ou previsão de comportamentos - os estudos incluídos nesta categoria procuram intervir no processo decisório, oferecendo recomendações operacionais para treinadores e analistas. Trata-se, portanto, de uma vertente aplicada da análise de dados, onde o objetivo não é apenas compreender o jogo, mas influenciá-lo através de decisões mais informadas.

Inserido nesta linha está o estudo de Suzuki et al. (2019), que apresenta uma abordagem interessante para estimar automaticamente táticas coletivas a partir de vídeo, recorrendo a modelos Deep Extreme Learning Machines (DELMs). O objetivo é identificar, em cada instante, qual das quatro táticas pré-definidas uma equipa está a utilizar - *Retreat, Fore check, Possession, Swift attack* - combinando informação posicional dos jogadores com características do próprio jogo.

A abordagem inicia-se com a extração automática de intervalos temporais relevantes, ilustrada na Figura 40, onde os autores mostram como filtram e estabilizam segmentos do jogo para identificar momentos táticos consistentes.

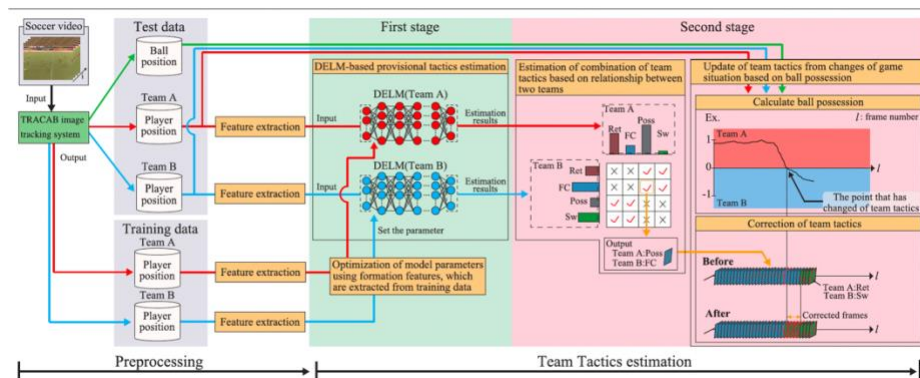


Figura 40 - Visão Geral do pré-processamento e do método proposto (Suzuki et al, 2019)

As características utilizadas pelo modelo incluem métricas de comportamento coletivo, como distâncias e variações entre jogadores, e indicadores de ocupação espacial. Estas divisões posicionais, fundamentais para a análise, surgem representadas na Figura 41, que mostra o campo particionado em 15 zonas, permitindo quantificar a distribuição dos jogadores por áreas.

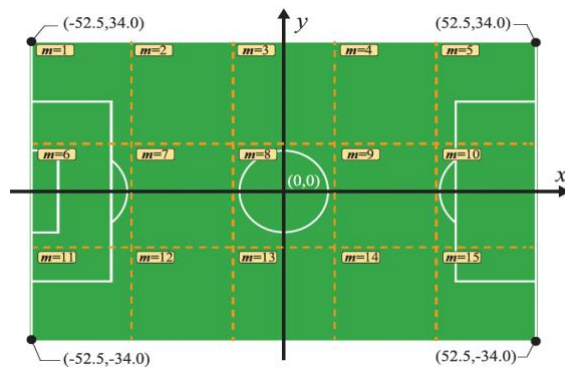


Figura 41 - Campo de jogo e as suas 15 regiões divididas (Suzuki et al, 2019)

A arquitetura DELM usada para classificar as táticas surge na Figura 42, mostrando um modelo pensado para ser leve e eficiente, evitando as necessidades de grandes volumes de treino típicas das CNNs.

