

Anais do



2023



**17 de novembro de 2023
Belo Horizonte | MG | Brasil**

Adriano C. M. Pereira (UFMG), Gabriel Valadão Meira (UFMG), Hugo Rios Neto (UFMG),
Leo M. Sá Freire (UFMG), Matheus Magalhães (UFMG), Thiago C. Porto (UFMG)
Wagner Meira Jr. (UFMG)

Patrocinador:



Organização:



DCC
DEPARTAMENTO DE
CIÊNCIA DA COMPUTAÇÃO

U F M G

Comissão Organizadora

A segunda edição do FAME - Football Analytics and Modeling Experience - contou com a colaboração e participação de um time de voluntários da Universidade Federal de Minas Gerais (UFMG), que listamos a seguir (em ordem alfabética):

- Adriano C. Machado Pereira
- Ana Clara Vasconcelos
- Bruno Martins
- Fabricio Cagnoni
- Gabriel Marra
- Gabriel Reis
- Gabriel Valadão Meira
- Henrique Magalhães de Oliveira Carvalho
- Hugo Rios-Neto
- João Lucas Lage Gonçalves
- Larissa Dolabella Gomide
- Leo M. Sá Freire
- Lucas Rocha Laredo
- Lucca Carvalho Augusto
- Lucca Stochiero
- Luiza Campos
- Matheus Magalhães
- Pedro Moreira
- Pedro Rausch
- Wagner Meira Jr.

Agradecimento

A Comissão organizadora agradece a todos que participaram, divulgaram e colaboraram neste evento. Cabe um agradecimento especial ao apoio e colaboração do Departamento de Ciência da Computação da Universidade Federal de Minas Gerais (DCC/UFMG) e à equipe do Centro de Atividades Didáticas 3 (CAD 3) da UFMG.

Aproveitamos para agradecer também à **GEMINI Sports Analytics** (<https://geminisports.ai>), que patrocinou nosso evento, tendo sido sua contribuição financeira fundamental para a realização e sucesso dessa edição do evento.

Patrocinador:



Organização:



DCC
DEPARTAMENTO DE
CIÊNCIA DA COMPUTAÇÃO

U F **m** G

Comitê de Programa

A segunda edição do FAME - Football Analytics and Modeling Experience – teve uma chamada de trabalhos técnicos, que foram apresentados no evento e, posteriormente, convidados para publicação dos trabalhos neste compêndio (Anais / Proceedings) de trabalhos do evento.

Este comitê de programa foi composto pelos seguintes colaboradores:

- Adriano C. Machado Pereira
- Gabriel Valadão Meira
- Hugo Rios-Neto
- Leo M. Sá Freire
- Matheus Magalhães
- Thiago C. Porto
- Wagner Meira Jr.

Os trabalhos submetidos constituíram de textos ou artigos acadêmicos, aplicações industriais e produtos tecnológicos de *startups* que utilizassem métodos quantitativos para solucionar problemas no esporte, preferencialmente no futebol. Sendo que foram sugeridos os seguintes temas, com possibilidade de outros temas relacionados:

- Análise tática e estratégica
- Análise de desempenho de jogadores e equipes
- Recrutamento e *valuation* de atletas
- Previsão e análise de resultados de jogos e competições
- Prevenção de lesões
- Otimização de treinamento e *performance* física

Cada trabalho enviou um resumo de até 1000 palavras, descrevendo o problema tratado e o objetivo geral, a metodologia, os resultados e as conclusões. E posteriormente convidados os(as) autores(as) a gerar uma versão final, com possibilidade de ser um resumo ou um artigo com mais detalhes do trabalho, a ser publicado em um compêndio do evento, que é este documento aqui apresentado. Esperamos que possam desfrutar da leitura dos trabalhos e aprender mais sobre essa temática do evento.

Cordialmente,

O Comitê de Programa – FAME'23.

Patrocinador:



Organização:



Sumário

Apresentamos o sumário dos trabalhos aprovados no evento e que aceitaram o convite para a publicação neste compêndio. No total, dos 20 trabalhos apresentados no evento FAME'23, dezesseis (16) fizeram uma revisão final dos trabalhos e submeteram para publicação, sendo: sete (7) trabalhos no formato de artigo e nove (9) no formato de resumo. Nesta seção listamos os trabalhos publicados:

Sumário - Artigos

Artigo	Título do Trabalho	Págs.
1	Predição de resultados de jogos do Campeonato Brasileiro de Futebol Feminino Ananda Mendes Souza, Anderson Almeida Ferreira, Renato Lopes Moreira	06-15
2	Analizando Robustez da Rede de Passes através da Dispersão de Hub e Authority Balthazar Paixão, Lucas Tavares, Pedro Henrique González, Glauco Amorim	16-22
3	Factorial Analysis Model of Football Teams Squads Gabriel Canda Blanco	23-32
4	Explainable xP: Solucionando Enviesamento de Modelos de Expected Pass com Tracking Gabriel R. G. Barbosa, Henrique M. O. Carvalho, Fernando T. C. Oliveira	33-43
5	Exploring the Relationship Between Scoring First, External and Internal Load, and Match Outcomes: A Case of Study of a Professional Brazilian Team during the 2021 Season Gabriel Colodel, Lucas Pereira, Vitor Bertoli	44-51
6	Uso de Parâmetros de Curvas Exponenciais para Classificação do Desempenho em Sprints de 30 Metros em Atletas de Futebol Sub-20 Lucas Pereira , Gabriel Colodel, Pedro Menezes, Vitor Bertoli Nascimento	52-60
7	Clusterizando Jogadores de Futebol em Diferentes Arquétipos Lucca Augusto, Jônatas dos Santos, Matheus Irias, Rodrigo Ferreira	61-67

Sumário - Resumo

Resumo	Título do Trabalho	Pág.
1	GoldFit Soccer: Sistema Especialista de Identificação de Talentos no Futebol Francisco Z. Werneck, Emerson F. Coelho, Marcelo O. Matta, Antônio J. Figueiredo, Rodrigo C.P. Silva	68-71
2	Comparação entre a formação de base no futebol de elite masculino e feminino no Brasil Guilherme Machado, Sixto González-Víllora, Israel Teoldo	72-76
3	Rankeando Jogadores de Futebol: Criação de um Índice de Performance Baseado em Aprendizado de Máquina José Américo Pereira Antunes	77-81
4	Importância da Avaliação Tridimensional do Chute e da Corrida em Atleta Profissional - Estudo de Caso Liria Okai-Nóbrega, Lucas Freitas, Vanessa Araújo, Diego Carvalho, Sabrina Cintra, Priscila Araújo, Renatha Carvalho, Aline Castro, Renata Frederico, Sérgio Teixeira da Fonseca, Thales Rezende de Souza	82-85
5	As estratégias biomecânicas de tronco e membro inferior para o chute máximo do futebol são complementares? Lucas de Freitas, Diego Carvalho, Sabrina Cintra, Gabriela Blaso, Brenda Gomes, Pedro Henrique Mello, Sérgio Teixeira da Fonseca, Juliana Ocarino, Priscila Araújo, Liria Okai-Nóbrega, Thales Rezende de Souza	86-90
6	Efeito do Treinamento Resistido de Músculos do Tronco e do Membro Superior na Cinemática do Chute do Futebol Miguel Arcanjo de Assis, Thiago Santos, Sergio Teixeira Fonseca, André Gustavo de Andrade, Priscila Araújo, Liria Akie Okai-Nóbrega, Thales Rezende de Souza, Renan Resende, Juliana Ocarino	91-93
7	Estudo sobre sprints em jogadores da Série B: Momentos e características Pedro Monteiro, Lucas Vinicius, Filipe Lourenço e Pedro Meneses	94-98
8	Utilizando o OpenPose para análise cinemática tridimensional do salto de goleiros de futebol no campo: efeito da lateralidade no pênalti Rafael Luiz Martins Monteiro, Paulo Roberto Pereira Santiago	99-102
9	Efeitos imediatos de uma órtese elástica no tronco na biomecânica e no desempenho do chute em praticantes de futebol: um estudo de viabilidade Sabrina P. Cintra, Sergio T. Fonseca, Diego C. Souza, Lucas V. Freitas, Gabriela M. Blaso, Luísa P. Antunes, Fabrício A. Magalhães, Priscila A. Araújo, Liria A. Nobrega, Thales R. Souza	103-106

Predição de resultados de jogos do Campeonato Brasileiro de Futebol Feminino

Ananda Mendes Souza¹, Anderson A. Ferreira¹, Renato L. Moreira²

¹Departamento de Ciência da Computação – Universidade Federal de Ouro Preto (UFOP)
Campus Morro do Cruzeiro – Ouro Preto – MG – Brasil

²Escola de Educação Física – Universidade Federal de Ouro Preto (UFOP)
Campus Morro do Cruzeiro – Ouro Preto – MG – Brasil

ananda.mendes@aluno.ufop.edu.br

{anderson.ferreira, renatolmoreira}@ufop.edu.br

Abstract. In the sports landscape, despite the notable growth in women's football, challenges persist in interpreting match results. This study aims to provide valuable insights into team performance, considering the influence of various factors. Utilizing Footystats data encompassing the 2022 Brazilian Women's Football Championship (134 matches), our methodology involves exploratory analyses and the construction of prediction models. Current analyses underscore the significance of a comprehensive database, preprocessing, and a careful selection of attributes for achieving more accurate results.

Resumo. No cenário esportivo, embora o futebol feminino tenha testemunhado um notável crescimento, persistem desafios na interpretação dos resultados das partidas. Este trabalho busca proporcionar informações valiosas sobre o desempenho das equipes, considerando a influência de diversos fatores. Com base em dados do Footystats, abrangendo os jogos do Campeonato Brasileiro de 2022 (134 partidas), a metodologia inclui análises exploratórias e a construção de modelos de predição. Até o momento, as análises ressaltam a importância de ter uma base de dados completa, pré-processamento e uma seleção criteriosa de atributos para alcançar resultados mais precisos.

1. Introdução

O futebol feminino tem ganhado cada vez mais destaque e reconhecimento nos últimos anos. Porém ainda enfrenta desafios em termos de análise de dados e previsão de resultados. A falta de informação disponível sobre partidas e jogadoras dificulta a criação de modelos precisos para prever o desempenho de equipes ou jogadoras individuais por analistas e treinadores. Além disso, a maioria dos bancos de dados disponíveis se concentra no futebol masculino, e apenas algumas plataformas começaram a incluir dados completos (Estatísticas dos campeonatos, das partidas, dos times e das jogadoras) do futebol feminino recentemente. No entanto, como afirma [Detoni 2022], o futebol feminino no Brasil é frequentemente subestimado e recebe pouca atenção da mídia e do público em geral. Isso resulta em recursos limitados disponíveis para coletar e analisar dados do futebol feminino.

A partir de 2019, observou-se uma melhora significativa no cenário do futebol feminino, impulsionada pelo sucesso da Copa do Mundo da França. Nesse ano, ocorreram

marcos importantes, como a inclusão de um grande número de clubes tradicionais no futebol masculino no cenário feminino, devido à obrigatoriedade de equipes femininas para os times masculinos da Série A do Brasileirão e da Libertadores. Além disso, houve o retorno das transmissões do Campeonato Brasileiro feminino na televisão [Nunes 2023]

A partir de 2022, o Brasileirão feminino passou a contar com três divisões, e o calendário nacional foi ampliado para incluir a Supercopa feminina. Ao longo desses últimos quatro anos, muitos clubes tradicionais no futebol masculino consolidaram sua presença na elite do futebol feminino. Atualmente, a Série A1 é composta por 13 times de renome, como Corinthians, Palmeiras, Flamengo, Cruzeiro, Santos, Internacional, Grêmio, São Paulo, Bahia, Athletico-PR, Atlético-MG, Avaí/Kindermann e Ceará. Mesmo que nem sempre o peso da camisa seja revertido em valorização e profissionalização para o feminino [Nunes 2023].

Com o crescimento do futebol feminino no país, tem havido um aumento significativo na quantidade de dados disponíveis referentes às competições e times. Atualmente, plataformas reconhecidas, inclusive sites de apostas, passaram a oferecer cobertura do Campeonato Brasileiro Feminino e outras competições [Estadao 2022]. Embora ainda haja lacunas e áreas a serem aprimoradas, esse avanço representa um importante começo para a área de análise de dados no futebol feminino, visto que a inclusão de plataformas com cobertura do futebol feminino proporciona aos analistas, treinadores e fãs acesso a estatísticas. Isso possibilita uma análise mais detalhada do desempenho das equipes e o desenvolvimento de modelos preditivos mais precisos [Rosa 2022].

Apesar da crescente disponibilidade de dados, ainda são escassos os estudos na área de análise de dados do futebol feminino no Brasil [Peconick 2018]. Mais especificamente, a utilização desses dados para desenvolver modelos de previsão ainda é um campo pouco ou quase nada explorado na modalidade feminina. Essa lacuna representa uma oportunidade para a expansão do conhecimento e o avanço da análise de dados no futebol feminino brasileiro. A criação de trabalhos dedicados à predição de resultados pode contribuir para identificar áreas em que as equipes precisam melhorar para competir em um nível mais alto. Isso pode ajudar os treinadores e jogadores a identificar áreas de fraqueza e desenvolver estratégias para melhorar o desempenho.

Buscando resolver essa problemática, neste trabalho será abordado a temática das previsões de resultados, por meio de aprendizado de máquina para jogos do Campeonato Brasileiro de Futebol Feminino série A, utilizando métodos de aprendizado de máquina que combinam vários modelos individuais em um modelo mais forte. Essa técnica visa melhorar a precisão e a estabilidade dos modelos de aprendizado de máquina, reduzindo o risco de *overfitting* (sobreajuste) e melhorando a precisão do modelo [Dietterich 2000]. A originalidade reside na exploração de técnicas de aprendizado de máquina treinadas com outras ligas do futebol feminino e aplicadas especificamente ao Campeonato Brasileiro de Futebol Feminino Série A, considerando as particularidades e desafios únicos dessa modalidade esportiva.

O objetivo principal deste trabalho é desenvolver e aplicar uma abordagem de previsões de resultados de jogos utilizando aprendizado de máquina para o Campeonato Brasileiro de Futebol Feminino, com o intuito de melhorar a análise de dados e fornecer percepções estratégicas para treinadores, analistas e gestores.

Portanto, este trabalho pode contribuir para o avanço da análise de dados no esporte e ajudar a preencher a lacuna existente na falta de informações sobre o desempenho das equipes e jogadoras do futebol feminino. Além disso, pode ajudar a identificar tendências e padrões de desempenho e permitindo decisões mais informadas. O uso de modelos de aprendizado de máquina também pode aumentar a visibilidade da modalidade e atrair investimentos e recursos para a sua melhoria.

2. Referencial Teórico

Conforme abordado anteriormente, há uma lacuna na literatura no que diz respeito aos estudos no âmbito do futebol feminino. Inicialmente, é realizada uma análise dos estudos de aprendizado de máquina voltados para o futebol masculino. Esses trabalhos se destacam pela abrangência, uma vez que contam com uma base de dados mais robusta, devido a quantidade de dados disponíveis, comparada ao futebol feminino. As principais variações encontram-se nos algoritmos e campeonatos utilizados durante sua concepção, garantindo uma abordagem mais completa sobre o tema. Segundo [Ganhor et al. 2020], apesar da quantidade de artigos disponíveis ser reduzida, é evidente um aumento relativo no interesse por essa linha de pesquisa nos últimos anos. Esse crescimento evidencia uma maior atenção e dedicação dos pesquisadores nesse campo específico, indicando o reconhecimento da importância das previsões de resultados no futebol e seu potencial para aprimorar a compreensão e o desempenho no esporte.

Os estudos realizados em [da Silva 2018] e [Silva 2022] abordaram o Campeonato Brasileiro Série A masculino de 2016 e os campeonatos entre 2003 e 2020, respectivamente. O segundo estudo se destaca por ter um conjunto de dados substancialmente maior, permitindo uma análise mais abrangente e detalhada. Ambos os trabalhos coletaram uma variedade de informações estatísticas relacionadas às partidas, incluindo passes, chutes e gols, que foram utilizadas como variáveis na análise. No primeiro estudo, foi necessário realizar um tratamento específico nos dados, aplicando a média de gols, uma vez que houve uma variação no formato do campeonato ao longo do tempo (de 24 para 20 clubes). Utilizando a regressão linear múltipla como ferramenta estatística, o primeiro trabalho obteve um R^2 de 93,09%, demonstrando alta correlação entre as variáveis analisadas e o resultado final do Campeonato Brasileiro de 2017. Por sua vez, o segundo trabalho alcançou uma taxa de acerto de 100% na previsão das posições das equipes, mas apresentou uma precisão de apenas 35% na previsão das pontuações. Esses resultados evidenciam a importância crucial da quantidade de jogos analisados, destacando-a como um fator determinante na qualidade das previsões. Assim, ambos os estudos contribuem significativamente para o aprofundamento e compreensão desse tópico específico no campo do aprendizado de máquina aplicado ao futebol.

Outra técnica empregada para inferir resultados das partidas é o uso de árvores de decisão. Um exemplo é o trabalho realizado por [Silva et al. 2020], que utilizou um modelo estatístico baseado em árvore de decisão para prever resultados de jogos de futebol em diferentes campeonatos ao redor do mundo durante o ano de 2019, incluindo o Brasileirão Série A e a Superliga Argentina de Futebol. Além da aplicação de árvore de decisão, a pesquisa utilizou técnicas de avaliação, como a validação cruzada e análise descritiva, como ferramentas estatísticas para análise de dados e resultados. Após o estudo, constatou-se que o modelo proposto acertou 57,5% dos jogos, o que sugere que a modelagem por meio de árvores de decisão é uma boa abordagem para a previsão de resultados

de jogos de futebol, uma vez que alcança resultados comparáveis à literatura.

A previsão de resultados de partidas também pode ser aplicada ao futsal. O estudo conduzido por [Duarte and Coppini 2021] utilizou técnicas de aprendizado de máquina para prever os resultados de jogos da Liga Nacional de Futsal (LNF)¹, com base nos dados gerados durante o primeiro tempo das partidas. Foram desenvolvidos dez modelos de previsão, sendo eles: *Random Forest* (RF) [Breiman 2001], *Gradient Boosting* [Friedman 2002], Regressão Logística (RL) [Cox 1972], *K-Nearest Neighbors* (KNN) [Cover and Hart 1967], *Support Vector Machine* (SVM) [Vapnik 1995], *Decision Tree* (DT) [Quinlan 1986], *Naive Bayes* [Bayes 1968], *Multi-Layer Perceptron* (MLP) [Rumelhart 1986], Análise Discriminante Linear (ADL) [Fisher 1936] e Análise Discriminante Quadrática (ADQ)[Fisher 1936]. Os resultados obtidos demonstraram que os modelos individuais apresentaram um bom desempenho na previsão de resultados específicos, como a vitória do time da casa, alcançando uma acurácia de até 95%. Além disso, foi criado um comitê de votação que combinava as previsões dos diferentes modelos. Esse comitê obteve um desempenho superior na previsão dos resultados gerais, alcançando uma acurácia de quase 79%. Esses resultados indicam que a combinação de vários modelos leva a uma resposta mais precisa no contexto geral.

O uso de modelos combinados também foi usado por [Schneider 2020], com o objetivo de avaliar o desempenho de diversos algoritmos de classificação quando os mesmos são utilizados para prever resultados de partidas de futebol contidas entre as temporadas 2000-01 e 2016-17 da Premier League. Para alcançar esse objetivo, o estudo avaliou o desempenho de vários algoritmos de classificação, incluindo Regressão Logística, Análise Discriminante Linear, Análise Discriminante Quadrática, *K-Nearest Neighbors*, *Naive Bayes*, *Support Vector Machine* com kernel linear (SVML), *Support Vector Machine* com kernel RBF (SVMR), *Random Forest*, *Extra Trees* (ET) [Geurts et al. 2006] e um classificador *ensemble* com classificadores que possuem distribuições de previsões de instâncias significativamente diferentes. Como métrica de avaliação de desempenho dos algoritmos, utilizou-se a acurácia, *F1-score* e análises da matriz de confusão. Ao se observar a matriz de confusão gerada através da análise de um conjunto de dados de teste, constatou-se que a acurácia do classificador ensemble foi 1% maior (57%) do que a constatada nos classificadores individuais (56% - SVML e KNN). Assim, a utilização de um ensemble de classificadores pode apresentar-se como uma alternativa viável para aumentar a robustez e capacidade de generalização de um modelo de previsões.

No que se refere ao futebol feminino, [Leitner et al. 2020] propõem uma abordagem híbrida para prever os resultados da Copa do Mundo Feminina da FIFA 2019, usando a técnica de floresta aleatória combinada com dois métodos diferentes de classificação: o método de classificação de *Poisson* [Karlis and Ntzoufras 2003] e as habilidades baseadas nas probabilidades dos *bookmakers* (empresas que oferecem serviços de apostas em eventos esportivos e outros eventos). O modelo híbrido foi ajustado a todos os jogos da Copa do Mundo Feminina de 2011 e 2015² e, com base nas estimativas resultantes, a Copa de 2019 foi simulada 100.000 vezes para determinar as probabilidades de vitória para todas as 24 equipes participantes. Os resultados mostraram que o modelo híbrido superou outros modelos de previsão em termos de precisão na previsão dos resultados dos jogos da

¹<https://lnficial.com.br>

²<https://www.fifa.com/fifaplus/pt/home>

Copa do Mundo Feminina da FIFA 2019. A abordagem proposta combina informações sobre as equipes, incluindo co-variáveis e habilidades estimadas, para melhorar as previsões. A validação cruzada foi usada para avaliar a acurácia do modelo híbrido proposto usando medidas de desempenho, como precisão e erro médio absoluto.

Este estudo compartilha semelhanças com trabalhos anteriores ao utilizar o aprendizado de máquina para desenvolver um modelo de previsão de resultados no futebol. No entanto, se destaca por focar exclusivamente no contexto do futebol feminino brasileiro, proporcionando uma perspectiva única e relevante para essa modalidade. Além disso, este trabalho oferece um ambiente propício para futuras pesquisas na aplicação da predição de resultados de jogos no campo esportivo, especialmente para equipes e setores de análise de desempenho, permitindo estudos específicos e personalizados para a realidade do futebol feminino no país. Em suma, esta pesquisa contribui de forma complementar, expandindo e aprimorando os resultados já existentes na aplicação do aprendizado de máquina no futebol feminino, oferecendo uma valiosa perspectiva no contexto brasileiro.

3. Metodologia

A pesquisa teve início com a análise de várias plataformas de dados que abrangem informações relacionadas ao Campeonato Brasileiro feminino, com o propósito de selecionar aquela que melhor se adequasse aos objetivos deste estudo, sendo escolhida a plataforma Footystats³. Na sequência, visando garantir a qualidade dos resultados e devido a pequena quantidade de jogos do Campeonato Brasileiro feminino disponível, foi escolhido 5 ligas para formar a base de dados: *National Women's Soccer League* (EUA), *FA Women's Super League* (Inglaterra), *Serie A Woman* (Itália), Liga F (Espanha), *UEFA Women's Champions League*. Posteriormente, foi feita a análise e pré-processamento dos dados, que inclui a eliminação de atributos irrelevantes, o tratamento de valores ausentes, a normalização dos valores dos atributos e a adição do atributo com o resultado da partida.

No processo de implementação, realizou-se a seleção de atributos com base em critérios bem definidos, estabelecendo as bases necessárias para o desenvolvimento da pesquisa. Dessa forma as jogadoras foram divididas conforme a posição (goleira, defensora, meio-campo ou atacante) e suas características específicas, em seguida, fez-se a média dos valores por time para compor a base de dados de treinamento.

Assim, foi compilado um conjunto de dados composto por 780 jogos referentes a temporada de 2022/23, com mais de 1300 atributos.

Os atributos utilizados podem ser categorizados da seguinte forma:

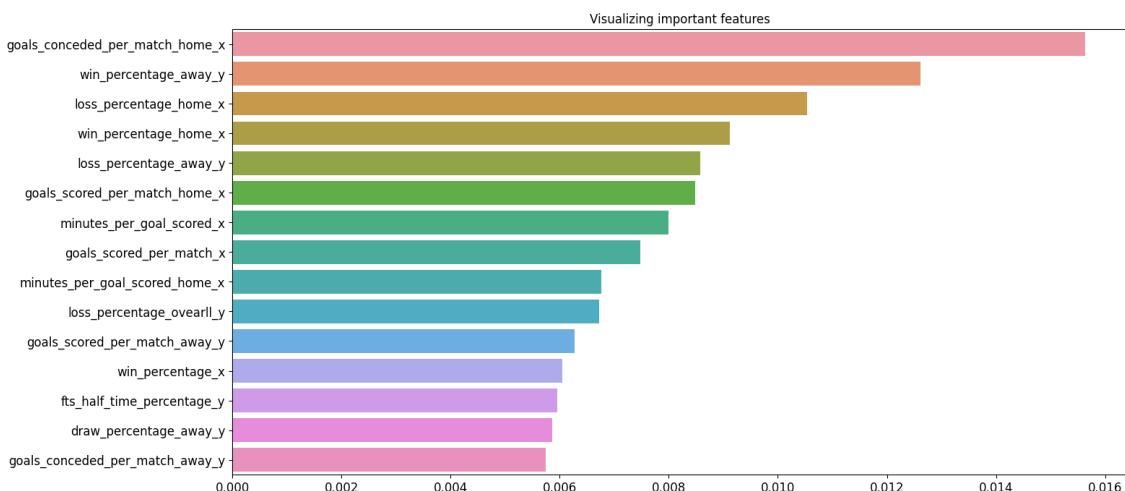
1. Atributos de Desempenho das Equipes: Estes atributos descrevem o desempenho histórico das equipes no campeonato. Eles incluem estatísticas como média de gols marcados e sofridos, quantidade de vitórias e derrotas, posição na tabela de classificação, entre outros.
2. Atributos de Condições de Partida: Inclui informações sobre os jogos, como os nomes dos times envolvidos, médias pré-jogo de pontos por partida para cada equipe, valores pré-jogo de expectativa de gols calculados com base no desempenho anterior, e outros fatores relevantes que afetam as condições de cada partida.

³<https://footystats.org/pt/>

3. Atributos de Desempenho Individual: Relacionados às estatísticas individuais das jogadoras, como número de gols, assistências, cartões amarelos, cartões vermelhos, e outros indicadores que destacam o desempenho individual durante a temporada.
4. Atributos de Tendências Temporais: Envolve dados que destacam tendências temporais ou sazonalidades, como a evolução ao longo do campeonato, variações no desempenho das equipes ou jogadores em diferentes fases do torneio, entre outros.

Para otimizar o conjunto de atributos, realizou-se uma análise, testando diferentes configurações, e a seleção final incluiu os 200 melhores atributos obtidos por meio de correlações e do método de identificação de *features* importantes. Alguns dos atributos são apresentados (Figura 1).

Figura 1. Os 15 melhores atributos selecionados



Fonte: Autora

Adicionalmente, foram treinados classificadores para prever vitória, empate ou derrota de um time. O treinamento foi realizado buscando os melhores valores de parâmetros para cada técnica de indução de classificadores, via *Grid Search*. A avaliação dos classificadores foi feita por meio da validação cruzada.

O fluxograma apresentado (Figura 2) esquematiza a metodologia adotada, destacando as etapas cruciais do processo.

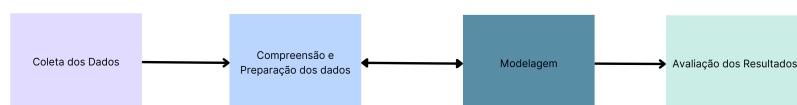


Figura 2. Metodologia utilizada neste trabalho

4. Resultados Iniciais

Para avaliar o desempenho preditivo sobre os resultados das partidas com os classificadores obtidos, usou-se a validação cruzada com 5 *folds*. As técnicas de indução de

classificadores utilizadas foram: *Random Forest*, *XGBoost*, KNN (*K-Nearest Neighbors*) [Zaki and Meira 2020].

A Tabela 1 mostra a acurácia média e o *F1-Score* obtidos por cada classificador, juntamente com o desvio padrão de ambos.

Tabela 1. Resultados dos classificadores

Classificador	Acurácia	Desvio Padrão	<i>F1-Score</i>	Desvio Padrão
RF	0.6136	0.090	0.5974	0.0910
KNN	0.4166	0.0690	0.3568	0.0676
XGB	0.5075	0.0897	0.4568	0.0898

Esses resultados destacam os desafios inerentes à tarefa de prever resultados de jogos de futebol. A acurácia média próxima a 50% ilustra a imprevisibilidade característica desses eventos esportivos e está de acordo com resultados obtidos por outros trabalhos [Schneider 2020], [Leitner et al. 2020].

A matriz de confusão (Figura 3), evidencia a notável dificuldade do modelo em prever empates (classe 0).

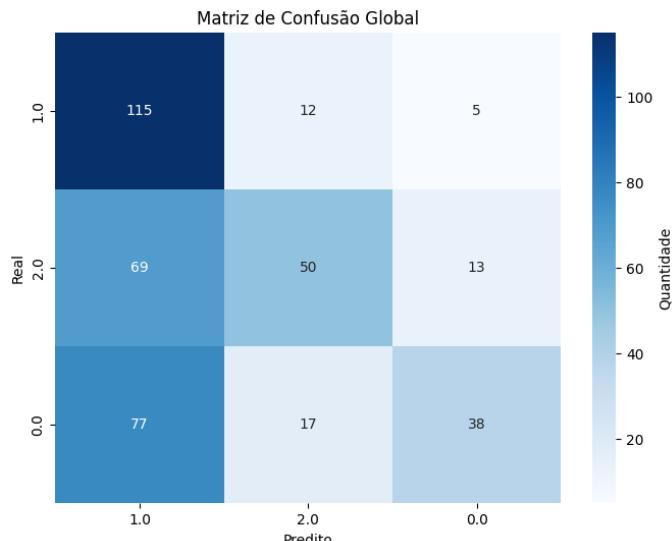


Figura 3. Matriz de Confusão dos Resultados

Ao ser submetido a testes em outras ligas, o modelo demonstrou resultados levemente semelhantes (veja Tabela 2). As discrepâncias observadas podem ser atribuídas à incompletude do conjunto de dados do Campeonato Brasileiro Feminino em comparação com as outras ligas testadas.

A interpretação combinada da acurácia, *F1-score* e do desvio padrão evidenciam uma consistência nos desempenhos dos classificadores.

5. Conclusão e Próximos passos

O presente estudo aborda a predição de resultados de jogos do Campeonato Brasileiro Feminino de Futebol com uma abordagem fundamentada em aprendizado de máquina.

Tabela 2. Melhores Resultados dos classificadores comparando com outras bases de dados

Base de dados	Acurácia	<i>F1-Score</i>
Brasileiro Feminino 2022	0.5125	0.4703
Brasileiro Masculino 2022	0.4745	0.4307
NSWL 2023	0.4135	0.3991

Os classificadores avaliados tiveram acurácia média variando de 41% a 61%, e desvio padrão variando de 6,9% a 9%. Entre os classificadores avaliados, o baseado em *Random Forest* obteve a maior acurácia média, embora com um desvio padrão um pouco mais alto, sugerindo previsões mais precisas, com variação moderada.

Em síntese, esta primeira fase do estudo ofereceu um panorama revelador sobre a complexidade da previsão de resultados em jogos de futebol. Ao identificar questões críticas enfrentadas pelos classificadores, estabeleceu-se uma sólida base para a fase subsequente da pesquisa.

Os próximos passos visam aperfeiçoar a habilidade preditiva dos modelos, como por exemplo, lidar com o volume reduzido de dados disponíveis devido ao alto número de atributos nulos, utilizar um *baseline* diferente dos métodos implementados e avaliar a capacidade preditiva do modelo em mais ligas.

6. References

Referências

- Bayes, T. (1968). Naive bayes classifier. *Article Sources and Contributors*, pages 1–9.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.
- da Silva, B. M. (2018). Multiple linear regression applied to football/ regressao linear multipla aplicada ao futebol. *Revista Brasileira de Futsal e Futebol*, 10(38):262.
- Detoni, H. O. (2022). Futebol feminino brasileiro e as dificuldades encontradas nesse subcampo esportivo.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. *Multiple Classifier Systems*, 1857:1–15.
- Duarte, R. and Coppini, J. (2021). Utilizando abordagens de aprendizado de máquina para prever resultados de jogos: o caso da liga nacional de futsal. *Revista Brasileira de Futsal e Futebol*, 13(53):275–283.
- Estadao (2022). Futebol feminino tem crescimento de mais de 130% em apostas esportivas em um ano. Disponível em: <https://www.estadao.com.br/esportes/futebol/>

apostas-esportivas-em-competicoes-de-futebol-feminino-crescem-mais-de
Acesso em 15 de agosto de 2023.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378.

Ganhor, J. P., Dalmazo, J. F., Morosini, J. A. S., Krombauer, M. E., and Celarino, A. L. S. (2020). Predição de resultados para partidas de futebol: um olhar para a produção científica em periódicos nacionais. *Revista Brasileira de Futsal e Futebol*, 18(72):58–67.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63:3–42.

Karlis, D. and Ntzoufras, I. (2003). Analysis of sports data by using bivariate poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):381–393.

Leitner, C., Zeileis, A., and Hornik, K. (2020). A hybrid random forest approach for predicting the results of the fifa women’s world cup 2019. *Journal of Applied Statistics*, 47(9):1875–1893.

Nunes, M. (2023). O que mudou no futebol feminino do brasil desde a última copa? Dibradoras website. Disponível em: <https://dibradoras.com.br/2023/03/22/o-que-mudou-no-futebol-feminino-do-brasil-desde-a-ultima-copa/>. Acesso em: 17 de julho de 2023.

Peconick, L. D. F. (2018). Inteligência artificial aplicada à previsão de jogos de futebol. Disponível em: https://bdm.unb.br/bitstream/10483/23782/1/2018_LauraDefrancoFerreiraPeconick.pdf.

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1:81–106.

Rosa, P. H. C. (2022). Mineração de dados aplicada a previsão de resultados de jogos de futebol. Disponível em: <https://repositorio.pucgoias.edu.br/jspui/bitstream/123456789/5163/1/TCC%20PEDRO%20HENRIQUE%20CANDIDO%20ROSA.pdf>.

Rumelhart, D. (1986). David e. rumelhart, geoffrey e. hinton, and ronald j. williams learning representations by back-propagating errors nature 323: 533-536. *nature*, 323:533–536.

Schneider, C. F. (2020). Aplicação de técnicas de machine learning para previsão de resultados de partidas de futebol. Projeto de diplomação em Engenharia Elétrica.

Silva, A. B. L., de Oliveira Barros, K. N. N., and Albuquerque, M. A. (2020). Modelagem via árvore de decisão para previsão de jogos de futebol. *Research, Society and Development*, 9(9):e204996869–e204996869.

Silva, B. M. d. (2022). Modelo preditivo aplicado ao futebol brasileiro. *RBFF - Revista Brasileira de Futsal e Futebol*, 14(58):291–297.

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer science & business media.

Zaki, M. J. and Meira, W. (2020). *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.

Analisando Robustez da Rede de Passes através da Dispersão de Hub e Authority

Balthazar Paixão^{a,*}, Lucas Giusti Tavares^a, Pedro Henrique González^b,
Glauco Fiorott Amorim^a

^a*PPCIC - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca,
CEFET/RJ, , Rio de Janeiro, 20271-110, RJ, Brazil*

^b*PESC - Universidade Federal do Rio de Janeiro, UFRJ/RJ, , Rio de
Janeiro, , RJ, Brazil*

Keywords: Complex Networks, Football/Soccer, Network Science

1. Resumo

O futebol, consagrado como o esporte mais popular do mundo, atrai bilhões de espectadores a cada ano [3, 7]. A capacidade de antecipar movimentos durante as partidas ou compreender os fatores determinantes para um gol são aspectos que fascinam torcedores, jogadores e profissionais técnicos.

A análise de dados no futebol é desafiadora devido à profusão de eventos em cada partida [2]. Uma estimativa sugere que, em um único tempo de jogo (45 minutos), um treinador consegue se recordar de apenas um terço dos lances ocorridos [2]. As análises esportivas, entretanto, têm o potencial de elucidar aspectos do jogo que podem passar despercebidos aos olhos dos analistas [11].

Apesar dos desafios, a busca por análises mais aprofundadas e o desenvolvimento de novas metodologias são imperativos para as equipes. Alguns focos de observação incluem a análise de passes, sequência de passes, agrupamentos e movimentos coletivos [4]. Além disso, há também uma categorização de

*Corresponding author.

Email addresses: balthazar.paixao@eic.cefet-rj.br (Balthazar Paixão), lucas.giusti@eic.cefet-rj.br (Lucas Giusti Tavares), pegonzalez@cos.ufrj.br (Pedro Henrique González), glauco.amorim@eic.cefet-rj.br (Glauco Fiorott Amorim)

táticas, como defensivas, ofensivas, individuais, coletivas e sistemas de jogo [8].

Uma vertente em ascensão nos estudos recentes é a aplicação da Ciência de Redes no futebol [1, 9, 10]. Esta abordagem visualiza os jogadores como vértices e suas interações, sejam passes ou proximidade, como arestas. Essa representação forma uma rede dinâmica que evidencia a relação entre os jogadores dentro de uma equipe e em relação ao adversário. Através dessa perspectiva, é possível extrair métricas como grau, centralidade, distância e detecção de comunidades, fornecendo *insights* valiosos sobre as nuances do jogo e as características das equipes.

Uma métrica interessante para analisar uma equipe é detectar a robustez da rede de passes que a mesma produz em um jogo. A ideia é que quanto mais robusta, mais tempo a equipe passa com a bola em todas as fases do jogo e, por consequência, mais próxima de marcar um gol está. Ichinose et al. [5] utilizaram a métrica para analisar o desempenho de equipes no campeonato japonês. Apesar de acharem uma correlação entre a métrica e o desempenho, eles consideraram a rede de passes como um grafo não dirigido.

Neste contexto, o presente trabalho tem por objetivo avaliar uma métrica que utiliza a representação da rede de passes por um grafo dirigido e usa a distribuição dos graus de entrada e saída dos jogadores para capturar a robustez de uma equipe. A métrica é aplicada em diferentes equipes ao longo de um campeonato, buscando correlacionar a robustez com as posições finais das equipes após o término do campeonato.

1.1. Metodologia

Nossa abordagem metodológica foca na construção de uma rede de passes entre os jogadores titulares de uma partida com o objetivo de compreender o nível de conectividade de um jogador em relação à rede como um todo. A partir dessa visão foi possível aplicar duas métricas para cada jogador: *Hub* e *Authority*(Autoridade). Essas métricas são variações das métricas apresentadas em [6].

1.1.1. Definição e Cálculo das Métricas

Neste trabalho, vamos considerar uma rede robusta quando todos os seus nós estão interconectados de maneira uniforme. Nesse cenário, a distância entre quaisquer dois nós é minimizada, garantindo que a rede seja resiliente a perturbações em um nó específico. Com base nesse entendimento, introduzimos duas métricas para avaliar a conectividade da rede estudada:

Distribuição da Conectividade de Hubs e Distribuição da Conectividade de Autoridades.

No contexto estudado, um "Hub" é definido como um jogador do qual, dada uma determinada ação, há uma probabilidade significativa de que a bola tenha partido dele. Por outro lado, "Autoridade" refere-se ao oposto: dada uma ação, qual a probabilidade de a bola ter sido direcionada a esse jogador. Ambas as métricas atribuem uma probabilidade específica a cada jogador.

A equação 1 apresenta a fórmula para a distribuição de Hubs, onde H_i é a pontuação de Hub para o jogador i , $Dout_i$ se refere ao grau de saída do jogador i , e $NDout$ à soma de todos os graus de saída da rede. Analogamente, a ref2 executa a mesma lógica se diferenciando que os graus analisados são os de entrada em um nó; A_i é a pontuação de Autoridade para o jogador i , Din_i se refere ao grau de entrada do jogador i , e $NDin$ à soma de todos os graus de saída da rede.

$$H_i = \frac{Dout_i}{NDout} \quad (1)$$

$$A_i = \frac{Din_i}{NDin} \quad (2)$$

A premissa de que uma rede mais robusta é aquela que está mais interconectada de forma homogênea nos levou à hipótese de que, quanto mais uniformemente distribuídas estiverem as medidas em relação à média, melhor será o desempenho de uma equipe no campeonato. Assim, calculamos a distância interquartil da distribuição de cada medida após cada partida. Quanto mais próximo de zero este resultado estiver, melhor. Finalmente, calculamos a média dessa distância interquartil para as equipes ao longo de todas as partidas, buscando compreender o desempenho geral.