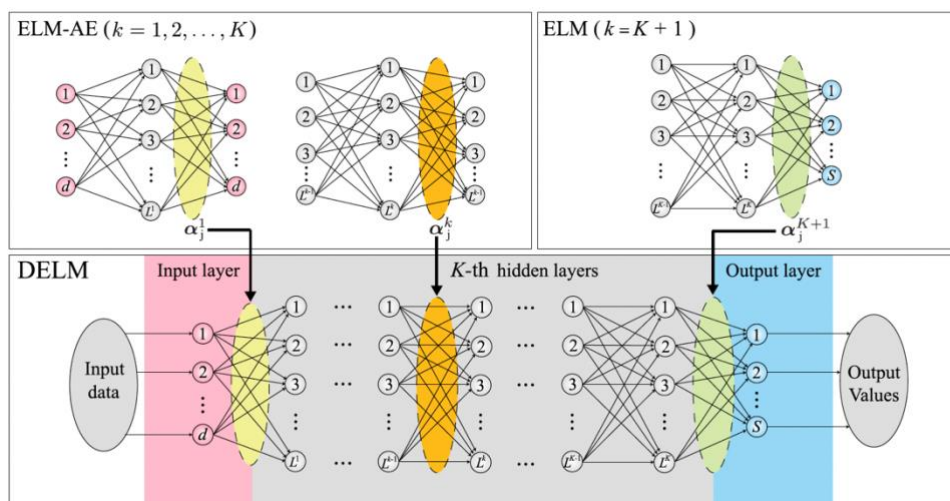


Figura 42 - Estrutura do Modelo DELM (Suzuki et al, 2019)

Após esta classificação inicial, o método passa por uma segunda etapa: ajustar as previsões com base em duas informações essenciais - a posse de bola e a relação entre as táticas das duas equipas. Esta lógica está representada na Figura 43 e ajuda a corrigir estimativas incoerentes, tornando a sequência de táticas mais consistente com aquilo que acontece realmente no jogo.

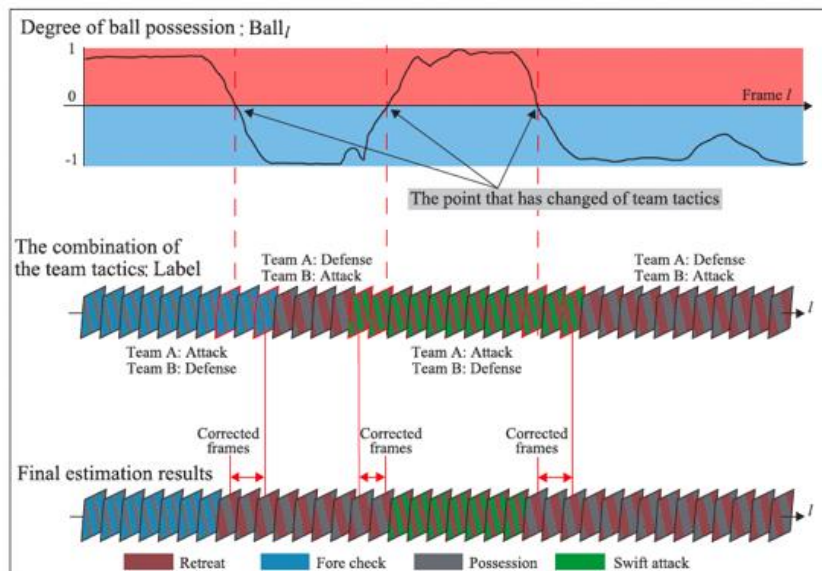


Figura 43 - Método para corrigir os resultados da estimativa com base na posse de bola (Suzuki et al, 2019)

Os resultados são claros. A Figura 44 mostra que o método supera várias abordagens alternativas, incluindo CNNs populares, tanto em *precision* como em *recall*.

Team tactics	Ours			Comp. 1			Comp. 2		
	R	P	F	R	P	F	R	P	F
Retreat	<b>0.928</b>	<b>0.917</b>	<b>0.923</b>	0.780	0.394	0.523	0.683	0.670	0.676
Fore check	0.871	<b>0.830</b>	<b>0.850</b>	0.408	0.658	0.504	0.611	0.651	0.630
Possession	<b>0.928</b>	<b>0.945</b>	<b>0.936</b>	0.653	0.471	0.547	0.699	0.720	0.709
Swift attack	0.791	<b>0.812</b>	<b>0.801</b>	0.112	0.565	0.187	0.637	0.644	0.640
Average	<b>0.880</b>	<b>0.876</b>	<b>0.878</b>	0.488	0.522	0.440	0.658	0.671	0.664