1.1.2. Dados

Para o tipo de análise proposto neste artigo, a preparação adequada dos dados é fundamental, pois o tratamento inadequado pode levar a obtenção de informações falsas e/ou imprecisas. Os dados utilizado neste trabalho são fornecidos gratuitamente pela Statsbomb, disponíveis em seu repositório no GitHub <https://github.com/statsbomb/open-data>. Estes dados contêm informações detalhadas sobre partidas, das quais selecionamos apenas os passes completos realizados por jogadores titulares que não se originaram de bolas paradas. Optamos por focar na Serie A do campeonato italiano durante a

temporada 2015/16. Nossa análise se baseia em uma rede direcionada, pois os passes entre jogadores têm intenções distintas, e desconsiderar a direção poderia levar a interpretações equivocadas.

1.2. Testes e Resultados

Os resultados parciais obtidos são promissores e mostram que a aplicação da métrica desenvolvida pode direcionar o entendimento de futuras análises no âmbito do futebol. Visando entender se as métricas trabalhadas estavam correlacionadas com a distribuição dos times na tabela, fez-se o uso do teste estatístico de Spearman. A Tabela 1 apresenta os resultados da correlação de cada métrica com a posição e a significância estatística utilizando o teste estatístico.

	Hub	Autoridade
Correlação	0.5236	0.6812
Significância	0.017	0.0009

Tabela 1: Resultado da Correlação e Significância das métricas Hub e Autoridade

Pode-se verificar que para a temporada de 2015/16 da Serie A, ambas as métricas de robustez apresentam correlação com a posição dos times na tabela. Este resultado fica claro na Figura 1. Ainda na Tabela 1 pode-se observar que esta correlação possui significância estatística de acordo com o *P-Value*.

A Figura 1 apresenta as pontuações de Hub como pontos em preto e as pontuações de Autoridade em amarelo. No eixo vertical é apresentado o valor da pontuação e no eixo horizontal a posição na tabela. É possível observar que quanto mais distante da primeira posição na tabela mais os times tendem a apresentar um pior resultado da métrica e vice-versa. Vale ressaltar que, como utilizamos a distância interquartil da distribuição de cada medida, quanto menor o valor do resultado, melhor.

	Serie A TIM		La Liga BBVA	
	Hub	Autoridade	Hub	Autoridade
Correlação	0.5236	0.6812	0.6857	0.5248
Significância	0.0170	0.0009	0.0008	0.0175

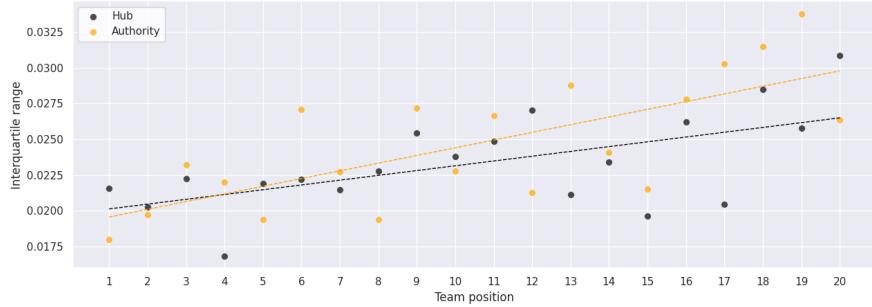


Figura 1: Gráfico representado as medidas de dispersão de Hub e Autoridade para as equipes da Série A do Campeonato Italiano. Os pontos em preto equivalem aos valores de Hub e os pontos em amarelo equivalem ao valor de Autoridade.

1.3. Conclusão

Neste resumo foi apresentada a utilização de uma métrica para avaliar a robustez de uma equipe de futebol e correlacionar esta métrica com o desempenho esportivo. A hipótese é de que quanto mais conectada é uma equipe, mais robusta ela é nas fases do jogo e, por isso, está mais próxima do resultado positivo. Esta interconexão se torna mais forte quando ela é homogênea na equipe. Para avaliar a hipótese de pesquisa foram usadas dois valores: *Distribuição da Conectividade de Hubs* e *Distribuição da Conectividade de Autoridades*. Os resultados dos testes são promissores e mostraram a existência de uma correlação com os valores obtidos e a classificação das equipes no campeonato. A pesquisa encontra-se em fase de evolução e, por isso, pode-se citar como atividades futuras a análise da métrica em outros campeonatos, além de expandir os conceitos de *Hub* e *Autoridade* para verificar quais jogadores e quais posições são mais influenciadas pelos valores.

Referências

- [1] Buldú, J.M., Busquets, J., Martínez, J.H., Herrera-Diestra, J.L., Echegoyen, I., Galeano, J., Luque, J., 2018. Using network science to analyse football passing networks: Dynamics, space, time, and the multilayer nature of the game. *Frontiers in Psychology* 9. URL: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01900>. doi:10.3389/fpsyg.2018.01900.

- [2] Estanislau, A., 2016. Comparação de diferentes sistemas de análise de desempenho no futebol: Proposta na utilização do software. Bachelor's thesis. Universidade Estadual de Campinas, Faculdade de Ciências Aplicadas.
- [3] Félix, L.G.S., Barbosa, C.M., Carvalho, I.A., da F. Vieira, V., Xavier, C.R., 2020. Forecasting soccer market tendencies using link prediction, in: Computational Science and Its Applications – IC-CSA 2020: 20th International Conference, Cagliari, Italy, July 1–4, 2020, Proceedings, Part I, Springer-Verlag, Berlin, Heidelberg. p. 663–675. URL: https://doi.org/10.1007/978-3-030-58799-4_48, doi:10.1007/978-3-030-58799-4_48.
- [4] Gudmundsson, J., Wolle, T., 2014. Football analysis using spatio-temporal tools. Computers, Environment and Urban Systems 47, 16–27. URL: <https://www.sciencedirect.com/science/article/pii/S0198971513000847>, doi:<https://doi.org/10.1016/j.comenvurbssys.2013.09.004>. progress in Movement Analysis – Experiences with Real Data.
- [5] Ichinose, G., Tsuchiya, T., Watanabe, S., 2021. Robustness of football passing networks against continuous node and link removals. Chaos, Solitons & Fractals 147, 110973. URL: <https://www.sciencedirect.com/science/article/pii/S0960077921003271>, doi:<https://doi.org/10.1016/j.chaos.2021.110973>.
- [6] Kleinberg, J.M., 1999. Authoritative sources in a hyperlinked environment. J. ACM 46, 604–632. URL: <https://doi.org/10.1145/324133.324140>, doi:10.1145/324133.324140.
- [7] Li, Y., Mateos, G., 2022. Networks of international football: communities, evolution and globalization of the game. Applied Network Science 7. doi:10.1007/s41109-022-00498-4.
- [8] Moura, F.A., 2011. Análise quantitativa da distribuição de jogadores de futebol em campo durante jogos oficiais. Ph.D. thesis. Universidade Estadual de Campinas. URL: <https://hdl.handle.net/20.500.12733/1616593>.

- [9] Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., Mendes, J., 2011. Networks as a novel tool for studying team ball sports as complex social systems. *Journal of Science and Medicine in Sport* 14, 170–176. doi:10.1016/j.jsams.2010.10.459.
- [10] Rodrigues, D.C.U.M., 2017. Complex network measurements in graph-based spatio-temporal soccer match analysis. Ph.D. thesis. Universidade Estadual de Campinas, Instituto de Computação. doi:10.47749/t/unicamp.2017.987776.
- [11] Rojas-Mora, J., Chávez-Bustamante, F., del Río-Andrade, J., Medina-Valdebenito, N., 2017. A Methodology for the Analysis of Soccer Matches Based on PageRank Centrality. Springer International Publishing, Cham. chapter 16. pp. 257–272. URL: https://doi.org/10.1007/978-3-319-63907-9_16, doi:10.1007/978-3-319-63907-9_16.

Factorial Analysis Model of Football Teams Squads

Gabriel Canda Blanco

Mestrado, Big Data Esportivo – UCAM, Universidad Católica San Antonio de Murcia,
Murcia, Spain

gabriel.candagcb9@gmail.com

Abstract. The creation of an analytical classification model aimed at predicting patterns and trends in the configurations of professional football team squads that lead to optimal sporting performance at a purely competitive level and within the respective league in which each football team is part.

1. Problem description

The case study situational problem brings together multiple conditions and requirements that turn the event into an arduously quantifiable amalgam.

The concept of 'Football', even though in its original beginnings it was a mere game created for the simple purpose of having fun and competing in an idle manner for both players and spectators, has reached a practically indescribable influence volume on the international scene up to quotas not only in sports.

Doing a simple meditation exercise without great fanfare in terms of ambition or mastery of the situation, two global relevance statements could be called that, unified, draw the gigantic monument that represents what many call the 'king sport'.

In the most influential approach in a further population number, even without being directly linked to the passion for this sport, we would talk about the tremendous socioeconomic impact that encompasses the community bordering on the global level, a fact clearly reflected in the jobs generation by LaLiga in the Spanish territory as can be seen from the professional services firm KPMG quote: "Spanish professional football generates 194,381 jobs and a turnover equivalent to the Spanish GDP 1.44%".

The second section is linked to financial investment, an epicenter driving force in football clubs when it comes to increasing the success chances of their respective sporting projects.

This event affects society in general to a greater or lesser extent, although in any case it is a perception especially applied in high-level sports and, with a marked feature in the salary section, even more so in the clubs budgetary calculation, in the players' football services acquisition and, by inherent drag, of their image rights commercially speaking.

The subsequent quote from the international organization FIFA reflects this: "The summer 2023 football transfer market recorded a historical spending record of 7.36 billion dollars".

Firstly, it is essential to understand the intended objective in business terms through the CRISP-DM methodology adoption (Cross Industry Standard Process for Data Mining). All phases are important, but this one allows to know the problem and how to solve it, being able to determine the metrics and algorithms to use, how to evaluate the models and the estimated implementation time.

The emergence originates from the sporting direction of a Spanish football club that requested to develop a machine learning model with which to observe patterns in the configurations of professional football team's squads that lead to optimal sporting performance at a purely competitive level within their respective league.

Through web scraping processes, a historical set of data from the 'TransferMarkt' and 'FBref' websites was available and, by predicting whether or not a team would achieve that performance, they could deduce conclusions regarding how to approach the design of the team's squad configuration and have greater advantage and context in the transfer market windows approaches, reinforcing decisional conviction, saving time and reducing long-term risk term on a sustained basis, inspired by the motto: "Unity makes strength".

2. General objective

The idea constitution that has a significant meaning, establishing itself as a value proposition, is structured in a foundational pillars trio that give essence and corporeality

to the project intrinsic desire itself. The cornerstones are summarized in the subsequent elementary terms, such as 'Mission', 'Vision' and, finally, 'Values'.

In reference to the initiative mission, the aim is to identify patterns and trends in the configurations of professional football team squads that lead to optimal sporting performance at a purely competitive level and within the respective league in which each football team is part.

Regarding the objective prism that promotes the vision, framing each analysis in the situational context in specificity as an evaluative measure of positive performance is essential to formalize a study with the degree of further plausibility with reality.

In reference to the constructive culmination that values provide, recognizing the essential factors weighted by virtue of their influence on increasing or decreasing the probabilities of sporting success supposes the emergence of applicable competitive advantages, with an eminently relational trait with the own and unique casuistries of each interested actor.

The synergistic casting of each axis of the fundamental bases trident by which the analytical candidacy motivations and intentions provided are supported, leads to a study exercise in constant evolution and improvement with a nuance of entrepreneurship of actions in sporting frameworks, by virtue of reinforce decisional conviction, the sustainability of long-term trust and, with a collaborative and disruptive perspective, the formation of purposeful and pioneering initiatives in the data democratization, cooperation to achieve a common good and the invention of new perspectives.

3. Implemented methodology

Every approach, in tune with its adhered significant value proposition, demands the fundamental presence of a methodological strategy as a means to achieve said ultimate goal.

3.1. Continuous execution cycle

The design of the action plan is essential in achieving emphasis on fair and required actions, avoiding inefficient use of the available time space and perpetrating recurring failures that block each aspect of the process.

The preliminary phase focuses on acquiring an understanding of the dimension of the specific situation as a global contextualization. This attitude allows us to understand and be able to describe the situational framework in which the project is placed.

This beginning will lay the foundations for potential success in the end, facilitating the complex and changing creative process that does not always make it possible to create great value propositions without having previously laid the contextual foundations.

The second stage is therefore defined as the moment of greatest possible feeling of inspiration in order to clarify the previous intention pursued by drawing the conceptual plane sketch of the idea that is intended to be achieved.

At the moment in which a concrete atmosphere has been generated in context and a series of visual references of the intended future scenario have been designed, the procedure requests commitment and determination in the concretization of the previous intention as the process third chapter.

This necessary dimensional transfer, starting from the conceptual level of a more theoretical nature to the realization level of the previously formulated aspirations, is probably the most complex and controversial step of all.

The exchange between the purely theoretical essence to the creation of material and applicable knowledge is the foundation of the reduction of limiting and unproductive barriers and feelings of procrastination, alluding to the proactive and purposeful spirit.

Maintaining an optimistic and flattering sense in each venture started, would be excessively decontextualized from reality if it is believed that the concretization phase conclusion represents the closure of the exposed continuous execution cycle.

Obtaining results closer to what was expected or further from what was really pursued, the adopted methodology fourth station seems significant to sustain a mentality of constant growth and insufficient learning.

This incessant reflective feeling is reflected in the iteration of a systematic process of analysis and adaptation of the results obtained. Standing at this point could be perfectly enough to restart the work cycle in a sustained manner over time and achieve growth.

However, the hunger to want to be relevant in the industry at a more global level, the awakening of that self-interest in each person to go beyond what is currently available, expand horizons and goals, and foster a belonging sense to a community united by virtue of the rise of the industry benefiting all interested parties, requires undertaking the fifth and final step of the cycle described.

The personality of wanting to be a pioneer and innovator in the sporting sector exemplifies the importance of the aforementioned step, thanks to the increasingly voluminous appearance of a collaborative spirit and a disruptive ambition in constant total feedback.

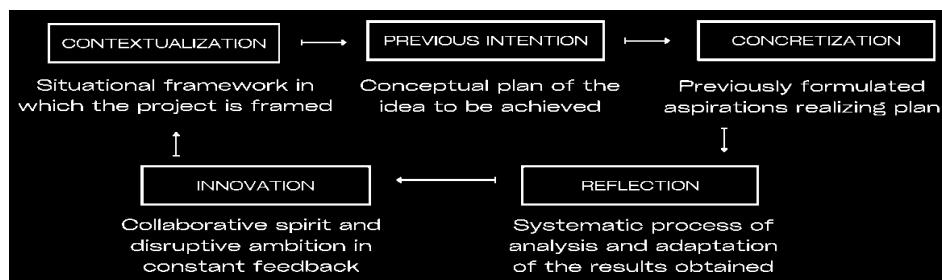


Figure 1. Schematic representation of the continuous execution cycle

3.2. Procedural architecture

The recently developed explanation regarding the methodological essence of the project involved is represented in a clear and simplified manner in a four-band procedural architecture composition that is easily identifiable by placing the gaze lightly and agilely on the subsequent image, divided into four other subsections respectively.

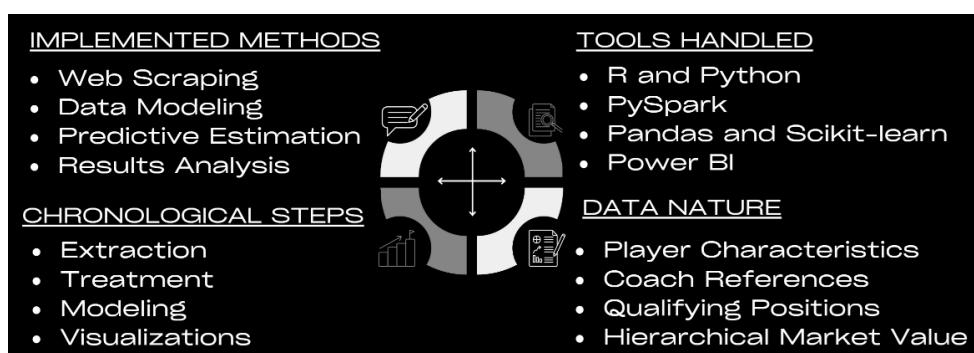


Figure 2. Graphical summary of the procedural architecture

The first to mention lies in the methods used containing the web scraping technique, the data modeling phase, the predictive estimation stage of the analytical model and the results conclusive analysis.

The second position consists of the tools used, including the programming languages R and Python, another language born from the last mentioned one, such as PySpark, the libraries also belonging to Python technology, such as Pandas and Scikit-learn, and, as a purely graphical resource, the Power BI visualization tool.

The third block lists the procedural chronology from the beginning to the end of the project, starting from the extraction of information, continuing with its treatment, following with the modeling phase until, concluding an iteration of the procedural architecture, the visualization moment.

Commented on the fourth instance, the section concerning the nature of the data appears, citing the players characteristics, the coaches references, the qualifying positions and, without downplaying their importance, the hierarchy by market values.

3.3. CRISP-DM analytical model

CRISP-DM (Cross Industry Standard Process for Data Mining) consists of the conception of an open standard process model that describes the common approaches used by data mining experts and provides a data analytics project life cycle description, equivalent to what is represented in software development projects.

Therefore, is the process that will serve as a guide for the data processing and, in this way, be able to work with predictive artificial intelligence techniques.

As a preliminary feature to the in-depth approach to the model development as such, it is convenient to mention the circular structure that is adopted in the applied methodological system.

When conceptualized as a figure with a circular format, the different stages that make up the model will be immersed in a constant iterative circuit.

This incessant loop will start from a solid and clarified base in terms of its complete understanding by virtue of the business model operation and its respective intrinsic goals and specific requirements.

The subsequent stage stands as elementary, endowed with the used adjectivation as it resides in the available information comprehension, being fully aware of the data being handled, its typology and nature, or simply the accessible tables depth.

The third section focuses its efforts on the data preparation and processing from the perspective of taking into account various factors to be considered in its management, from data cleaning to the new metrics generation born from old ones.

Phase four begins the predictive artificial intelligence techniques implementation, which completes three other sub-phases such as the ideal method selection thanks to the data training stages under the different techniques operations until it reaches the optimal model, the predictive capacity improvement by setting the appropriate hyper parameters and adjusting them in order to be moved to the next macro phase.

The suggested subsequent fifth stage confronts the model with a data smaller amount with which it has never trained, in order to test it and see its accuracy at a predictive level.

To this penultimate step, as a way of strengthening the analytical model, a selection strategy is applied to the definitive variables that contain the greatest weight, both positively and negatively in the approach or not to success in terms of sporting performance and, potentially, refine the model prediction skills and capabilities.

The final movement makes possible the option of displaying the obtained model, and in the opposite case being able to return it to the process original point, or even getting currently satisfactory results, but aiming for a predictive capacity closer to the reality. The desire to pursue growth in a constant and permanent proportion validates the meaning of a constant iterative circuit. The next screenshot shows this methodological philosophy.



Figure 3. Circular diagram of the constant iterative circuit

3.4. Algorithmic system implementation

The methodological system is conceptualized by an algorithmic system implementation, starting from a complete business model understanding and its intrinsic specific requirements, as previously mentioned.

The subsequent stage is essential by knowing the available information typology and depth. The match history encapsulates values from the 2018/2019 season to 2022/2023, both included, of the Big Five European Leagues. The 2017/2018 season was used to materialize calculations by going back to the respective previous season for the metrics construction. The data type is distributed in teams, players and coaches variables.

For players, there are numerical statistics, marks identifying whether they were debutants in the league that year and their transfer history. Of coaches, references such as age, nationality, dates of arrival and departure if applicable, and information from previous seasons to know if they had already been in the club. The third collective element brings two perspectives, the clubs competition table qualifying order this season and the qualifying hierarchy by the players accumulated market values.

This section also focuses on the data treatment, considering the categorical variables transformation into values, the null values possible cases to be eliminated or completed, or the non-binary numerical characteristics scaling to operate with volumetric ranges comparable to each other proportionally.

The third phase is based on the metrics enhancement, providing an enrichment by constructing new unique variables, governed by the volumetrics proportional distribution method, ensuring that each value is measured regarding to the teams quantitative situation.

These new variables are intermediated among themselves through a elements fusion system, generating new metrics based on other new metrics. All of this, using facts from past seasons and the player and coach variables combination, parameterizes a compendium grouping at the squad level, differentiating by season and competition, and extracting 490 unique entries.

Phase four introduces the analytical model as a Supervised Learning problem and the Classification model tries to predict a category:

“If a team achieved a qualifying performance according to the competition that is higher to the ordering by market value (1) or not (0), fulfilling it in the same way by being in the first position in the ordering by market value and managing to maintain the place

at level of competition, also considered as successful cases of not being able to rise a league table position beyond first place.”

The procedure follows fragmented sub-phases, the target variable stipulation, the 80% information training, the classification technique selection, the Fine-Tuning results consolidation, the appropriate hyper parameters settings improving the predictive capacity, the remaining 20% information testing phase, the elementary characteristics conservation that contribute the greatest positive and negative importance in whether or not they approach performance success and the model exactitude optimization, acquiring a reliable prediction result within the unavoidable uncertainty.

The Machine Learning techniques tested were Logistic Regression, K-Nearest-Neighbor, SGD, Naive Bayes, SVM, Decision Tree and Random Forest, complemented with 'Cross Validation' and achieving a truth-optimized performance metric.

4. Obtained results

The optimal technique was Logistic Regression ($C=1$, $random_state=42$), by virtue of the 65.30% accuracy score and the confusion matrix, achieving positive predictive proportions whether or not the goal was reached and one of the lowest standard deviation.

There were obtained six final characteristics weights, whose three negative and three positive coefficients values show the importance with both impacts on achieving the target.

The positive ones reflect metrics such as whether the squad is veteran players based (1.018342), whether the coach was at the club in the past (0.752595) and the number of days in a row in charge (0.535323), and the negative such as the league squad young debutants percentage proportion (-0.251582), the team market value sum in the forward position (-0.325496) and whether the squad players number volume is over the competition average (-0.603099).

The following data spatial distribution groups the Big Five European Leagues, filtering only the teams whose squad players volume is under than the competition average, showing the differences between having achieved optimal sporting performance or not.

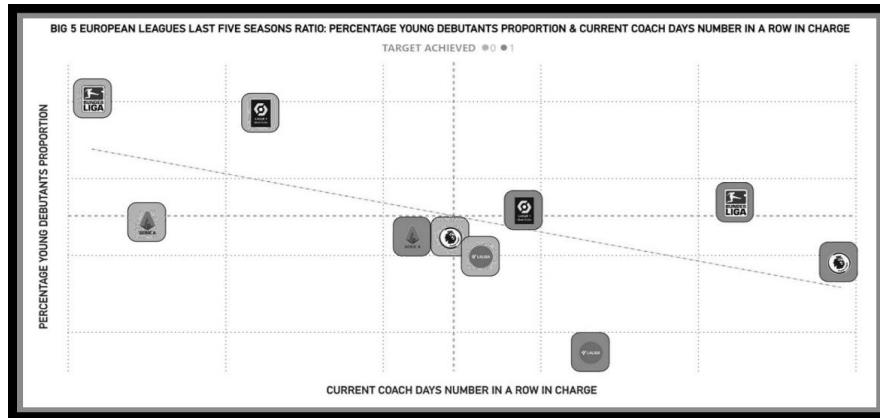


Figure 4. Explanatory Big Five European Leagues comparatives scatter plot

It is significantly denoted that the leagues overall performance was superior when they included solely the teams whose percentage young debutants proportion was low and the current coach days number in a row in charge was high, as the model teaches.

5. Deducted conclusions

The study object gives three main lessons learned by the deductive analysis exercise of the patterns and trends shown, providing an intuition aspect based on previous experience.

The first lesson is represented by the support continuity commitment through the attitude of endorsing and promoting the sporting success achievement conviction based on the project long-term sustainability, reflecting trust in the process, support the coach figure and avoid past stages desperate resorts.

Secondly, this learning enunciates the mentality of appreciating the experience. Without leaving aside the new young emerging talents appearances, it is proclaimed to deservedly estimate the immense value provided by the significant experiences accumulation in various situational contexts in the professional conjuncture, from the previous baggage intrinsic essence to the facilitated adaptation in new competitions.

Finally, it reinforces the willingness to consolidate the group despite the fact that the differential figures will continue to mark the success or not margins, the collective spirit takes precedence over the individual, strengthening the squad unity by virtue of the human factor reinforcement, constituting a reduced members block, avoiding the award-winning positions personified figures, prioritizing the forces and efforts balance and entrusting success based on the teamwork conviction and core transcendence.

Explainable xP: Solucionando Enviesamento de Modelos de Expected Pass com Tracking

Gabriel R. G. Barbosa¹, Henrique M. O. Carvalho², Fernando T. C. Oliveira²

¹ Departamento de Engenharia Elétrica – Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte, MG – Brazil

²Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte, MG – Brazil

grgb@ufmg.br, {henriquecarvalho, fernandotonucci}@dcc.ufmg.br

Abstract. *To fully understand the dynamics of a football match, it is crucial to analyze the passes made during the game. In this context, the concept of Expected Pass Completion (xP) emerged as a means to quantify the likelihood of a pass being completed, allowing for a better understanding of each of these events within a match. However, existing xP models in the literature often utilize post-pass informations, which can lead to biased results. In this work, we propose a new method for calculating xP using only pre-pass information. Additionally, we employ the Explainable Boosting Machine (EBM) to enhance the interpretability of our results, allowing us to conduct in-depth analyses on, for example, the individual impact of each defender in a specific play.*

Resumo. *Para compreender o panorama de uma partida de futebol, é essencial fazer uma análise efetiva dos passes nela realizados. Nesse contexto, o Expected Pass Completion(xP) surgiu como uma forma de metrificar a probabilidade de completude de um passe, possibilitando, portanto, um melhor entendimento sobre cada um desses eventos dentro de um jogo. Modelos de xP existentes na literatura, no entanto, utilizam informações posteriores ao passe, de forma que seus resultados podem se tornar enviesados. Neste trabalho, nós propusemos um novo método para calcular o xP que parte apenas de informações pré-passes. Adicionalmente, utilizamos o Explainable Bossting Machine(EBM) como forma de aumentar a interpretabilidade dos resultados, permitindo-nos realizar análises profundas sobre, por exemplo, o impacto individual de cada defensor em um dado lance.*

1. Introdução

Os passes são a ação básica que define o andamento de uma partida de futebol. Em média, cerca de 11 passes são trocados por minuto [Knutson 2017], o que os torna o evento mais comum. Com isso, o seu estudo é crucial para compreender o contexto de um jogo, já que, por ser um esporte de baixa pontuação, os gols não são um evento frequente o suficiente para essa análise. Além disso, uma boa troca de passes que gere um gol pode ser determinante para a vitória, o que reforça a relevância dessa ação. No intuito de qualificá-los, surgiram os modelos de *Expected Pass Completion(xP)*, que buscam prever a probabilidade de completude dos passes.

Em geral, os modelos de xP na literatura possuem três principais limitações que afetam a qualidade dos resultados. (1) Algumas abordagens ignoram informações relevantes para o problema, notoriamente, a posição dos jogadores em campo. (2) Mesmo trabalhos que tratam essa questão [Power et al. 2017, Anzer and Bauer 2022], introduzem dados de momentos pós-passe, e isso permite que o modelo estabeleça relações causais a partir de suas consequências, situação considerada como um vazamento de dados. (3) Outro problema comum é o uso de algoritmos pouco interpretáveis, como um Gradient-Boosted Ensemble of Decision Trees (GBM) usado em [Kullowatz 2018] e um Extreme Gradient Boosting (XGBoost) usado em [Anzer and Bauer 2022]. Esses métodos, embora atinjam bons valores nas métricas de classificação, têm suas aplicações limitadas devido a nebulosidade de seus resultados.

Portanto, o objetivo deste trabalho é propor um novo modelo de xP que aborda os problemas supracitados do seguinte modo: (1) introduzindo dados de tracking associado a eventos; (2) adicionando apenas informações anteriores ao passe, a fim de evitar vazamento; (3) utilizando o Explainable Boosting Machine(EBM) [Nori et al. 2019], modelo aditivo generalizado baseado em árvores, que tem como principal benefício a interpretabilidade.

Aplicamos nossa abordagem para realizar análises que nos ajudam a melhor compreender os impactos dos atacantes e defensores em campo. Para isso, utilizamos o PAX (passes completados além do esperado) [Gallagher 2021] para verificar a habilidade dos passadores, avaliamos a influência do posicionamento dos marcadores na probabilidade de completude dos passes e compararmos o estilo de jogadores baseado em características de seus passes.

2. Trabalhos Relacionados

Analizando trabalhos científicos e não-científicos de relevância no tema, é possível agrupá-los de acordo com os tipos de dados utilizados para a previsão da probabilidade de acerto do passe:

- Geométricos: compreendem dados numéricos geométricos do passe (ex: distância do passe, localização dos jogadores, ângulo do passe, velocidade do passe).
- Contexto de Passe: dados que caracterizam o passe (ex: parte do corpo, pressão do adversário).
- Eventos Passados: eventos que antecedem imediatamente o passe (ex: tempo de condução antes do passe, tempo de posse de bola do time)
- Contexto de Jogo: contextos específicos da partida (ex: time com jogador a mais, passe feito por jogador do mandante)
- Tracking Estático: uso de dados de tracking como extensão dos dados de evento (ex: jogadores próximos ao passador)
- Tracking Dinâmico: uso do alto frame rate de coleta de dados de tracking para simular a física da bola, prever posição futura de jogadores, etc.

Tabela 1. Tipos de dados usados por cada trabalho relacionado. G: Geométricos; CP: Contexto (Passe); CJ: Contexto (Jogo); EP: Evento Passados; TE: Tracking Estático; TD: Tracking Dinâmico;

Fonte	G	CP	CJ	EP	TE	TD
[Power et al. 2017]	x	x	x	-	-	x
[Kullowatz 2018]	x	x	x	-	-	-
[Gallagher 2021]	x	x	-	x	-	-
[Anzer and Bauer 2022]	x	x	-	-	-	x
Nosso Modelo	x	x	-	-	x	-

A Tabela 1 fornece uma visão geral das semelhanças e diferenças na abordagem de cada fonte. [Power et al. 2017] propuseram utilizar tanto dados de evento quanto tracking de alta taxa de coleta. O modelo escolhido para a tarefa foi a regressão logística, já que os autores prezaram pela interpretabilidade, ainda que com valores inferiores em métricas. Por meio do tracking, sugeriram a aplicação do recebedor esperado em situações em que o passe foi interceptado, cálculo feito com base no ângulo e distância dos companheiros de time ao passador. Ainda assim, são utilizadas informações referentes às posições finais de evento, dado proveniente do momento pós-passe.

O modelo da American Soccer Analysis [Kullowatz 2018] usa o GBM, que recebe informações contextuais do partida e do passe para calcular o xP. O uso deste modelo black-box, embora aumente a acurácia das previsões, diminui a interpretabilidade do processo, dificultando, consequentemente, a extração de análises a partir dos dados. Já o modelo da Opta [Gallagher 2021], que também usa o GBM, tem dois grandes diferenciais: a introdução de eventos passados como forma de caracterização do passe e a aplicação do SHAP(Shapley Additive Explanations) [Lundberg and Lee 2017] para aumentar a explicabilidade ao modelo.

[Anzer and Bauer 2022] propuseram múltiplas técnicas para aumentar acurácia de seu modelo. Usando tracking com alta taxa de coleta, estimaram o jogador para quem o passador tinha a intenção de enviar a bola. Isso foi feito por meio dos primeiros 0.4 segundos após o passe, cujos frames são utilizados para prever fisicamente a trajetória da bola se não houvesse interceptações. Além disso, usam velocidade e direção inicial dos jogadores para prever quem estaria mais próximo do ponto de chegada da bola. Para prever o quanto difícil é um passe, além dos dados geométricos e de contexto, é realizado um grid search variando a direção e velocidade do passe para determinar o quanto preciso o passe precisa para ser completado. Essa variedade de técnicas gera um custo computacional muito alto, o que dificulta uma análise rápida, durante ou imediatamente após o jogo, por exemplo.

Em todos os modelos da literatura, a distância entre passador e fim do evento é utilizada como feature, de algum modo. Contudo, essa é a distância efetiva percorrida pela bola após o passe; ou seja, passes errados não contém informação de para quem o passe foi direcionado, apenas onde ele foi interceptado. No modelo proposto por [Anzer and Bauer 2022], ainda que não se use diretamente a distância percorrida pela bola, foi construída uma feature que informa se os passes foram curtos, médios, ou longos, a partir dela.

Dado esse contexto de trabalhos recentes, propomos o uso de tracking estático como forma de extensão dos dados de evento. Além disso, o projeto minimizou vazamentos ao restringir seu contexto ao instante anterior ao passe, evitando o uso, por exemplo, de informações relacionadas ao final do evento.

3. Metodologia

Este projeto foi implementado em Python, usando dados fornecidos pela empresa SkillCorner¹ referentes à temporada 2020/21 da Premier League. Cada dado de evento foi associado a um único frame de tracking, permitindo uma solução eficiente computacionalmente. A métrica de xP foi construída utilizando o EBM [Nori et al. 2019].

Para minimizar o enviesamento por vazamento de dados, foram escolhidas features do momento *pré-passer*. Ou seja, foram descartadas a posição do recebedor no momento da recepção e a distância total percorrida pelo passe, uma vez que essas variáveis são conhecidas apenas no momento pós-passe. Em seus lugares, foram usadas informações de posição e distância referentes ao jogador mais provável de receber o passe - o recebedor esperado (seção 3.1). Ademais, os dados de tracking foram usados para metrificar os impactos do time defensivo no sucesso do passe. Nesse âmbito, as ações dos defensores foram divididas em três momentos distintos: na realização do passe, na viagem da bola, e no domínio do recebedor.

O EBM foi treinado com 90% dos dados e testado com o restante. A avaliação do modelo levou em consideração acurácia e AUC Score, já que as classes são desbalanceadas. O modelo final considerou as seguintes features:

- Distância passador - recebedor esperado
- Pressão ao passador
- Pressão ao recebedor esperado
- Pressão à linha do passe
- Ângulo passador - recebedor esperado em relação ao sentido de jogo
- Distância entre passador e gol adversário
- Distância entre passador e linha lateral mais próxima
- Coordenada x do passador
- Coordenada y do passador
- Coordenada x do recebedor esperado
- Coordenada y do recebedor esperado
- Parte do corpo

3.1. Recebedor Esperado

Para permitir a criação do modelo pré-passe, foi necessário calcular o *recebedor esperado* de cada passe. Ou seja, dada a direção da bola e a posição dos companheiros de equipe do passador no momento do passe, é calculado o provável recebedor do passe caso ele não seja interceptado. Na Figura 1, podem ser visualizados os passos para esse procedimento. São eles:

- I. Traçar uma projeção do passe a partir do ponto de interceptação, para sabermos a direção pretendida pelo passador.

¹<https://skillcorner.com/>

- II. Determinar a distância euclidiana entre cada um dos potenciais recebedores e a projeção da linha de passe.
- III. Determinar o ângulo entre potencial recebedor, passador e direção de passe.
- IV. Aplicar um peso $w(\theta)$ a cada ângulo, de modo a obter uma distância ponderada $d \cdot w(\theta)$, usada para selecionar o recebedor mais provável.

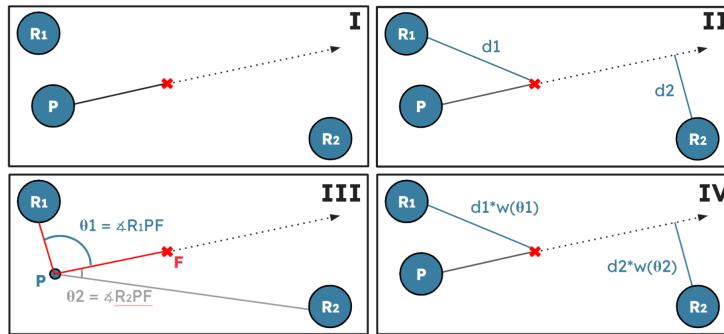


Figura 1. Passos para determinar o recebedor esperado de um passe. Neste exemplo, o passador P realizou um passe incompleto, interceptado no ponto F , em vermelho. R_2 seria o recebedor esperado do passe, devido à sua menor distância à linha de passe ponderada pelo ângulo agudo entre R_2, P e F .

O peso $w(\theta)$ é determinado por uma sigmoide, que faz com que ângulos mais agudos sejam menos penalizados do que ângulos obtusos.

3.2. Pressão dos Defensores

Para obter a pressão exercida pelos defensores, partimos da hipótese que, quanto mais rápido um marcador chega em certa região do campo, maior sua influência nela. Assim, por meio de um tempo estimado do deslocamento de cada defensor, calculamos sua pressão individual exercida no passador, na linha do passe, e no recebedor. Combinando os valores obtidos de todos os defensores, temos a pressão total da equipe.

3.2.1. Tempo de Deslocamento dos Jogadores

A fim de estimar o tempo de deslocamento dos jogadores para as regiões de interesse, foi empregado um modelo físico simplificado, que leva em consideração a velocidade e direção inicial dos jogadores. Para isso, o cálculo foi dividido em dois momentos: primeiro, calculamos a movimentação do jogador enquanto ele ajusta sua direção para o novo alvo; em seguida, calculamos sua movimentação, em linha reta, até seu destino.

Para inferir a movimentação durante a mudança de direção, partimos da velocidade inicial v_{ini} do jogador. Considera-se que ele reduz sua velocidade de maneira proporcional ao ângulo α , que representa o grau de desvio da rota inicial. Assim, a redução de velocidade Δv foi obtida pela seguinte fórmula:

$$\Delta v = \frac{v_{ini} * \alpha}{\pi}$$

Em seguida, divide-se Δv pela desaceleração $da = -7m/s^2$ para encontrar o tempo t_{dir} de redirecionamento do jogador. Sua movimentação mov em t_{dir} é dada por:

$$mov = t_{dir} * \frac{v_{ini} + (v_{ini} - \Delta v)}{2} * \theta$$

onde θ é um vetor normalizado que indica o ângulo médio da movimentação do atleta ao longo da curva, dado por $\frac{dir_{ini} - dir_{final}}{2}$. Note que mov será um vetor de dois elementos x e y, que indicam a distância que o atleta percorreu em cada eixo no tempo t_{dir} . Para obter a distância a ser percorrida pelo atleta na segunda etapa dos cálculos, basta somar mov ao vetor d que contém a distância nos eixos x e y entre o atleta e seu destino.

Na segunda etapa, tem-se um movimento acelerado retilíneo com a distância, velocidade inicial ($v_{ini} - \Delta v$) e aceleração ($a = 2.5m/s^2$ para este trabalho) conhecidos. Assim, o cálculo do tempo t_{reto} pode ser feito trivialmente partindo da equação de Torricelli. No entanto, tendo em vista que a equação considera uma aceleração constante, caso a velocidade máxima do jogador seja atingida, o restante do tempo é dado por $t = \frac{d}{v}$. Neste trabalho, consideramos $v_{max} = 7m/s$.

Por fim, adiciona-se o tempo de reação $r = 150ms$ no cálculo do tempo total t_{des} , dado por

$$t_{des} = t_{dir} + t_{reto} + r$$

3.2.2. Tempo de Deslocamento da bola

Para estimar o tempo que a bola demora para percorrer uma certa região do campo é essencial distinguir passes rasteiros de passes aéreos. Dado que o modelo proposto não usa informações posteriores ao passe, foi necessário assumir que passes curtos (i.e. menos de 20m) são rasteiros, enquanto os demais são aéreos. Além disso, foi preciso também assumir, para cada tipo de passe, a força inicial colocada na bola, de maneira a usar exclusivamente informações pré-passe.

Para passes rasteiros, consideramos $da_{grass} = -8m/s^2$ a velocidade de desaceleração da bola na grama, e uma velocidade inicial $v_{grass} = I_1 + \frac{d}{C_1}$, onde $I_1 = 13m/s$ representa uma velocidade mínima, e $C_1 = 3$ é uma constante para controlar o impacto da distância na força do passe. Nesse caso, o tempo de deslocamento pode ser obtido novamente pela equação de Torricelli.

Para passes aéreos, consideramos uma velocidade inicial de $v_{air} = I_2 + \frac{d}{C_2}$, com $I_2 = 12m/s$ e $c_2 = 6$. Note que, neste caso, a desaceleração da bola não seria constante, mas sim proporcional à velocidade da bola e a resistência do ar. Para simplificar o cálculo, no entanto, foi utilizada uma desaceleração constante da_{air} dada por $da_{air} = \frac{\rho * v_{air}^2}{2}$, onde ρ é a resistência do ar. Para estimar a resistência do ar, consideramos uma densidade do ar de $1.2Kg/m^3$, o coeficiente da bola como 0.18, a área da bola como $0.15m^2$ e a massa da bola como $400g$. Com a desaceleração constante, usamos, novamente, a equação de Torricelli para determinar o tempo de deslocamento.