Team tactics	Comp. 3			Comp. 4			Comp. 5		
	R	P	F	R	P	F	R	P	F
Retreat	0.657	0.672	0.665	0.600	0.616	0.608	0.877	0.825	0.850
Fore check	0.660	0.628	0.644	0.605	0.581	0.593	0.633	0.779	0.699
Possession	0.671	0.672	0.671	0.617	0.612	0.615	0.881	0.827	0.854
Swift attack	0.664	0.647	0.656	0.610	0.592	0.601	0.643	0.752	0.693
Average	0.663	0.655	0.669	0.608	0.600	0.604	0.759	0.796	0.774

Figura 44 - Comparação da Estimativa das Táticas das Equipas vs. Métodos Comparativos (Suzuki et al, 2019)

Enquanto as Figuras 45 e 46 mostram ainda que, depois da segunda etapa, as confusões entre táticas próximas - como *Retreat* e *Possession* - diminuem significativamente.

		Estimation results			
		Ret	FC	Poss	Sw
Gt	Ret	1,396	48	110	19
	FC	31	510	39	52
	Poss	152	67	1,720	20
	Sw	57	12	44	334

Figura 45 - Confusion matrix obtido a partir dos resultados da estimativa antes do cálculo da combinação ótima (Suzuki et al, 2019)

		Estimation results			
		Ret	FC	Poss	Sw
Gt	Ret	1,425	43	50	17
	FC	9	527	22	47
	Poss	80	54	1,897	14
	Sw	40	11	38	337

Figura 46 - Confusion matrix obtido a partir dos resultados da estimativa depois do cálculo da combinação ótima (Suzuki et al, 2019)

No conjunto, Suzuki et al. (2019) demonstram que integrar aprendizagem automática com características estruturais do jogo, como posse de bola e dependências entre táticas, permite melhorar significativamente a estimativa automática de comportamentos coletivos. O método é simples, eficiente e destaca-se pela capacidade de incorporar contexto, algo frequentemente ausente em abordagens puramente estatísticas ou baseadas apenas em padrões de movimento.

Por outro lado, o estudo de Nouraie et al. (2023) apresenta o ITFPS, um sistema de apoio à decisão que visa auxiliar treinadores na definição de equipas e na escolha de jogadores para cada posição, combinando deep learning, teoria dos grafos e algoritmos de otimização. A lógica do método assenta em três passos principais: (1) atribuição de pontuações a jogadores para cada posição, (2) escolha da formação ótima através do algoritmo húngaro e finalmente, criação de conjuntos de candidatos para cada posição, permitindo ao treinador ajustar a equipa conforme necessidades contextuais.

O primeiro passo consiste em treinar uma rede neuronal para prever a adequação de cada jogador às 21 posições consideradas. Embora os modelos apresentem erros distintos por posição Nouraie et al. (2023) mostram que a rede neuronal é mais eficaz a distinguir perfis posicionais do que alternativas como regressão linear ou *random forest*.

No segundo passo, os autores utilizam um grafo bipartido entre jogadores e posições, atribuindo a cada aresta o score previsto pela rede. O algoritmo húngaro identifica então a combinação de jogadores que maximiza o rendimento esperado da formação. Para avaliar o alinhamento entre ITFPS e decisões reais, Nouraie et al. (2023) aplicam o método às equipas do Manchester City, Liverpool, Newcastle e Leeds (época 2021/22). A comparação entre treinador e algoritmo é ilustrada na Figura 47, mostrando que Guardiola e Klopp apresentam

as maiores similaridades, sugerindo decisões alinhadas com padrões ótimos baseados em atributos.