3.2.3. Cálculo da Pressão

Para determinar a pressão que o time defensor exerce sobre o passe, calculamos a pressão individual de cada marcador em três momentos distintos da jogada: na realização do passe, no percurso da bola e no domínio do recebedor. Em todos os casos, utilizamos o tempo de deslocamento t_j do jogador adversário j como a base da estimativa da pressão, sob a hipótese que, quanto mais rápido o marcador chega no lance, maior seu impacto. A pressão exercida ao passador se dá de forma simples por meio da fórmula:

$$P_p(j) = 1 - \tanh(t_j)/B_1$$

onde $B_1 = 1.5$ regula o impacto de altos tempos de deslocamento na pressão. A função \tanh foi escolhida empiricamente por se assemelhar a uma variação intuitiva que se esperaria da alteração do impacto dos marcadores conforme a distância aumenta. Após isso, a pressão do total da equipe adversária é calculada pela fórmula:

$$P_p(all) = 1 - \prod_J (1 - P_p(j))$$

No cálculo da pressão defensiva, opta-se pela multiplicação em vez da soma por dois motivos principais: primeiro, a multiplicação garante que o valor total da pressão seja mantido entre 0 e 1, de forma que eles permaneçam normalizados para serem usados pela rede; segundo, a multiplicação reduz a pressão gerada por um grande número de defensores situados a uma distância média do lance relativo a um único jogador próximo ao lance.

Para obter a pressão ao recebedor do passe, não é factível considerar apenas o tempo de deslocamento do marcador, uma vez que, em passes mais distantes, ele tem um tempo adicional para chegar até o recebedor enquanto a bola se desloca entre os atacantes. Assim, deve-se considerar, também, o tempo de deslocamento da bola t_b nesse cálculo, de modo que a pressão ao recebedor é dada por:

$$P_r(j) = 1 - \tanh\left(\frac{t_j}{t_b}\right)$$

Similarmente, a pressão à linha de passe é determinada por meio da mesma fórmula, porém, aqui, t_j é o tempo de deslocamento entre o jogador j e o ponto mais próximo a ele na linha de passe.

4. Resultados

As métricas de qualidade do modelo foram comparadas com as métricas de um Baseline, construído utilizando o XGBoost, treinado com apenas dados de evento e permitindo dados do momento *pós-passe* (ou seja, enviesamento maior). Tais escolhas aproximam o Baseline da maioria dos modelos de xP atualmente.

O Baseline é comparado com dois modelos na Tabela 2. Ambos utilizam EBM treinado com dados de evento e tracking. O modelo *pré-passe* minimiza o enviesamento

utilizando da engenharia de features. Já o modelo *pós-passe* usa dados de distância final do passe como feature e, portanto, é beneficiado por vazamento de dados.

Tabela 2. Resultados dos diferentes modelos treinados com diferentes features.

	Baseline	Pré-Passe	Pós-Passe
Acurácia	86.135	84.735	87.709
Precisão	88.682	87.328	90.537
Revocação	93.854	94.346	94.278
AUC	87.597	88.237	91.362

4.1. Modelo pré-passe x Modelo pós-passe

Utilizando o EBM, temos grande interpretabilidade da relação entre cada variável e a variável alvo. Como o modelo final é aditivo generalista, é gerada uma função para cada feature, que pode ser visualizada para compreender seu impacto na predição.

A percepção mais relevante levantada pelo projeto vem da importância de se considerar apenas features do momento pré-passe. No modelo proposto, a distância entre passador e o recebedor esperado se relaciona com a probabilidade de completude do passe, de acordo com a função da Figura 2. Podemos observar uma coerência entre o que se vê no campo e a função: passes mais curtos são menos errados do que passes longos.



Figura 2. Função da distância pré-passe.

Contudo, se utilizamos a distância pós-passe (ou seja, a distância percorrida de fato pela bola), a função que relaciona a feature com a probabilidade de completude está mostrada na Figura 3. Pela função, conclui-se que passes muito curtos são mais difíceis do que passes médios, o que não faz sentido. Isso se dá porque a maioria das interceptações ocorrem próximas ao passador, e isso enviesa os dados de passes errados.

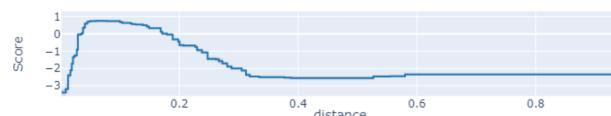


Figura 3. Função da distância pós-passe.

4.2. Passes Completed Above Expected (PAx)

O PAx é uma métrica introduzida por [Gallagher 2021], calculada para cada passe segundo a fórmula $PAx = [Resultado - xP]$, em que *Resultado* representa o resultado real do passe: 0 para um passe mal sucedido e 1 para um passe bem sucedido. A motivação dessa medida é atribuir valor aos passes com base na sua dificuldade.

O PAx foi utilizado em nossos dados para gerar uma estatística agregada de passes por jogador durante toda a temporada. O somatório proveniente do PAx de todos os passes de um jogador foi normalizado pelos minutos jogados por ele na temporada, informação retirada de uma base de dados de estatísticas agregadas da Premier League em 2020/21². O resultado dessa análise é mostrado na Tabela 3.

Tabela 3. Top 5 PAx na temporada 20/21 da Premier League.

Nome	Time	Minutos jogados	PAx	PAx Normalizado
Rodri	Manchester City	2748	96.306	0.035
John Stones	Manchester City	1933	67.142	0.034
Thiago Silva	Chelsea	1935	64.598	0.033
Mateo Kovacic	Chelsea	1815	60.295	0.033
César Azpilicueta	Chelsea	2188	71.541	0.032

4.3. Influência da marcação

Métricas de avaliação em momentos com a posse de bola são grandes limitadores na análise de dados no futebol, visto que, segundo [Anderson and Sally 2013], jogadores ficam em média 53.4 segundos com a posse de bola em um jogo. Avaliar um jogador apenas por esse período, portanto, é insuficiente para indicar o desempenho. Desse modo, propomos um método de avaliação do posicionamento de jogadores baseado na influência deles no resultado dos passes. Essa métrica consiste em analisar o quanto a posição dos defensores dificulta a conclusão de um passe. Para tal, o cálculo foi realizado removendo cada um dos marcadores e recalculando o xP. A diferença entre o valor base - com todos os defensores - e o valor sem um defensor específico foi considerada a influência desse defensor no passe.

A exemplo disso, foi selecionado um evento de passe entre Kevin De Bruyne e Phil Foden, em uma partida do Manchester City contra o Chelsea. É interessante, na Figura 4b, ver como o jogador mais próximo do passador(Kevin De Bruyne) tem uma influência de 12.4% no passe, isto é, sem ele na marcação, a chance de completude do passe aumentaria em 12.4%. Nesse caso, o defensor é o lateral César Azpilicueta.

4.4. Outras Aplicações

Além do PAx e da influência do marcador, o modelo de xP permite diversas outras análises. Usando dados de recebedor esperado, até então apenas calculáveis em modelos muito complexos e com tracking dinâmico, podemos caracterizar jogadores não somente com relação a seus passes mas também com relação ao estado do recebedor do passe. Isso pode ser observado no gráfico de radar da Figura 5. Nele, nota-se que De Bruyne costuma dar passes estando mais pressionado (*Pressão ao passador*), enquanto Alexander-Arnold dá passes para recebedores mais pressionados (*Pressão ao recebedor*). Essa ferramenta auxilia em diversos processos de tomada de decisão, desde análise de desempenho do elenco até processos de scouting de atletas.

²<https://www.kaggle.com/datasets/rajatrc1705/english-premier-league202021>



Figura 4. Influência dos defensores em um passe entre De Bruyne e Foden.

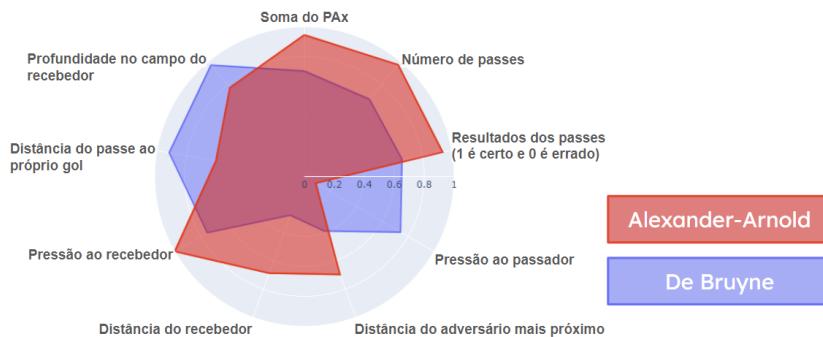


Figura 5. Gráfico de radar comparando dois jogadores da PL.

5. Conclusão

Neste trabalho, foi proposto um modelo para o problema de previsão da probabilidade de completude de um passe (xP) que traz inovações com relação ao uso de recebedor esperado e pressão detalhada ao passe, sem o uso de métodos excessivamente complexos computacionalmente. A interpretabilidade do EBM ajudou a justificar a importância do uso do recebedor esperado e features pré-passe para a criação de um modelo menos enviesado. O desempenho do modelo pós-passe superou o Baseline; e o modelo com features pré-passe soluciona um grave problema de enviesamento mantendo bom desempenho, porém abaixo do modelo pós-passe, como esperado.

Trabalhos futuros podem ser realizados com o objetivo de expandir a capacidade do modelo, como a introdução de dados de outras ligas e temporadas, bem como a utilização de tracking por GPS em vez de técnicas de visão computacional limitadas pelas câmeras de transmissão. Essa abordagem pode oferecer uma visão mais abrangente e precisa dos movimentos dos jogadores e da dinâmica do jogo, superando as limitações dos dados de tracking utilizados neste trabalho.

Referências

- Anderson, C. and Sally, D. (2013). *The Numbers Game: Why Everything You Know About Football is Wrong*. Penguin Books Limited.
- Anzer, G. and Bauer, P. (2022). Expected passes. *Data Mining and Knowledge Discovery*, 36:1–23.
- Gallagher, J. (2021). Introducing expected pass completion (xp).
- Knutson, T. (2017). Passing Percentages Are Mostly Useless – Quantifying Passing Ability.
- Kullowatz, M. (2018). An updated expected passing model.
- Lundberg, S. and Lee, S.-I. (2017). A unified approach to interpreting model predictions.
- Nori, H., Jenkins, S., Koch, P., and Caruana, R. (2019). Interpretml: A unified framework for machine learning interpretability. *arXiv preprint arXiv:1909.09223*.
- Power, P., Ruiz, H., Wei, X., and Lucey, P. (2017). Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’17, page 1605–1613, New York, NY, USA. Association for Computing Machinery.

Exploring the Relationship Between Scoring First, External and Internal Load, and Match Outcomes: A Case of Study of a Professional Brazilian Team during the 2021 Season

Gabriel Colodel¹, Lucas Pereira¹, Vitor Bertoli Nascimento^{1,2}

¹Pontifícia Universidade Católica do Paraná

² Universidade Federal do Paraná

colodel.gab@gmail.com, lucas.pereira1610@gmail.com,
vitor.bnascimento@pucpr.br

Abstract. This article aims to analyze, within the context of professional football, the relationships established between the team that scores the first goal and their respective metrics associated with external load (Inertial Movement Analysis - IMA) and internal load (TRIMP method) throughout a match. Based on these data, we can investigate their connection with the final outcome of a match, deepening our understanding of the game and its probabilities.

Resumo. Este artigo se propõe a analisar, contextualizado ao futebol profissional, as relações estabelecidas entre a equipe que marca o primeiro gol, e suas respectivas métricas associadas a carga externa (Inertial Movement Analysis - IMA) e de carga interna (método TRIMP) ao longo de uma partida. Com base nesses dados podemos investigar sua conexão com o placar final de uma partida, compreendendo de forma mais profunda o jogo e suas probabilidades.

1. INTRODUCTION

Soccer has high demand on different aspects. That requires teams to exhibit tactical, technical, physical, and psychological prowess to secure victory (Garganta, 1997). The game spans 90 minutes, during which the team that scores more goals than their opponent emerges as the winner (Eggels, 2016; Junior, 2015). The interplay of physical, physiological, and tactical metrics plays a crucial role in understanding the game dynamics, designing training programs, and predicting outcomes (Teoldo, et al, 2021).

Research conducted in various global soccer championships consistently demonstrates a significant increase in the likelihood of the team scoring the first goal winning the match (Farias, 2018; Garcia-Rubio, 2017; Armatas et al, 2009b; Molinuevo and Bermejo, 2012; Vergonis et al., 2019). For example, during the 2014 World Cup, 74.08% of matches resulted in the team scoring the first goal winning (Leite, 2015). In Major League Soccer from 2015 to 2016, the home team had a 72% probability of victory if they scored the first goal within the opening 15 minutes of the first half (AltarribaBarté, 2020). A detailed analysis of the entire Spanish women's championship during the same period showed that scoring the first goal increased the probability of victory by a factor of 2.8 to 11.1 (Ibañez, 2018).

Within the domain of elite Brazilian soccer, marked by the involvement of renowned clubs in numerous high-stakes championships each season, each match introduces unique strategic components. (Pereira, 2006). The dynamics of being the first team to score or concede a goal and the positioning on the scoreboard can lead to significant changes in tactical strategies and physiological metrics (Lago-Peñas, 2010; Almeida, 2014; Shakida, 2019; Silva, 2023)

2. OBJECTIVE

The aim of this case of study is to investigate and compare the likelihood of winning when a team scores first and its connection to external and internal load.

3. METHODOLOGY

We conducted an analysis of 64 matches involving a single professional Brazilian team during the 2021 season across four championships: the Brazilian Championship ($n = 37$, 57.8%), Brazilian Cup ($n = 8$, 12.5%), Sul American Cup ($n = 13$, 20.3%), and State Championship ($n = 6$, 9.3%). We recorded instances when the team scored first. In each match, players wore GPS + IMU units (*Catapult Vector S7, Catapult Innovations, Australia & SPI Elite, GPSports, Australia*) securely fastened within fitted vests, positioned between the upper sides of the scapula blades. Due to certain arenas lacking GPS signals, we relied on IMU variables to capture external load, utilizing IMA units. To assess internal load, we employed the TRIMP method. For statistical analysis, we calculated the percentage and probability of the team scoring first using multinomial logistic regression. To compare outcome of match, we utilized a generalized mixed model with a Poisson distribution for IMA, considering it as a count variable. For TRIMP, we employed the Gama distribution. In all our models, we included 'score first' as a random effect. All statistical analyses were performed using *Jamovi* (version 2.3.12)

4. RESULTS

The probability to winning, losing, and drawing for each championship is presented in table 1.

Table 1. Estimates probability for championship and standard error an interval confidence

Championship	Result	Probability	SE	Lower	Upper
Brazilian Championship	Lose	7.3%	3.8%	0.4%	15.0%
Brazilian Championship	Draw	0.0%	0.0%	0.0%	0.0%
Brazilian Championship	Win	92.7%	3.8%	85.0%	100.0%
Brazilian Cup	Lose	0.0%	0.0%	0.0%	0.0%
Brazilian Cup	Draw	37.4%	20.6%	5.0%	79.7%
Brazilian Cup	Win	62.6%	20.6%	20.3%	100.0%
Stadual	Lose	18.4%	8.1%	1.8%	34.9%
Stadual	Draw	25.2%	19.2%	14.3%	64.8%
Stadual	Win	56.4%	23.0%	9.0%	100.0%
Sul American Cup	Lose	0.0%	0.0%	0.0%	0.0%
Sul American Cup	Draw	0.0%	0.0%	0.0%	0.0%
Sul American Cup	Win	100.0%	0.0%	100.0%	100.0%

On table 2 we present comparisons between outcomes Win (estimated mean: 2106.12; IC: 2040.43-2173.95), Draw (estimated mean: 2023.13; IC: 1981.25 – 2065.91), Lose (estimated mean: 1992.75; IC 1953.52 – 2032.77).

Table 2. IMA comparison between outcomes

Outcome		exp(B)	SE	z	p
Draw	Win	0.961	0.012	-3.352	<.001
Lose	Draw	0.985	0.01	-1.527	0.127
Lose	Win	0.946	0.011	-4.773	<.001

On table 3 we present comparisons between outcomes Win (estimated mean: 2732.21; IC: 2564,03 – 2900.40), Draw (estimated mean: 2715.66; IC: 2441.02 - 2990.31), Lose (estimated mean: 2610.95, IC: 2398.21 – 2823.68)

Table 3. TRIMP comparison between outcomes

Outcome		Difference	SE	z	p
Draw	Win	-16.552	161.054	-0.103	0.918
Lose	Draw	-104.717	173.732	-0.603	0.549
Lose	Win	-121.269	135.620	-0.894	0.375

5. Discussion

The primary goal of this study was to investigate the probability that a soccer team had, throughout a season, of winning a match after scoring the first goal. This was coupled with an examination of the correlation between external and internal loads metric's of its athletes. According to the statistical analyses conducted, the probability of win indeed increased, in line with previously published studies. Concerning the external load, it was higher in comparison between the results of Win - Draw and Win - Lose, while no difference was observed between the results of Defeat - Draw. However, regarding the internal load, no differences were identified in any of the result comparisons.

According to table number 1, there was indeed a higher probability for the team, after scoring the first goal of the match, to end up with a win outcome. Additionally, according to the data, we found that this probability of achieving a win after scoring first varied depending on the championship being contested. In the Brazilian championship, the probability was 92.7%, in the Brazilian Cup it was 62.6%, in the Stadual 56.4%, and finally, in the South American Championship, it was 100%.

In 2021, the team reached the semifinals of the State Championship, the first tournament of the year, in which the team traditionally mixes professional athletes with those from the youth categories. They also made it to the finals of the Brazilian Cup, a knockout tournament involving several teams from Brazil, with a significant financial prize. In the Brazilian Championship, a league tournament, the team finished in 14th place out of 20 teams. And in the South American Championship, one of the most prestigious tournaments on the continent, also highly lucrative, but with a small sample of games, the team became the champion for the second time. This information can lead to understand the difference between the probabilities.

In accordance with the article published by Leite (2015), about the relationship between the winning team and the one that scored the first goal of the match was analyzed

the following international championships: FIFA World Cups, UEFA Euro, CONMEBOL America Cups, AFC Asian Cup, CAF Africa Cups, FIFA Confe. Cups. In each championship, there was also a variation in the probability of winning after the first goal was scored, but nevertheless, in all of them, the probability of winning increased; the overall average of this probability stood at 71.17%. Although our study analyzes data from a single team, indicating a notably smaller sample size, the probability aligns closely with our findings.

Analyzing the table number 2, we came to conclusion that between the outcomes Win-Draw and Win-Lose there was a difference in the variable Inertial Movement Analysis (IMA), in correlation with scoring first. Already in the outcomes Lose-Draw, there was no notable difference.

According to Holme (2015), this technology, known as "Inertial Movement Analysis" (IMA), not only assesses frequency but also computes the magnitude and direction of agility and change of direction movements. Moreover, it has the capability to categorize events by intensity and differentiate among forward, backward, and lateral movements both to the left and right. The variables associated with IMA hold promise in identifying critical aspects of gameplay and in distinguishing between individual players' distinct agility patterns. Understanding about the athlete's external load and physical effort.

Upon a thorough analysis of the data, a discernible trend emerges: Inertial Movement Analysis (IMA) tends to register higher values when a team is leading in a match. This underscores the critical importance of sustaining a heightened level of physical performance, particularly after scoring the initial goal, to secure a victory. In accordance with the findings of Moala (2017), variables such as low intensity running and the total distance covered, integral components of the external load, exhibit higher values in matches resulting in wins as opposed to draws or losses. This implies that a truly successful team must consistently uphold and manage a superior standard of physical performance.

In examining the internal load, we utilized the TRIMP (Training Impulse) method, which incorporates heart rate data. Nevertheless, our results demonstrated a lack of evident correlation between the TRIMP metric and any particular match outcome. Similarly, Kanope (2018) conducted an analysis of TRIMP in under-20 youth athletes during official matches, shedding light on the potential limitations of this metric in intermittent sports due to heart rate variability. It is worth noting that both studies were constrained by relatively small sample sizes.

6. Conclusion

The analyses highlight the pivotal role of the team scoring the initial goal, and the correlation with the match outcome is notably influenced by the specific championship under consideration. Inertial Movement Analysis (IMA) stands out as a responsive metric closely tied to the match outcome, contingent on which team secures the first goal. Elevated IMA values correspond to an increased likelihood of securing victory. This valuable insight can guide the customization of the team's training regimen and strategic approach across diverse championships. On the other hand, in relation to the TRIMP method, no discernible correlation was observed between the match outcome and variations in this metric.

References

- Almeida, C. H., Ferreira, A. P., and Volossovitch, A. (2014). Effects of match location, match status and quality of opposition on regaining possession in UEFA champions league. *J. Hum. Kinet.* 41, 203–214. doi: 10.2478/hukin-2014-0048
- Altarriba-Bartés, A., Calle, M. L., Susín, A., Gonçalves, B., Vives, M., Sampaio, J., et al. (2020). Analysis of the winning probability and the scoring actions in the American professional soccer championship. [Análisis de la probabilidad de ganar y de las acciones que conducen al gol en la liga americana de fútbol profesional]. RICYDE. *Rev. Int. cienc. deporte* 16, 67–84. doi: 10.5232/ricyde2020.05906
- Armatas, V., Yiannakos, A., Papadopoulou, S., and Skoufas, D. (2009b). Evaluation of goals scored in top ranking soccer matches: Greek “Super League” 2006-07. *Serb. J. Sports Sci.* 3, 39–43.
- Eggels, H (2016). Expected Goals in Soccer. Eindhoven University of Technology.
- Farias, V. M. et al. (2018) Goal scoring frequency in soccer in different age groups. *Revista Brasileira de Ciência e Movimento*, v. 26, n. 3, p. 47, 15 nov. .
- Pereira, L. (2006) Modelação do jogo de futebol. Universidade do Porto. Faculdade de Ciências do Desporto e de Educação Física
- Garganta, J. (1997). Modelação táctica do jogo de Futebol. Estudo da organização da fase ofensiva em equipas de alto rendimento. Dissertação de Doutoramento em Ciências do Desporto. Universidade do Porto: Faculdade de Ciências do Desporto e de Educação Física
- Holme, B.R. (2015). Wearable microsensor technology to measure physical activity demands in handball: a reliability study of inertial movement analysis and player load.
- Ibañez, S. J. et al. (2018) The impact of scoring first on match outcome in women’s professional football. *International Journal of Performance Analysis in Sport*, v. 18, n. 2, p. 318–326, 4 mar.
- Junior, N; Paulo, S. (2015) Evidências científicas sobre o gol do futebol: uma revisão sistemática. *Revista brasileira de futebol e futsal*. v. 25, p. 297–311, [s.d.].
- Kanope, T., Morandi, R. F., & Pimenta, E. M. (2018). Quantificação da carga de jogos oficiais de futebol através do TRIMP. *Revista Brasileira De Educação Física E Esporte*, 32(3), 351-359. <https://doi.org/10.11606/1807-5509201800030351>
- Lago-Peñas, C., Lago-Ballesteros, J., Dellal, A., and Gómez, M. (2010). Game-related statistics that discriminated winning, drawing and losing teams from the Spanish soccer league. *J. Sports Sci. Med.* 9:288.
- Leite, W. (2015). The influence of the first goal on the final result of the football match. *Balt. J. Sport, Health Sci.* 3, 29–35. doi: 10.33607/bjshs.v3i98.91.
- Liu, T., Fournier-Viger, P., and Hohmann, A. (2018). “Using diagnostic analysis to discover offensive patterns in a football game,” in *Recent developments in data science*

- and business analytics. eds. M. Tavana and S. Patnaik (Cham, Switzerland: Springer). 381–386.
- Molinuevo JS, Bermejo JP. (2012). The effect of scoring first and home advantage in professional spanish football and indoor soccer leagues. *Rev Psicol Deporte*; 21: 301-8
- Moalla, Wassim & Fessi, Mohamed & Makni, Emna & A, Dellal & Filetti, Cristoforo & Salvo, Valter & Chamari, Karim. (2017). Association Of Physical And Technical Activities With Partial Match Status In A Soccer Professional Team. *Journal of Strength and Conditioning Research*. 32. 1. 10.1519/JSC.00000000000002033.
- Pratas, J. M.; Volossovitch, A.; Carita, A. I. (2018) Analysis of Scoring Sequences in Matches of the Portuguese Premier League. *Journal of Human Kinetics*, v. 64, n. 1, p. 255–263, .
- Sánchez-Murillo, P., Antúnez, A., Rojas-Valverde, D., & Ibáñez, S.J. (2021). On-Match Impact and Outcomes of Scoring First in Professional European Female Football. *International Journal of Environmental Research and Public Health*, 18.
- Silva, Hugo & Nakamura, Fabio & Casamichana, David & Barba, Eider & Castellano, Julen & Marcelino, Rui. (2023). Match peak speeds, maximum accelerations, and maximum decelerations differ in young football players: expression of maximal capacities is dependent of match context. *Sport Sciences for Health*. 1-6. 10.1007/s11332- 023-01124-1.
- Teoldo da Costa, Israel & Guilherme, José & Garganta, Júlio. (2021). Football Intelligence: Training and Tactics for Soccer Success. 10.4324/9781003223375
- Vergonis, Alexandros & Michailidis, Yiannis & Mikikis, Dimitrios & Semaltianou, Eleni & Mavrommatis, George. (2019). Techinical analysis of goal scoring patterns in the 2018 FIFA WORLD CUP in Russia . *Facta Universitatis, Series: Physical Education and Sport*. 17. 181-193. 10.22190/FUPES190612019V.
- Zhou, C., Gómez, M.-Á., and Lorenzo, A. (2020). The evolution of physical and technical performance parameters in the Chinese Soccer Super League. *Biol. Sport* 37:139. doi: 10.5114/biolspor.2020.93039

Uso de Parâmetros de Curvas Exponenciais para Classificação do Desempenho em Sprints de 30 Metros em Atletas de Futebol Sub-20

Lucas Pereira¹, Gabriel Colodel¹, Pedro Menezes³, Vitor Bertoli Nascimento^{1,2}

¹Pontifícia Universidade Católica do Paraná

²Universidade Federal do Paraná

³Clube de Regatas do Flamengo

lucas.pereira1610@gmail.com, colodel.gab@gmail.com,
pedro.menezes@flamengo.com.br, vitor.bnascimento@pucpr.br

Abstract. This study, titled “Uso de Parâmetros de Curvas Exponenciais para Classificação do Desempenho em Sprints de 30 Metros em Atletas de Futebol Sub-20”, aims to introduce a graphical classification method using exponential curve parameters to model the velocity curve during a 30-meter sprint. Involving 22 under-20 soccer players, the study utilizes GPS + IMU units for velocity measurement and software R for data analysis. Findings highlight the utility of velocity curve graphs in understanding individual sprint characteristics, offering a nuanced approach to training program development. The graphical representation aids in the comparative analysis of athletes, enhancing understanding and training customization based on individual profiles.

Resumo. O presente estudo propõe uma metodologia de classificação gráfica para a análise do desempenho em sprints de 30 metros em atletas de futebol sub-20, utilizando parâmetros de curvas exponenciais. Participam 22 jogadores, com dados de velocidade coletados por dispositivos GPS + IMU e analisados através do software R. A pesquisa demonstra como a representação gráfica da curva de velocidade pode facilitar a interpretação dos dados e auxiliar treinadores e atletas no desenvolvimento de programas de treinamento mais eficazes. Os resultados enfatizam a importância dos gráficos de quadrante para a compreensão das características individuais de sprint, permitindo uma abordagem mais personalizada e eficiente no treinamento esportivo.

1. Problema

O futebol é uma modalidade esportiva que apresenta diversas demandas físicas (GOMEZ, 2019). Durante os 90 minutos de uma partida, um jogador percorre, em média, uma distância de 10 quilômetros. Dentro dessa quilometragem, entre 8% e 18% são percorridos em alta velocidade, muitas vezes em momentos cruciais da partida (EKBLOM et al. 1986). O entendimento da capacidade de sprint dos atletas pode ser explorado por meio da curva de velocidade.

A curva de velocidade durante um sprint pode ser modelada por meio de uma equação exponencial (FURUSAWA; HILL; PARKINSON, 1927; VOLKOV; LAPIN, 1979) expressa pela equação: $y = a \cdot e^{b \cdot x}$. Nessa, 'a' representa a magnitude ou amplitude da curva, ou seja, quanto maior o valor de 'a', maior será a amplitude da curva exponencial, ou seja, um valor mais elevado de 'a' resultará em uma curva mais íngreme. Por sua vez, o parâmetro 'b' corresponde à taxa da função exponencial, determinando a taxa de crescimento ou decaimento da curva exponencial. Um valor positivo de 'b' resultará em um crescimento exponencial, enquanto um valor negativo de 'b' resultará em um decaimento exponencial. Portanto, quanto maior o valor absoluto de 'b', mais rapidamente a curva se afastará da linha de base.

Além desta forma também podemos modelar a velocidade máxima por meio da seguinte equação : $v(t) = MVS \times \left(1 - e^{-\frac{t}{TAU}}\right)$, onde MVS é máxima velocidade de sprint (m/s), TAU tempo para aceleração relativa (s). Desta podemos obter máxima aceleração (m/s^2) $MAC = \frac{MVS}{TAU}$. Tais medidas foram recentemente popularizadas por (SAMOZINO et al., 2012) e podem ser aplicados a dados de tracking de GPS (CLAVEL et al., 2022).

Apesar da análise desses parâmetros serem uma ótima ferramenta para compreensão de sprints no futebol, permitindo orientar os treinadores na elaboração de programas de treinamento mais eficazes. A representação gráfica dessas informações matemáticas pode apresentar desafios na sua interpretação no dia a dia do treinamento esportivo. Nesse sentido, a utilização de ferramentas gráficas se torna essencial para aprimorar a compreensão desses dados tanto por atletas quanto por profissionais envolvidos no treinamento esportivo se faz necessária.

Uma abordagem eficaz para a visualização de duas variáveis e classificação de observações é o uso de gráficos de quadrante, que também são utilizados para diagnósticos clínicos (THOMAS et al., 2006). Esta abordagem divide as observações em quatro quadrantes distintos, o que pode facilitar a interpretação e a tomada de decisões informadas no contexto do treinamento esportivo.

2. Objetivo

O objetivo deste estudo foi explorar a modelos exponenciais aplicados ao sprints em atletas de futebol sub-20, visando uma melhor compreensão das dinâmicas de aceleração e velocidade máxima.

3. Metodologia

Participaram do estudo 22 jogadores pertencentes à categoria sub 20, com idades aproximadas de 19.88 ± 1.66 anos, pesos médios de 77.41 ± 8.55 quilogramas, e estaturas médias de 181.91 ± 6.41 cm. Deste n = 6 atacantes, n = 5 laterais, n = 4 meio-campo, n = 4 volantes, n = 4 zagueiros. Para a realização do teste, foi empregado um percurso de 30 metros demarcado com cones para fins de avaliação. A velocidade foi avaliada por meio de dispositivos GPS + IMU units (*Catapult Vector S7, Catapult Innovations, Australia & SPI Elite, GPSports, Australia*), com uma frequência de aquisição de 10Hz.

Os dados coletados foram analisados no software R (Versão 4.0.2, *R Foundation for Statistical Computing*) por meio de uma rotina personalizada desenvolvida para esta finalidade utilizada em outros estudos (CLAVEL et al., 2022; JOVANOVIĆ; VESCOVI, 2022). Para criação dos gráficos de quadrante foi calculado o valor médio da variável de cada eixo plotado em gráficos de dispersão para ser ponto central do gráfico.

Para comparação dos valores de 'a', 'b', MVS, MAC e TAU entre posições táticas foi utilizado ANOVA considerando $p < 0.05$. O tamanho do efeito (d de Cohen) foi calculado a partir da razão entre a diferença média e o desvio padrão combinado e interpretado da seguinte forma: trivial < 0.2 ; $0.2 \leq$ pequeno < 0.6 ; $0,6 \leq$ moderado < 1.2 ; $1.2 \leq$ grande $< 2,0$; muito grande $> 2,0$ (HOPKINS et al., 2009).

4. Resultados

Na figura 1 apresentamos as curvas de velocidade de cada atleta e agrupadas por posição tática.

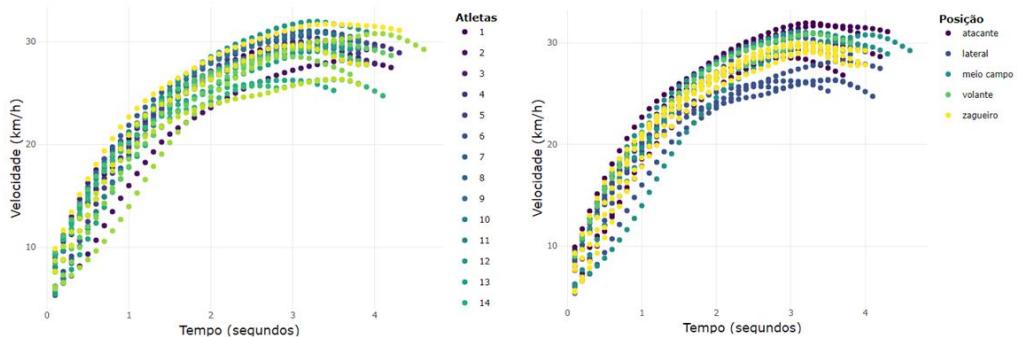


Figura 1. Curvas de velocidade por posição tática e individual

Apesar de não serem observadas diferenças para valores 'a' ($p = 0.658$) e 'b' ($p = 0.918$) entre as posições táticas. Observamos efeitos moderado entre atacantes e laterais, laterais e volantes e volantes e zagueiros. (tabela 1).

Tabela 1. Apresentando valores de média comparação para 'a' entre as posições

Comparação		Intervalo de Confiança a 95%						
Posição	Posição	Erro-padrão	t	p-valor	d de Cohen	Lim. Inferior	Superior	Efeito
Atacante	Lateral	0.922	1.282	0.217	0.776	-0.532	2.084	Moderado
	Meio	0.983	0.647	0.526	0.418	-0.952	1.788	Pequeno

	campo							
	Volante	1.244	-0.232	0.819	-0.19	-1.914	1.534	Pequeno
	Zagueiro	0.922	0.895	0.383	0.542	-0.751	1.834	Pequeno
Lateral	Meio campo	1.022	-0.534	0.6	-0.358	-1.779	1.063	Pequeno
	Volante	1.274	-1.154	0.264	-0.966	-2.765	0.834	Moderado
	Zagueiro	0.963	-0.37	0.716	-0.234	-1.571	1.103	Pequeno
Meio campo	Volante	1.319	-0.702	0.492	-0.608	-2.448	1.233	Moderado
	Zagueiro	1.022	0.185	0.856	0.124	-1.292	1.540	Pequeno
Volante	Zagueiro	1.274	0.874	0.394	0.732	-1.053	2.517	Moderado

Nota. Comparações baseadas nas médias marginais estimadas

Os efeitos para valores de 'b' entre atacantes e laterais, laterais e volantes e volantes e zagueiros foram pequenos (tabela 2).

Tabela 2. Apresentando valores de média, erro padrão e d de Cohen comparação para 'b' entre as posições

Comparação		Intervalo de Confiança a 95%						
Posição	Posição	Erro-padrão	t	p	d de Cohen	Lim. Inferior	Superior	Efeito
Atacante	Lateral	0.018	0.072	0.943	0.044	-1.234	1.321	Pequeno
	Meio campo	0.019	-0.375	0.712	-0.242	-1.607	1.122	Pequeno
	Volante	0.024	-0.176	0.863	-0.143	-1.867	1.580	Pequeno
	Zagueiro	0.018	-0.815	0.427	-0.493	-1783	1.000	Pequeno
	Meio campo	0.009	-0.426	0.675	-0.286	-1.705	1.133	Pequeno

Lateral	Volante	0.025	0	0.826	-0.187	-1.954	1.579	Pequeno
	Zagueiro	0.019	-0.849	0.408	-0.537	-1.885	1.000	Pequeno
Meio campo	Volante	0.026	0.114	0.91	0.099	-1.729	1.926	Pequeno
	Zagueiro	0.020	-0.374	0.713	-0.251	-1.669	1.167	Pequeno
Volante	Zagueiro	0.020	-0.418	0.681	-0.35	-2.120	1.420	Moderado

Atacantes e volantes apresentam maiores valores para 'a', seguidos de meio campistas, zagueiros e laterais, enquanto para 'b' maiores valores são apresentados por zagueiros, meio campistas, atacantes e laterais. Os valores próximos de 'b' mostram similaridades entre as posições, no entanto, para 'a' tais diferenças são mais pronunciadas. Tal resultado pode ser explicado a categoria analisada, porém tal hipótese precisa ser testada com amostra maior.

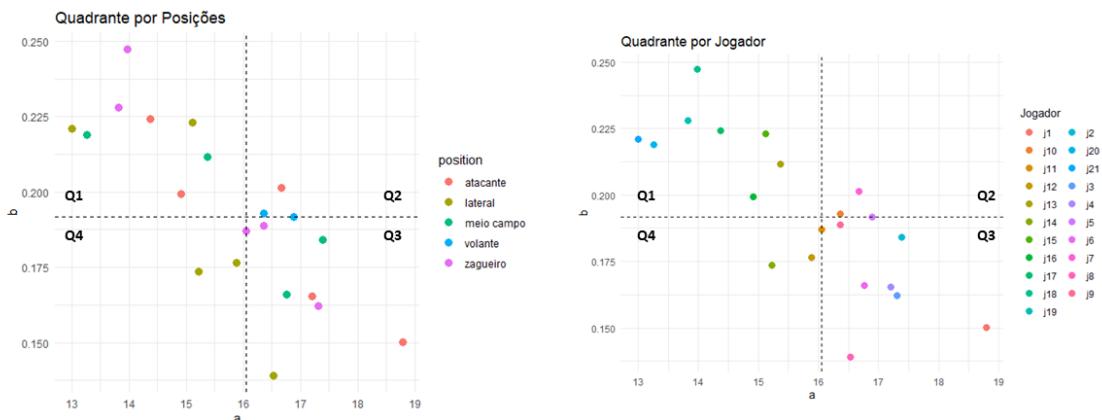


Figura 2. Gráfico de Quadrante por posição e individual

No quadrante 1 temos atletas que tem grande valor de 'a', porém baixo valor de 'b'. Nesse quadrante estão atletas que apresentam grande aumento da velocidade, porém No Quadrante 2 temos atletas com grande valor de 'a' e 'b'. No quadrante 3 temos atletas com grande valor de 'b' e menor de 'a', no quadrante 4 temos atletas com baixo valor de 'a' e 'b'. Não foi observada concentração de atletas de uma posição nos quadrantes.

Os p-valor e tamanhos de efeito da diferença de MSS entre as posições atacante (6.00 m/s, IC: 5.52-6.48), zagueiro (6.05 m/s, IC: 5.47-6.64), volante (6.87 m/s, IC: 6.26-7.43), lateral (5.80 m/s, IC: 5.55-6.39) e meio-campo (6.43 m/s, IC: 5.85-7.02) são apresentados na tabela 3.

Tabela 3. Comparação dos valores de posição MSS

Comparação	Intervalo de Confiança a 95%
------------	------------------------------

Posição	Posição	Diferença Média	Erro-padrão	t	p	d de Cohen	Lim. Inferior	Superior	Efeito	
Atacante	Lateral	0.2	0.358	0.559	0.583	0.361	-1.007	1.729	Pequeno	
	Meio campo	-0.432	0.358	-1.209	0.243	-0.78	-2.171	0.611	Modera do	
	Volante	-0.842	0.358	-2.355	0.031	-1.520	-2.989	-0.051	Grande	
	Zagueiro	-0.052	0.358	-0.147	0.885	-0.095	-1.457	1.268	Trivial	
Lateral	Meio campo	-0.632	0.392	-1.614	0.125	-1.141	-2.689	0.407	Grande	
	Volante	-1.042	0.392	-2.660	0.017	-1.881	-3.521	-0.241	Grande	
	Zagueiro	-0.252	0.392	-0.644	0.528	-0.456	-1.956	1.045	Pequeno	
Volante	Meio campo	Volante	-0.41	0.392	-1.046	0.31	-0.74	-2.255	0.776	Modera do
	Zagueiro	Zagueiro	0.38	0.392	0.97	0.346	0.686	-0.827	2.198	Modera do
	Volante	Zagueiro	0.79	0.392	2.016	0.06	1.425	-0.153	3.004	Grande

Nota. Comparações baseadas nas médias marginais estimadas

Os p-valor e tamanhos de efeito da diferença de MAC entre as posições atacante (4.77 m/s^2 , IC: 4.13-5.41), zagueiro (4.28 m/s^2 , IC:3.50-5.06), volante (3.98 m/s^2 , IC:3.20-4.76), lateral (3.65 m/s^2 , IC:2.87-4.43) e meio-campo (4.19 m/s^2 , IC:3.41-4.97) são apresentados na tabela 4.