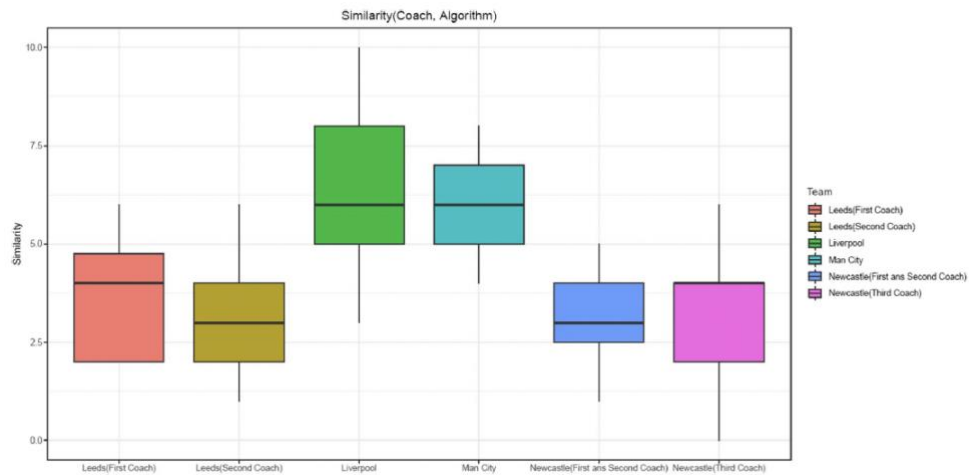


Figura 47 - Boxplot dos resultados da pontuação de similaridade (Nouraie et al, 2023)

O terceiro passo introduz um elemento flexível: em vez de selecionar um único jogador para cada posição, o ITFPS cria conjuntos de candidatos, definidos com base nos desempenhos relativos (parâmetro  $\theta$ ). Isso permite ao treinador gerir rotações, lesões e preferências táticas sem perder coerência estrutural. A eficácia deste passo é avaliada através da Figura 48 onde se observa que equipas como Liverpool e Manchester City apresentam maiores valores de “Accordance”, indicando seleções de jogadores mais próximas do cenário ótimo sugerido pelo sistema.

Team	$\theta$	Min	1st	Med	3rd	Max	Mean
<b>Manchester City</b>	$\theta=1$	4	6	6	6.75	8	6.02
	$\theta=3$	5	7	7	7	8	6.89
<b>Liverpool</b>	$\theta=1$	1	5	6	7	8	5.78
	$\theta=3$	1	6	7	8	9	6.91
<b>Newcastle United(First Coach)</b>	$\theta=1$	1	2	3	3	4	2.45
	$\theta=3$	4	4.5	5	6.5	7	5.36
<b>Newcastle United(Second Coach)</b>	$\theta=1$	2	3	4	4	5	3.48
	$\theta=3$	3	5	6	6	7	5.33
<b>Leeds United(First Coach)</b>	$\theta=1$	2	4	5	5	8	4.53
	$\theta=3$	4	5.25	6	7	9	6.23
<b>Leeds United(Second Coach)</b>	$\theta=1$	4	4	5	6.25	8	5.41
	$\theta=3$	6	7	7	7	9	7

Figura 48 - Valores de mínimo, 1º quartil, mediana, 3º Quartil, máximo e desvio-padrão para a métrica Accordance entre o treinador e o ITFPS (Nouraie et al, 2023)

Para combinar seleção e posicionamento, os autores propõem o Ratio Metric, cujo desempenho está detalhado na Figura 49. Este indicador mostra que treinadores de topo, como Guardiola ou Klopp, tendem a maximizar simultaneamente escolha e posicionamento, enquanto equipas em dificuldades, como o Newcastle antes da mudança de treinador, exibem valores inferiores.

Team	$\Theta$	Min	1st	Med	3rd	Max	Mean
Manchester City	$\Theta=1$	0.5	0.75	0.75	0.87	1	0.78
	$\Theta=3$	0.55	0.7	0.76	0.77	0.88	0.74
Liverpool	$\Theta=1$	0.2	0.71	0.85	0.87	1	0.78
	$\Theta=3$	0.25	0.85	1	1	1	0.91
Newcastle United(First Coach)	$\Theta=1$	0.33	0.4	0.42	0.55	0.75	0.48
	$\Theta=3$	0.57	0.69	0.71	0.8	0.87	0.73
Newcastle United(Second Coach)	$\Theta=1$	0.42	0.66	0.71	0.8	1	0.73
	$\Theta=3$	0.6	0.71	0.75	0.85	0.87	0.76
Leeds United(First Coach)	$\Theta=1$	0.33	0.5	0.62	0.7	0.88	0.59
	$\Theta=3$	0.44	0.63	0.7	0.77	0.9	0.69
Leeds United(Second Coach)	$\Theta=1$	0.44	0.5	0.59	0.85	1	0.67
	$\Theta=3$	0.66	0.7	0.73	0.77	1	0.76