Tabela 4. Comparação dos valores de MAC por posição

Comparação							Intervalo de Confiança a 95%	
Posição	Posição	Diferença Média	Erro-padrão	t	p	d de Cohen	Lim. Inferior	Superior
Atacante	Lateral	1.121	0.479	2.341	0.032	1.511	0.043	2.978
	Meio campo	0.578	0.479	1.208	0.244	0.780	-0.611	2170
	Volante	0.788	0.479	1.646	0.118	1.063	-0.353	2478

	Zagueiro	0.488	0.479	1.020	0.322	0.658	-0.724	2.041
Lateral	Meio campo	-0.542	0.525	-1.034	0.316	-0.731	-2.246	0.784
	Volante	-0.333	0.525	-1	0.535	0.448	-1.949	1052
	Zagueiro	-0.632	0.525	-1.206	0.244	-0.853	-2.376	0.671
	Volante	0.210	0.525	0	0.694	0.283	-1.212	1.778
Meio campo	Zagueiro	-0.090	0.525	-0.172	0.866	-0.121	-1.614	1.371
	Volante	-0.300	0.525	-1	0.575	-0.404	-1.903	1.095

Nota. Comparações baseadas nas médias marginais estimadas

Na Figura 3 apresentamos gráficos de quadrante entre MVS e MAC. No quadrante 1 (Q1) temos atletas com maior MAC e maior MVS, atacantes(n=3), meio campistas (n=2), zagueiro (n=2). No quadrante 2 (Q2) temos atletas com maior MVS, mas menor MAC, volantes (n= 3), lateral (n= 1), meio campo(n=1) e atacante (n=1). No quadrante 3(Q3) temos atletas com menor MVS e menor MAC do grupo, lateral (n=3), zagueiro (n=3), meio campo (n=1). No quadrante 4 temos atletas com maior MAC e menor MVS, atacante (n=2), zagueiro (n=1).

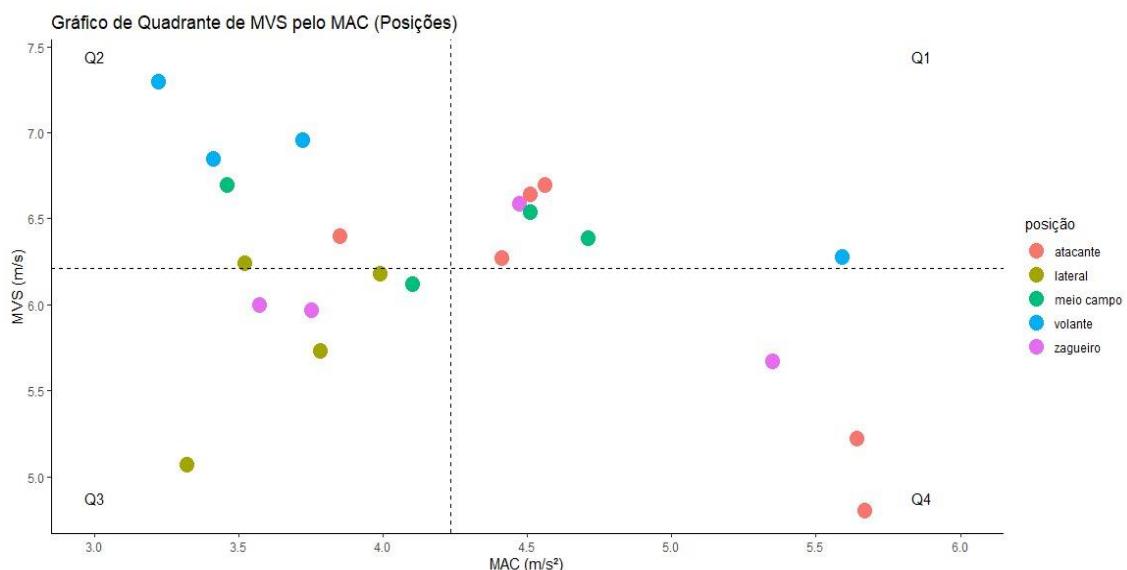


Figura 3.3 Gráfico de Quadrante de MVS e MAC por posição tática

5. Discussão

A análise dos dados apresentados nas tabelas do nosso estudo revela que os volantes apresentaram a maior velocidade máxima de sprint (MSS), destacando-se na manutenção de altas velocidades por períodos prolongados, uma habilidade crucial para cobrir grandes áreas do campo. Essa característica se mostrou significativamente superior aos atacantes, laterais e zagueiros. Por outro lado, os laterais, com o menor MSS, parecem equilibrar a velocidade com demandas táticas e físicas específicas da posição, resultando em um impacto negativo em comparação com meio-campistas e volantes.

Ao examinar a Figura 3, observamos variações nas características de sprint entre as diferentes posições. Atacantes, meio-campistas e zagueiros, localizados no Quadrante 1, demonstraram maior MVS e MAC, sugerindo uma forte orientação para a força inicial e eficácia mecânica em velocidades crescentes. Em contraste, jogadores em outras posições, como volantes e laterais, apresentaram uma combinação de maior MVS e menor MAC, refletindo as necessidades e habilidades físicas únicas de cada posição.

Tais necessidades demostradas em cada quadrante permitem aos treinadores utilizarem métodos específicos para desenvolvimento da capacidade de sprint. Atletas com grande aceleração e baixa velocidade são recomendados métodos assistivos como corrida em declive, e tracionados por elástico. Por outro lado, atletas velozes, mas com pouca aceleração se beneficiariam de métodos como treino com peso corporal assim como proposto por Hicks et al. (2023). Com isso apesar das diferenças individuais, podem ser contempladas durante o treinamento,

Em resumo, este estudo destaca a importância de uma abordagem personalizada no treinamento de atletas, onde estratégias de aceleração podem ser adaptadas com base nas características individuais dos sprinters. As diferenças observadas entre as posições de jogadores sub-20 em termos de MSS, MVS e MAC ressaltam a necessidade de treinamentos específicos que considerem essas variações. Além disso, a análise comparativa com estudos externos fornece uma base mais sólida para futuras pesquisas e aplicações práticas em treinamento e performance atlética.

6. Conclusão

Com base nos gráficos da curva de velocidade, é possível analisar as características de sprint dos atletas. Portanto, torna-se viável comparar o perfil dos atletas entre si, compreendendo de forma suas características. Apoiado no gráfico de quadrantes, temos as diferenças individuais destacadas, auxiliando a compreender diferentes características para planejamento, execução e direcionamento de treinamentos.

Este estudo realça a importância de uma abordagem personalizada no treinamento de atletas, onde estratégias de aceleração podem ser adaptadas com base nas características individuais dos sprinters. Ademais, os resultados apontam para a necessidade de mais pesquisas que explorem a relação entre as técnicas de aceleração e o desempenho em sprints, especialmente considerando as variáveis individuais dos atletas.

Referências

- CLAVEL, P. et al. Concurrent Validity and Reliability of Sprinting Force–Velocity Profile Assessed With GPS Devices in Elite Athletes. *International Journal of Sports Physiology and Performance*, v. 17, n. 10, p. 1527–1531, 11 Jul. 2022.

- EKLOM, B. Applied physiology of soccer. *Sports Med* 3. p. 50–60, 1986.
- FURUSAWA, K.; HILL, A. V.; PARKINSON, J. L. The Dynamics of “Sprint” Running. *Proceedings of The Royal Society B: Biological Sciences*, v. 102, p. 29–42, 1927.
- HICKS, D. et al Individualization of Training Based on Sprint Force-Velocity Profiles: A Conceptual Framework for Biomechanical and Technical Training Recommendations. *Strength and Conditioning Journal* 45(6), p. 711-715, 2023.
- HILL A.V. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol. Suppl.* 40:iv-vii, 1910.
- HOPKINS, W. G. et al. Progressive statistics for studies in sports medicine and exercise Science. *Medicine and Science in Sports and Exercise*, v. 41, n. 1, p. 3–12, jan. 2009.
- JOVANOVIĆ, M.; VESCOVI, J.: An R Package for Modeling Short Sprints. *International Journal of Strength and Conditioning*, v. 2, n. 1, 22 Jul. 2022.
- SAMOZINO, P. et al. Optimal Force–Velocity Profile in Ballistic Movements—Altius. *Medicine & Science in Sports & Exercise*, v. 44, n. 2, p. 313–322, 2012.
- THOMAS, C. et al. The diagnostic plot: A concept for identifying different states of iron Deficiency and monitoring the response to epoetin therapy. *Medical Oncology*, v. 23, n. 1, p. 23–36, 2006.
- VOLKOV, N. I.; LAPIN, V. I. Analysis of the velocity curve in sprint running. *Medicine And science in sports*, v. 11 4, p. 332–7, 1979.

Clusterizando jogadores de futebol em diferentes arquétipos

Lucca Augusto - lucca.aug@gmail.com

Resumo

Muitas vezes é necessário reposicionar uma peça no elenco de um clube, seja internamente, devido ao jogador titular ter se machucado ou até ter sido vendido, ou externamente, contratando um novo jogador. Em times bem encaixados, com uma padrão de jogo consolidado e eficaz, uma simples troca pode gerar sérios problemas, pelo novo jogador não conseguir desempenhar a função necessária ou pelo time se adaptar ao melhor possível do novo componente e mudar muito sua estrutura principal.

Para isso, queremos trocar uma peça por um jogador o mais próximo possível daquele que está de saída, e isso nem sempre é bem feito utilizando categorias mais simples como posição, é comum que um meia central desempenhe uma função de criação e em outro time tal função seja desempenhada por um volante, meia ofensivo ou ponta.

Achar tais reposições não é tão trivial com as categorias já existentes, portanto é proposto aqui a clusterização de jogadores em novos arquétipos, além da criação de uma métrica de distância entre eles.

1 Introdução

Em competições de futebol de elite, é comum notar jogadores que compartilham características semelhantes em seu estilo de jogo, como, por exemplo, uma participação ativa em jogadas, movimentação constante pelo campo ou a forma como se deslocam no gramado. O conceito de "posição" no futebol atual está em constante evolução, já que a presença de flutuações de jogadores em campo e diferentes tarefas para as mesmas posições tornam essa definição cada vez mais complexa. Nesse sentido, discutir o arquétipo de jogador pode ser útil para entendermos como os jogadores se encaixam em diferentes funções e estilos de jogo, contribuindo para a formação de equipes equilibradas e eficientes em campo.

Com isso, a busca por um jogador para completar um time deixou de ser uma simples questão de preencher um espaço vago, passando a envolver diferentes níveis de análise. Atualmente, é necessário levar em conta não apenas as características técnicas do jogador, mas também seu estilo de jogo, habilidades táticas e como ele se encaixa em um modelo de jogo específico. Dessa forma, é possível prever qual jogador seria o melhor contribuinte para o time em questão.

Em outras palavras, a escolha de um jogador passa a ser uma decisão estratégica baseada em dados, em vez de uma escolha intuitiva. Nesse sentido, o objetivo desse trabalho é utilizar diferentes dados para identificar os tipos comuns de jogadores e quais tarefas são mais relacionadas a cada arquétipo. Através de uma análise cuidadosa do desempenho dos jogadores em campo, é possível identificar padrões de comportamento que se repetem e associá-los a diferentes perfis de jogadores. Assim, é possível traçar um perfil mais preciso do tipo de jogador que melhor se adequa a cada posição e modelo de jogo, contribuindo para a formação de times mais competitivos e eficientes.

Esse tipo de modelo de agrupamento de jogadores apresenta uma vasta variedade de aplicações, podendo ser utilizado para contextualizar as estatísticas dos jogadores e permitindo que se comparem desempenhos entre diferentes posições. É possível analisar o desempenho de equipes e ligas ao longo do tempo, entendendo como as formações e funções evoluíram e se adaptaram. Outra aplicação importante é a possibilidade de identificar as necessidades de um clube em contratações, auxiliando na tomada de decisões estratégicas para a contratação de jogadores mais adequados ao perfil da equipe.

2 Trabalhos Relacionados

Existem trabalhos relacionados que abordam o mesmo problema de pesquisa proposto neste estudo, como Khobdeh, Yamaghani e Sareshkeh 2021 que faz tal agrupamento no ambiente do basquete, ou Ven 2018 e Poli, Ravenel e Besson 2022 que estão no ambiente futebolístico e seguem a ideia de agrupar jogadores em novos tipos, sendo o primeiro uma tarefa de clusterização e o segundo de classificação. As novas descobertas obtidas neste estudo têm relevância em relação aos trabalhos existentes, pois proporcionam uma perspectiva adicional sobre o agrupamento e análise de jogadores com base em seus padrões de desempenho durante as partidas de futebol, sendo assim, vamos discutir como essas descobertas se relacionam com os trabalhos anteriores, confirmado, refutando ou expandindo o conhecimento atual.

No trabalho Ven 2018, foi proposto o uso de algoritmos de agrupamento para clusterizar os jogadores com base em suas principais habilidades e pontos fracos. Nossos resultados confirmam essa abordagem ao aplicar técnicas de agrupamento e identificar padrões de desempenho distintos em cada cluster. No entanto, diferentemente do estudo mencionado, não nos concentrarmos exclusivamente nas habilidades dos jogadores, mas sim em uma ampla gama de eventos durante as partidas.

Da mesma forma, Poli, Ravenel e Besson 2022 categorizou os jogadores de acordo com seu perfil técnico e principais ações em campo. Nossos resultados estão em consonância com essa abordagem, pois também consideramos as estatísticas dos jogadores e seus eventos específicos para agrupá-los. Além disso, levamos em conta o contexto da equipe ao normalizar as estatísticas de acordo com a média geral de cada jogador, minimizando as diferenças causadas pelo ambiente em que atuam. Nossas descobertas buscam expandir o conhecimento

atual, fornecendo uma visão mais abrangente dos padrões de desempenho dos jogadores. Em vez de nos concentrarmos apenas nas habilidades ou características técnicas, analisamos um conjunto diversificado de eventos, como duelos, chutes, passes, entre outros, essa abordagem permite uma compreensão mais completa do desempenho dos jogadores durante as partidas e sua contribuição para os resultados do jogo.

Este estudo utiliza a análise de componentes principais, definida em Maćkiewicz e Ratajczak 1993, para reduzir a dimensionalidade dos dados e facilitar a visualização dos clusters, diferentemente das outras referências utilizadas. Essa técnica permitiu uma representação bidimensional dos jogadores, destacando as diferenças entre os grupos de forma mais clara.

As descobertas deste estudo buscam confirmar a validade das abordagens anteriores ao agrupar os jogadores com base em seus padrões de desempenho, também chamados de arquétipos. Ao mesmo tempo, expandem o conhecimento atual, considerando um conjunto diversificado de eventos e introduzindo técnicas de redução de dimensionalidade.

3 Metodologia

3.1 Tratamento dos Dados

Inicialmente foram removidos todos os jogadores com menos de 1000 minutos na temporada, de forma a evitar dados inconstantes de jogadores que atuaram pouco e portanto não teriam um padrão à ser extraído dos dados. Outra limpeza que foi feita é na posição dos jogadores, devido às características super específicas da posição, foram removidos todos os goleiros para que a comparação seja feita apenas entre os jogadores de linha. Obtendo uma tabela com todos os eventos gerados por jogadores de linha com pelo menos 1000 minutos jogados.

Agora já com as instâncias finais, é feito um tratamento nos atributos dessas instâncias, foram removidos os dados de evento dos grupos: 'Foul', 'Interruption', 'Goalkeeper leaving line', 'Offside' e 'Save attempt'. Tais atributos são considerados não muito úteis na caracterização do estilo de jogo de um jogador.

Finalmente nossa tabela final possui os seguintes atributos, onde o padrão é '{grupo_do_evento}_{sub-evento}': 'duel_Air duel', 'duel_Ground loose ball duel', 'duel_Ground defending duel', 'duel_Ground attacking duel', 'free kick_Corner', 'free kick_Free Kick', 'free kick_Free kick cross', 'free kick_Free kick shot', 'free kick_Throw in', 'pass_Cross', 'pass_Head pass', 'pass_High pass', 'pass_Launch', 'pass_Simple pass', 'pass_Smart pass', 'shot_Shot', 'others_Clearance'

Com esta tabela os dados são normalizados por jogo, a cada 90 minutos, de forma a equilibrar jogadores que tenham participação diferente, seja em quantidade de jogos ou em minutagem por jogo.

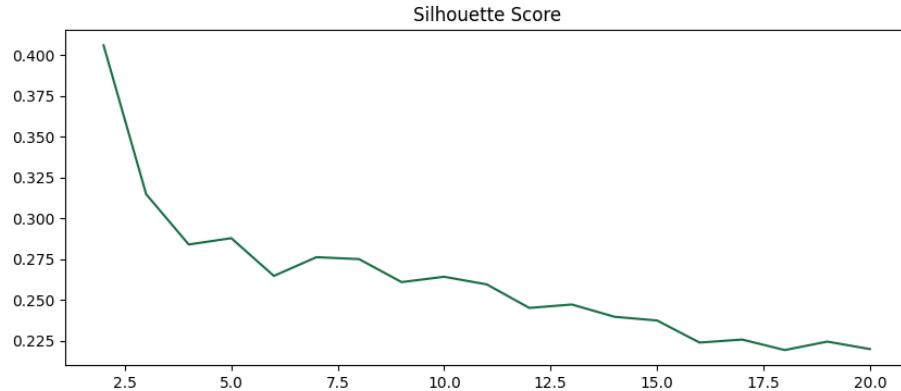


Figura 1: Silhouette Score para K variando de 2 à 20

3.2 Modelos

O modelo escolhido para a definição dos arquétipos dos jogadores foi o K-Means, por ser um algoritmo de clusterização simples e eficiente. Tal algoritmo necessita da definição de um hiperparâmetro essencial que é o K, número de clusters a serem agrupados. Para isso rodamos o algoritmo com múltiplos valores e calculamos o Silhouette Score, obtendo o gráfico da Figura 1.

Partindo da Figura 1 foi decidido utilizar $K = 6$ já que a partir deste valor o erro não cai tanto e mantém o número de clusters mais baixo, resultando em um maior número de jogadores no mesmo cluster. Para visualização foi feita uma PCA de dois componentes entre todos os atributos, obtendo a Figura 2

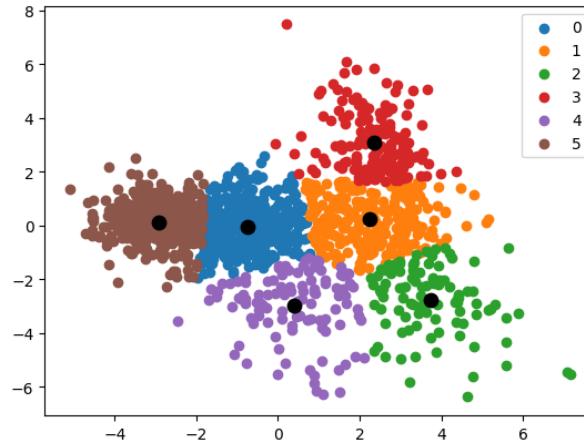


Figura 2: Distribuição em duas dimensões onde o ponto preto representa o centroide do cluster e os pontos coloridos as instâncias

4 Resultados

Dados os resultados obtidos e apresentados na subseção 3.2, é possível identificar os principais fatores que caracterizam os clusters e interpretar quais suas funções, como uma espécie de Explanable AI. Seguem as análises de cada grupo a baixo onde a referência é o centroide do respectivo cluster.

0 (Referência: Marco D'Alessandro): Possuem grande participação em chutes, duelos ofensivos, passes inteligentes e cruzamentos. São jogadores claramente ofensivos capazes de fazer diversas funções no ataque, mostrando bastante versatilidade nas ações realizadas.

1 (Referência: Fallou Diagne): Altos índices de cortes, passes de cabeça e lançamentos, além de uma quantidade de duelos defensivos acima da média. Essas características indicam a tendência de que jogadores deste grupo tenham um papel mais defensivo em seus times, sem grandes contribuições no ataque.

2 (Referência: Kevin Lasagna): Se destaca pela quantidade de duelos, especialmente aéreos e ofensivos. Também apresenta um alto número de finalizações. Jogadores deste grupo atuam mais próximos do gol e não participam muito da construção de jogadas, possuindo tarefas mais terminais, tipicamente, mas não necessariamente, realizadas pelos centroavantes.

3 (Referência: Manu Trigueros): Também apresentam valores altos em situações de bola parada e possuem foco na realização de passes. Uma interpretação sobre a composição desse grupo é que são jogadores equilibrados, que não são incisivos ofensivamente, mas que também não são especialistas na defesa.

4 (Referência: Isco): Caracterizado principalmente pelas ações de bola parada, como escanteios e faltas, tanto em chutes diretos quanto em cruzamentos, é o grupo que mais distribui passes inteligentes e cruzamentos. É um grupo mais restrito dos jogadores que são os principais cobradores de bolas paradas em seus clubes, combinando essa característica com a tendência de serem jogadores criativos e que participam frequentemente da armação do jogo.