Figura 49 - Valores de mínimo, 1º quartil, mediana, 3º Quartil, máximo e desvio-padrão para a métrica Ratio Metric entre o treinador e o ITFPS (Nouraie et al, 2023)

De forma geral, Nouraie et al. (2023) demonstram que é possível transformar o processo de construção de equipas num problema matemático bem definido, oferecendo soluções otimizadas que se aproximam de decisões tomadas pelos melhores treinadores da Premier League. O método é particularmente útil enquanto ferramenta de apoio à decisão, fornecendo alternativas objetivas e justificadas para seleção de jogadores e definição de formações.

Por último e para complementar as abordagens anteriores de previsão tática e apoio à decisão, o estudo de Berman et al. (2022) acrescenta uma perspetiva particularmente relevante ao focar-se na análise temporal do jogo e na recomendação de substituições em tempo real. Ao combinar métricas dinâmicas de redes de passe com modelos preditivos, os autores exploram como a evolução do comportamento coletivo pode informar decisões estratégicas durante o próprio encontro.

Assim os autores desenvolveram um sistema de análise temporal do jogo que combina métricas de redes de passe com modelos de machine learning, com o objetivo de identificar tendências táticas ao longo do encontro e recomendar substituições em tempo real. A base do método é o modelo NMELSA, que constrói redes de passe atualizadas continuamente em janelas de 10 minutos, permitindo capturar flutuações no comportamento coletivo da equipa.

A Figura 50 exemplifica uma rede de passes gerada para o “El Clásico” espanhol, destacando como estruturas de circulação variam consoante o momento do jogo.

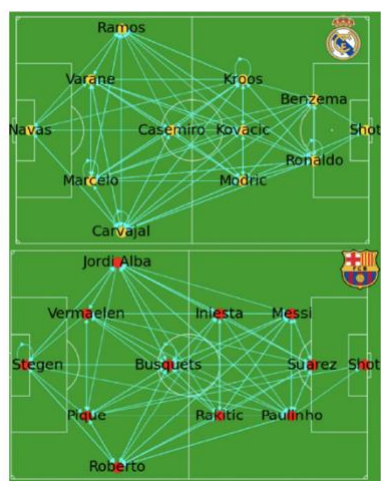


Figura 50 - Rede de passes da primeira parte do El Clásico entre o Real Madrid (em cima) e o Barcelona (em baixo) (Berman et al, 2022)

Sobre estas redes, Berman et al. (2022) calculam seis métricas centrais - *out-degree*, *betweenness*, *clustering*, *closeness*, *eigenvector centrality* e *entropia* - cuja evolução temporal é apresentada na Figura 51, mostrando que certos padrões métricos antecedem momentos críticos, como períodos de domínio que se correlacionam com golos do Barcelona.

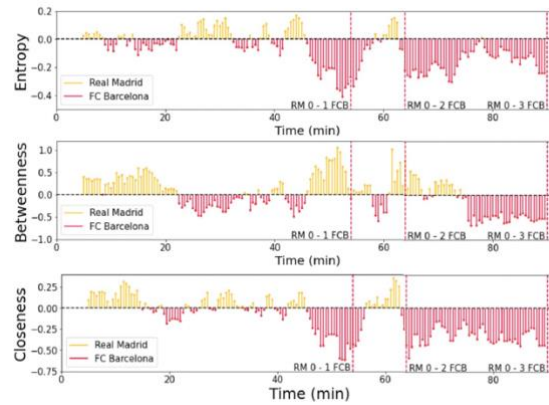


Figura 51 - Série temporal de três das seis métricas de rede para o jogo Barcelona vs. Real Madrid. As linhas tracejadas assinalam os golos do Barcelona (Berman et al, 2022)

Com base nestas séries temporais, os autores testam modelos preditivos de resultado - win, draw e loss - e embora métodos como Time Series Forest e ROCKET apresentem limitações, como mostra a Figura 52, a combinação de métricas de rede com algoritmos como XGBoost e fDNN melhora a precisão de previsão, atingindo valores próximos de 67%

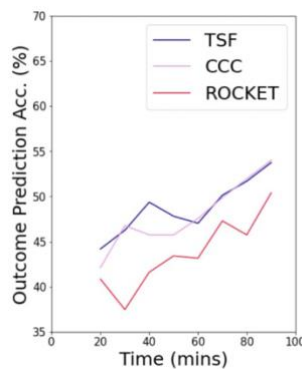


Figura 52 - Precisão da previsão de desfecho do jogo, dos três classificadores, em determinados momentos da partida (Berman et al, 2022)

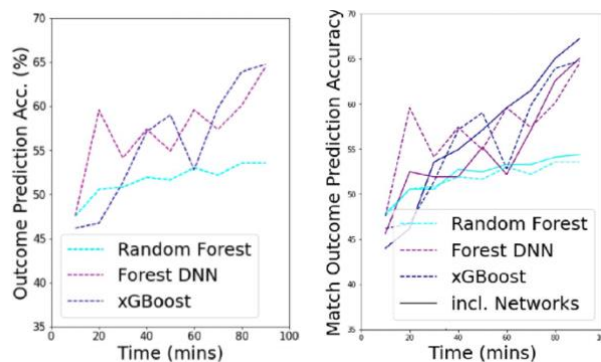


Figura 53 - Precisão da previsão de desfecho do jogo, utilizando algoritmos baseados em características (linhas tracejadas) e os mesmos algoritmos, mas com métricas de rede incluídas (linhas contínuas) (Berman et al, 2022)

A partir destas previsões e das métricas individuais extraídas da rede, Berman et al. (2022) desenvolveram dois modelos de substituição. O primeiro identifica, aos 60 minutos, o jogador com pior pontuação e simula como diferentes substitutos poderiam alterar a probabilidade de vitória, recomendando a troca que maximiza esse ganho. Por exemplo, o modelo mostrou que substituir Vermaelen por Aleix Vidal aumentaria substancialmente as probabilidades de vitória do Barcelona. Já o segundo modelo agrupa jogos em seis clusters táticos, via K-Means e recomenda substituições que aproximem a equipa do “cluster ideal” para enfrentar o estilo do adversário, como mostra a Figura 54.

Opponent Cluster	Desired Cluster	Match-ups	Point Ratio
0	3	86	0.80
1	5	89	0.82
2	3	94	0.43
3	5	49	0.84
4	5	18	0.96
5	4	78	0.50
6	3	15	0.47

Figura 54 - Relação entre os 6 Clusters (Berman et al, 2022)

Um dos resultados mais interessantes é que 71% dos jogadores que o modelo recomenda substituir acabam por ser substituídos pelo treinador na vida real, e 78% dos sugeridos para entrar também o fazem, sugerindo que o sistema capta tendências reais de desempenho e leitura tática do jogo. No entanto, apenas cerca de 6% das substituições coincidem exatamente - jogador out/jogador in - refletindo diferenças naturais entre decisão humana e algoritmos generalistas.

Embora eficaz, o modelo apresenta limitações reconhecidas pelos autores: trabalha sempre “em retrospectiva”, isto é, simula como seria o jogo se a substituição tivesse ocorrido antes, não prevê a evolução futura do jogo e não incorpora variáveis contextuais fundamentais, como o resultado, cartões, desgaste físico ou intenção tática do treinador. Ainda assim, Berman et al. (2022) demonstram de forma convincente que métricas de redes e modelos de previsão podem alimentar sistemas de apoio à decisão com relevância prática.

Em síntese, os estudos classificados nesta categoria revelam que o apoio à decisão baseado em dados está a aproximar-se progressivamente da prática operacional, oferecendo ferramentas capazes de orientar escolhas críticas como a formação inicial, as substituições ou a adaptação tática. Apesar dos avanços, permanece evidente que a adoção prática destes sistemas depende da capacidade de integrar dados quantitativos com o conhecimento subjetivo e experiencial dos treinadores, numa abordagem híbrida que reconheça as limitações inerentes à modelação de fenómenos altamente humanos e contextuais.

## **5. CONCLUSÃO**

### **5.1. SÍNTESE DO TRABALHO DESENVOLVIDO**

Este trabalho teve como objetivo compreender de que forma a análise de dados está a transformar a avaliação de desempenho no futebol, tanto ao nível individual como coletivo, e também no apoio à decisão antes e durante o jogo. Através de uma revisão sistemática conduzida pelo método PRISMA, foi possível reduzir um conjunto inicial muito amplo de estudos a um núcleo de artigos que representam bem o estado atual da investigação nesta área.

A análise teve o objetivo de dar resposta a três grandes tópicos: (1) a avaliação do desempenho individual, (2) a análise tática da equipa e (3) a previsão tática com apoio à decisão. Em cada um deles observou-se uma evolução clara, desde modelos que procuram quantificar ações individuais com maior rigor, até métodos que analisam padrões coletivos complexos, passando por abordagens preditivas capazes de antecipar formações, estilos de jogo ou até substituições. No conjunto, estes estudos mostram que o futebol está cada vez mais dependente da forma como os dados são recolhidos, tratados e interpretados, dando ao staff técnico ferramentas mais objetivas para fundamentar decisões.

De um modo geral, os objetivos definidos foram atingidos. A revisão permitiu perceber que dados e modelos analíticos já têm um papel central no processo de decisão, e que a tecnologia utilizada - seja tracking, redes de passe, deep learning ou métricas compostas - está a tornar o jogo mais mensurável e mais previsível do ponto de vista tático.

### **5.2. LIMITAÇÕES**

Como qualquer revisão sistemática, também esta ficou limitada pelos critérios aplicados na pesquisa, pelo período temporal definido e pelo facto de ter sido realizada numa única base de dados. Além disso, muitos clubes não partilham publicamente as ferramentas e métodos que usam internamente, o que restringe a capacidade de compreender plenamente como a análise de dados influencia o processo de decisão no contexto profissional. A heterogeneidade dos estudos - em dados, metodologias e métricas - torna igualmente difícil comparar resultados ou tirar conclusões totalmente generalizáveis

### **5.3. TRABALHO FUTURO**

Para investigações futuras, seria útil explorar metodologias complementares e integrar diferentes tipos de dados - técnicos, táticos, físicos e contextuais - num mesmo modelo. Outra direção passa por aproximar a investigação académica da realidade dos clubes, estudando mais de perto que ferramentas são usadas, como são interpretadas e de que forma influenciam decisões concretas. Finalmente, seria relevante avaliar o impacto real destas abordagens na performance das equipas, testando se decisões suportadas por dados conduzem, de facto, a melhores resultados dentro de campo.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Andreff, W. (2011). Some comparative economics of the organization of sports: competition and regulation in north American vs. European professional team sports leagues. *The European Journal of Comparative Economics*, 8(1), 3-27. <https://shs.hal.science/halshs-00677436>
- Andrienko, G., Andrienko, N., Anzer, G., Bauer, P., Budziak, G., Fuchs, G., ... & Wrobel, S. (2021). Constructing spaces and times for tactical analysis in football. *IEEE Transactions on Visualization and Computer Graphics*, 27(4), 2280-2297. <https://doi.org/10.1109/TVCG.2019.2952129>
- Andrzejewski, M., Chmura, J., Pluta, B., & Konarski, J. M. (2015). Sprinting Activities and Distance Covered by Top Level Europa League Soccer Players. *International Journal of Sports Science & Coaching*, 10(1), 39-50. <https://doi.org/10.1260/1747-9541.10.1.39>
- Barbon Junior, S., Pinto, A., Barroso, J. V., Caetano, F. G., Moura, F. A., Cunha, S. A., & Torres, R. da S. (2022). Sport action mining: Dribbling recognition in soccer. *Multimedia Tools and Applications*, 81(3). <https://doi.org/10.1007/s11042-021-11784-1>
- Berman Y., Mistry S., Mathew J. and Krishna A. (2022). "Temporal Match Analysis and Recommending Substitutions in Live Soccer Games,". *IEEE International Conference on Web Services (ICWS)*, Barcelona, Spain, 2022, pp. 397-404. <https://doi.org/10.1109/ICWS55610.2022.00066>
- Borges, P. H., Garganta, J., Guilherme, J., Jaime, M. D. O., Menegassi, V. M., Rechenchosky, L., ... & Rinaldi, W. (2019). Tactical efficacy and offensive game processes adopted by Italian and Brazilian youth soccer players. *Motriz: Revista de Educação Física*, 25, e101922. <https://doi.org/10.1590/s1980-6574201900020017>
- Bransen, L., Van Haaren, J. (2019). Measuring Football Players' On-the-Ball Contributions from Passes During Games. In: Brefeld, U., Davis, J., Van Haaren, J., Zimmermann, A. (eds) *Machine Learning and Data Mining for Sports Analytics. MLSA 2018. Lecture Notes in Computer Science*, vol 11330. Springer. [https://doi.org/10.1007/978-3-030-17274-9\\_1](https://doi.org/10.1007/978-3-030-17274-9_1)
- Cefis, M., & Carpita, M. (2025). A new xG model for football analytics. *Journal of the Operational Research Society*, 76(1), 1–13. <https://doi.org/10.1080/01605682.2024.2323669>
- Cummins, C., Orr, R., O'Connor, H., & West, C. (2013). Global positioning systems (GPS) and microtechnology sensors in team sports: a systematic review. *Sports medicine*, 43(10), 1025-1042. <https://doi.org/10.1007/s40279-013-0069-2>
- Gudmundsson, J., & Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys (CSUR)*, 50(2), 1-34. <https://doi.org/10.1145/3054132>

Lee, G. J., & Jung, J. J. (2022). DNN-based multi-output model for predicting soccer team tactics. *PeerJ Computer Science*, 8, e853. <https://doi.org/10.7717/PEERJ-CS.853>

Mandorino, M., Gabbett, T. J., Tessitore, A., Leduc, C., Persichetti, V., & Lacombe, M. (2025). The Interaction of Fitness and Fatigue on Physical and Tactical Performance in Football. *Applied Sciences*, 15(7), 3574. <https://doi.org/10.3390/app15073574>

Memmert, D., Lemmink, K. A., & Sampaio, J. (2017). Current approaches to tactical performance analyses in soccer using position data. *Sports medicine*, 47(1), 1-10. <https://doi.org/10.1007/s40279-016-0562-5>

Mohammed, B., Said, E. M., Lotfi, Z., Nourddine, E., & Fatima-Zahra, G. (2025). Tactical and Physical Profiling of the Moroccan National Football Team at the FIFA World Cup Qatar 2022: A Data-Driven and Artificial Intelligence-Assisted Analysis. *Applied Sciences*, 15(18), 9994. <https://doi.org/10.3390/app15189994>

Nouraie, M., Eslahchi, C., & Baca, A. (2023). Intelligent team formation and player selection: A data-driven approach for football coaches. *Applied Intelligence*, 53(24), 30250–30265. <https://doi.org/10.1007/s10489-023-05150-x>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, 372. <https://doi.org/10.1136/bmj.n71>

Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., & Giannotti, F. (2019). PlayeRank: data-driven performance evaluation and player ranking in soccer via a machine learning approach. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(5), 1-2. <https://doi.org/10.1145/3343172>

Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*, 5(1), 1410. <https://doi.org/10.1186/s40064-016-3108-2>

Robertson, Roland & Giulianotti, Richard. (2006). Football, Globalization and Glocalization. *Revista Internacional De Sociologia LXIV*. <https://doi.org/10.3989/ris.2006.i45.14>

Saseendran, S., Thanalakshmi, S. P. V., Prabakaran, S., & Ravisankar, P. (2023). Analysis of player tracking data extracted from football match feed. *Romanian Journal of Information Technology and Automatic Control*, ISSN 1220-1758, vol. 33(2), pp. 89-102. <https://doi.org/10.33436/v33i2y202307>

Suzuki, G., Takahashi, S., Ogawa, T., & Haseyama, M. (2018). Team Tactics Estimation in Soccer Videos via Deep Extreme Learning Machine Based on Players Formation. 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), 116–117. <https://doi.org/10.1109/GCCE.2018.8574814>

Tianbiao, L., & Andreas, H. (2016). Apriori-based diagnostical analysis of passings in the football game. 2016 IEEE International Conference on Big Data Analysis (ICBDA), 1–4. <https://doi.org/10.1109/ICBDA.2016.7509795>



**NOVA**

**IMS**

Information  
Management  
School