5 (Referência: Adam Smith): Maior destaque para as ações de duelos defensivos e cobranças de lateral. É possível inferir que esses jogadores atuam nas regiões laterais do campo e se dedicam mais ao lado defensivo do jogo.

Analizando a distribuição das instâncias e jogadores-chave, que são jogadores mais famosos e que já possuem o estilo de jogo amplamente conhecido, podemos ver alguns padrões relativamente esperados, como por exemplo: a mesma clusterização entre os jogadores Karim Benzema e Luis Suárez, que atuam como centroavantes que voltam pra participar da construção da jogada ou até o jogador mais próximo de Neymar Júnior ser o Lionel Messi, pontas de qualidade no drible que também possuem alta criatividade na criação de jogadas, ambos no cluster em que Isco é o centroide, outro jogador com características semelhantes.

Chama a atenção alguns dados um pouco contraditórios com o que se pensa o senso-comum, como o jogador zagueiro Marquinhos ter Leon Goretzka como jogador mais próximo, sendo ele um meia com algumas atuações até em setores mais ofensivos do campo. Outro caso ainda mais curioso é o jogador de menor distância de Cristiano Ronaldo ser o Ahmed Hegazy, um zagueiro egípcio que

atuava pelo West Brom nesta temporada, time que foi o lanterna da Premier League na ocasião.

5 Conclusão

Os principais pontos do projeto eram de fornecer um método que fosse capaz de agrupar jogadores em funções táticas definidas, independente da posição dos jogadores, e de que tendo essas funções táticas definidas fosse possível calcular a distância estatística entre os jogadores.

Esse trabalho é importante porque fornece uma ferramenta que contribui em diversas situações dentro de um clube de futebol. O método que desenvolvemos, pode ser utilizado para auxiliar em tomadas de decisão já que é possível calcular qual o jogador mais próximo a um outro jogador. Assim, por exemplo, quando é preciso escolher um jogador para substituir um outro que acabou deixando o elenco, ou onde deseja-se profissionalizar um jogador da base, ou ainda para auxílio nas funções de scouting, nós sabemos pelo método quais são os jogadores mais próximos e também quais são os jogadores que estão mais distantes estatisticamente de modo a descartar algumas opções e fornecer as mais assertivas quanto ao papel do jogador desejado.

Uma proposta de continuidade de estudo para este projeto seria de formular diferentes conjuntos de atributos que foram utilizados para treinar o algoritmo. Dessa forma, seria possível estudar e comparar os impactos resultantes desses diferentes conjuntos nos indivíduos presentes em cada cluster e na distância estatística entre eles, visando responder algumas perguntas, por exemplo, Como essa escolha influencia na definição das funções táticas ? Os indivíduos que estavam agrupados se mantêm próximos quando são mudados os conjuntos de atributos ? Além disso, será possível analisar se existem relações entre os clusters formados de acordo com as diferentes escolhas de conjuntos de atributos iniciais, com a finalidade de entender se existe um método de escolha que seja o melhor possível.

Referências

- Khobdeh, Soroush Babaee, Mohammad Reza Yamaghani e Siavash Khodaparast Sareshkeh (2021). “Clustering Of Basketball Players Using Self-Organizing Map Neural Networks.” Em: URL: <https://api.semanticscholar.org/CorpusID:237612596>.
- Ven, Eline van de (2018). “Clustering soccer players to find the drivers of soccer team performance”. Em: URL: <https://api.semanticscholar.org/CorpusID:56123257>.
- Poli, Raffaele, Loïc Ravenel e Roger Besson (2022). *Technical profiling of football players*. URL: <https://football-observatory.com/IMG/sites/mr/mr72/en/>.

Maćkiewicz, Andrzej e Waldemar Ratajczak (1993). “Principal components analysis (PCA)”. Em: *Computers Geosciences* 19.3, pp. 303–342. ISSN: 0098-3004. DOI: [https://doi.org/10.1016/0098-3004\(93\)90090-R](https://doi.org/10.1016/0098-3004(93)90090-R). URL: <https://www.sciencedirect.com/science/article/pii/009830049390090R>.

GoldFit Soccer: Sistema Especialista de Identificação de Talentos no Futebol

Francisco Z. Werneck¹, Emerson F. Coelho¹, Marcelo O. Matta², António J. Figueiredo³, Rodrigo C.P. Silva⁴

¹Escola de Educação Física – Universidade Federal de Ouro Preto (UFOP) – Ouro Preto
– MG – Brasil

²Projeto Futebol UFJF – Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora,
MG, Brasil

³Universidade de Coimbra (UC) – Coimbra – Portugal

⁴Departamento de Computação
Universidade Federal de Ouro Preto (UFOP) – Ouro Preto, MG – Brasil

{f.zacaron@ufop.edu.br, emersoncoelho@hotmail.com, marcelo.matta@ufjf.edu.br,
afigueiredo@fcdef.uc.pt,rodrigo.silva@ufop.edu.br}

1. Introdução

Identificar corretamente jovens futebolistas talentosos com potencial para se tornarem jogadores profissionais é um desafio constante para os treinadores e uma necessidade para os clubes. Estudos têm fornecido evidências empíricas sobre a relevância prognóstica das baterias de testes e do “olho do treinador” neste processo. No entanto, pouco é sabido sobre a validade de modelos multidimensionais para identificar os atletas com maior potencial, levando-se em conta a maturação biológica.

2. Objetivo

O objetivo do estudo foi criar um modelo de avaliação do potencial de jovens futebolistas, testar sua validade preditiva e desenvolver a versão web de um sistema inteligente de identificação de talentos no futebol.

3. Métodos

Participaram do estudo 132 jovens futebolistas brasileiros do sexo masculino das categorias sub13 ao sub17 de um programa de desenvolvimento de talentos ($14,7 \pm 1,4$ anos; nível competitivo regional). Os atletas realizaram a bateria de testes do Projeto Atletas de Ouro® para medida de indicadores antropométricos (massa corporal, estatura, envergadura, dobras cutâneas), físico-motores (agilidade, CMJ, velocidade 20m, YoYo IR-1), psicológicos (ASCI-28, Competência percebida, SOQ), ambientais (apoio familiar, experiência esportiva), maturacionais (Porcentagem atingida da estatura adulta prevista, idade prevista no pico de velocidade de crescimento (PVC), técnicos (Drible) e táticos (TACSSIS). Os treinadores avaliaram seus atletas quanto aos aspectos

intangíveis do potencial esportivo e a expectativa de sucesso futuro. O follow-up dos atletas foi de 5 anos. O critério de sucesso dos jogadores após este período foi operacionalizado por alcançarem pelo menos a categoria sub20 de um clube do Campeonato Brasileiro ou ter se profissionalizado (a taxa de sucesso foi de 15,9%). Todas as variáveis quantitativas foram padronizadas em escores Z, utilizando valores de referência da literatura internacional e, em seguida, foram transformadas em valores percentis. A partir da análise exploratória dos dados e do modelo de performance no futebol, criou-se um índice híbrido, composto por testes objetivos, a percepção do atleta e o “olho do treinador” para estimativa do potencial dos atletas, chamado Gold Score Soccer [GSS = (agilidade*2,0 + resistência aeróbica*3,0 + drible*3,0 + habilidades táticas*2,0 + confiança/motivação*1,0 + suporte familiar*1,0 + aspectos intangíveis*5,0 + expectativa de sucesso futuro*2,0)/19)], com variação de 0 a 100. Os pesos das variáveis do GSS foram escolhidos a partir da análise dos dados, evidência científica e opinião de experts. Os atletas foram classificados quanto ao GSS em Bronze (GSS<40), Prata (40<GSS<70) ou Ouro (GSS≥70). Uma regressão logística binária estimou a probabilidade de sucesso (sim ou não) dos jogadores, utilizando como preditores o GSS, a lateralidade (ambidestro) e a idade do PVC.

4. Resultados

Confirmando a validade prognóstica univariada, os futebolistas de sucesso apresentaram melhores resultados comparados aos não-sucesso em relação ao teste de velocidade 20m ($p=0,009$), agilidade ($p=0,04$), salto contramovimento ($p=0,04$), Yo-Yo IR-1 ($p=0,001$), drible ($p<0,001$), habilidades psicológicas ($p<0,05$) e “olho do treinador” ($p<0,001$). Futebolistas que obtiveram sucesso apresentaram maior GSS quando comparados aos não sucessos ($60,0\pm8,3$ vs. $45,0\pm13,6$; $p<0,001$; $d=1,10$). Os futebolistas classificados como Ouro tinham uma chance 20 vezes maior de obterem sucesso. A probabilidade de sucesso futuro dos atletas foi obtida pelo seguinte modelo logístico: $P(Y) = 1 / [1 + \exp (-(-23,825 + 0,163*(Gold\ Score\ Soccer) + 1,664*(Ambidestro: não = 0; sim=1) + 0,872* (PVC, em anos)))]$, onde valores de $P(Y)>0,16$ classifica o atleta como Sucesso e valores $<0,16$ como não sucesso. Este modelo logístico multidimensional foi denominado GoldFit Soccer. De acordo com a área sobre a curva ROC, o modelo apresentou alta validade prognóstica (sensibilidade = 85,7%; especificidade = 83,8%; acurácia = 84,1%) e excelente validade discriminante (AUC = 0,93 [0,87–0,98]). Para organização, análise, armazenamento dos dados e apresentação de resultados, este modelo foi implementado em sistema Web. Para o front end da aplicação foram utilizadas a linguagem JavaScript e a biblioteca React; para o desenvolvimento da interface e para o back end foram utilizados a linguagem Python e o framework Django, para o desenvolvimento do banco de dados (Figura 1).

5. Conclusões

Foi desenvolvida uma ferramenta inovadora de identificação de talentos no futebol chamada GoldFit Soccer, que teve 84% de precisão na previsão de sucesso futuro dos atletas. Os resultados demonstraram que o GoldFit Soccer é uma ferramenta científica multidimensional válida que proporciona alta validade prognóstica e pode ser útil para melhorar o processo de identificação de talentos no futebol. Numa área tão imprevisível como a identificação e desenvolvimento de talentos, onde os maus

julgamentos são sempre um problema para os treinadores, o GoldFit Soccer é uma ferramenta útil para minimizar o impacto de potenciais erros.

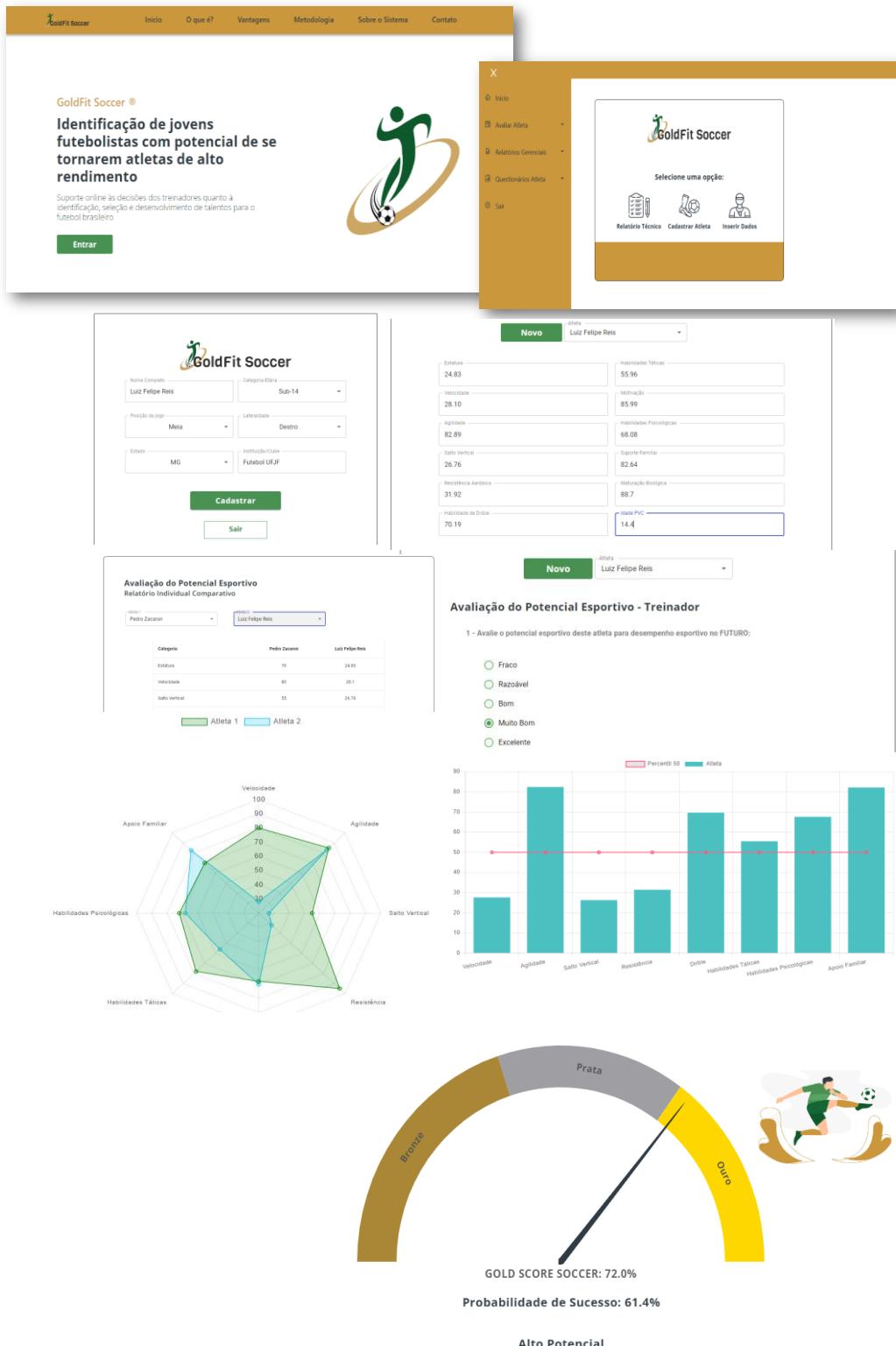


Figura 1. Imagem representativa da versão Web do GoldFit Soccer.

6. Aplicações Práticas

O GoldFit Soccer foi desenvolvido para que os grandes clubes de futebol identifiquem jovens talentos com maior assertividade o mais precocemente possível, através de uma análise completa dos indicadores de potencial dos atletas. É uma ferramenta dinâmica que auxiliará treinadores e demais profissionais tanto na identificação de talentos quanto no processo de seleção de jovens jogadores de futebol. O clube que utilizar o GoldFit Soccer pode obter vantagens competitivas, investindo com maior segurança nos jovens jogadores e direcionando a comissão técnica para o desenvolvimento de jovens atletas com maior potencial de sucesso. As aplicações práticas do GoldFit Soccer são: (i) identifica jovens futebolistas com maior potencial de sucesso futuro; (ii) minimiza os erros de seleção e a perda de potenciais talentos; (iii) emite relatórios individualizados, observando os pontos fortes e fracos a serem desenvolvidos; (iv) emite relatórios gerenciais que auxiliam os treinadores na gestão do potencial esportivo dos seus atletas, otimizando o processo de treinamento; (v) monitora os efeitos do treinamento e a evolução do potencial esportivo do atleta; (vi) pode ser usado em qualquer centro de treinamento/clube do país; (vii) maximiza investimentos na formação de novos talentos para o futebol brasileiro. Esta ferramenta vem sendo aplicada pelo nosso grupo de pesquisa no Fluminense Futebol Clube e os resultados preliminares encontrados indicam que o GSS é valido para discriminar a seleção de jovens futebolistas de elite sub14 de uma temporada para outra e para a seleção brasileira sub15. Após o seguimento destes atletas pelo período de 5 anos, será verificada a validade do GoldFit Soccer na previsão de sucesso futuro.

References

- Sieghartsleitner, R, Zuber, C, Zibung, M, Conzelmann, A. (2019). Science or coaches' eye? – both! Beneficial collaboration of multidimensional measurements and coach assessments for efficient talent selection in elite youth football. *Journal of Sports Science & Medicine*, 18(1): 32-43.
- Vidigal, E.C, Silva, F.F, Rodrigues, T.L.A, ... Coelho, E.F., Werneck, F.Z. (2022). “Coach’s Eye”: Psychological and tactical skills discriminate sporting potential of young soccer players. *Revista Brasileira de Cineantropometria & Desempenho Humano*, 24. <https://doi.org/10.1590/1980-0037.2022v24e91439>
- Williams, A.M, Reilly, T. (2000). Talent identification and development in soccer. *Journal of Sports Sciences*, 18(9): 657–667.
- Werneck, FZ; Coelho, EC; Ferreira, RM. (2020). Manual do Jovem Atleta: Da Escola ao Alto Rendimento. Curitiba: CVR.
- Werneck FZ, Coelho EF, Miranda L. (2022). Projeto Atletas de Ouro®: an innovation to sporting talent detection. *E-legis*. 15:89-104.
- Werneck FZ, Coelho EF (2022). Gold Score Soccer: um modelo científico para identificação de talentos no futebol. *The Brazilian Journal of Soccer Science*. 2022;15(supl):17.

Comparação entre a formação de base no futebol de elite masculino e feminino no Brasil

Guilherme Machado^{1,2,3}, Sixto González-Víllora², Israel Teoldo¹

¹ Núcleo de Pesquisa e Estudos em Futebol (NUPEF), Universidade Federal de Viçosa (UFV), Brasil.

² EDAF Research Group, Facultad de Educación, Universidad de Castilla-La Mancha (UCLM), Espanha.

³ Federação Paulista de Futebol (FPF), Brasil.

1. Introdução

O futebol feminino está em crescente desenvolvimento no cenário internacional e nacional, levando a um maior aporte financeiro e à estruturação e implantação de processos já presentes mais comumente no futebol masculino (WILLIAMS; FORD; DRUST, 2020). No entanto, é importante reconhecer que homens e mulheres têm diferenças biológicas, psicológicas e sociais que podem afetar sua participação e desempenho no futebol (TEOLDO; MACHADO; CARDOSO, 2023). Portanto, é necessário adaptar a abordagem de treinamento e desenvolvimento para atender às necessidades específicas das jogadoras de futebol. Dessa forma, para otimizar a adaptação da estrutura do futebol de base do futebol masculino já consolidadas no Brasil às especificidades do futebol feminino é necessário conhecer as semelhanças e diferenças organizacionais de ambas. Estudos recentes realizados com atletas de elite no Brasil também mostraram diferenças socioeconômicas no contexto de desenvolvimento de talentos no futebol masculino (TEOLDO; CARDOSO, 2021) e feminino (TEOLDO et al., 2023), ao identificarem o “mapa do talento” no Brasil. Uma diferença importante entre o futebol masculino e feminino é a quantidade de horas vivenciadas durante a formação em atividades que são fundamentais para a formação esportiva, como: 1) treinamento estruturado no futebol (também conhecido como prática deliberada); 2) futebol recreativo (também conhecido como jogo deliberado ou o “futebol de rua”); 3) participação em competição (jogos de campeonatos ou torneios); e 4) treinamento estruturado no futsal. Estudos recentes realizados com atletas de futebol de elite feminino no contexto internacional e nacional têm demonstrado que a interação entre as três primeiras atividades citadas anteriormente (treinamento, recreação e competição no futebol) é fundamental ao longo da formação para que as atletas alcancem a profissionalização (MACHADO; GONZÁLEZ-VÍLLORA; TEOLDO, 2023).

2. Objetivo

O objetivo deste trabalho foi comparar a vivência acumulada (ambiente recreativo, de treinamento e competitivo) ao longo da formação esportiva no futebol feminino e masculino.

3. Métodos

Participaram deste estudo 180 jogadores de futebol de elite sub-18 do Brasil de futebol feminino ($n = 98$) e masculino ($n = 82$) participantes de equipes de base de clubes da primeira divisão do futebol masculino e de clubes das séries A1 e A2 do futebol feminino. O trabalho foi aprovado pelo Comitê de Ética da Universidade Federal de Viçosa, protocolo 4.924.597. O

Questionário de Histórico do Jogador (FORD; YATES; WILLIAMS, 2010) foi utilizado para coletar informações sobre a participação em diferentes atividades de futebol ao longo da formação. A validade e confiabilidade deste questionário demonstrou ser aceitável para diferentes culturas (ingleses, alemães e suíços) (FORD; YATES; WILLIAMS, 2010; GULLICH, 2014; HAUGAASEN et al., 2014). Cinco variáveis relacionadas a participação em diferentes atividades esportivas foram avaliadas: 1) competição no futebol; 2) treinamento de futebol; 3) jogo deliberado no futebol (atividades recreativas sem a presença de um adulto, como jogar futebol na rua com os amigos); 4) treinamento de futsal; e 5) treinamento de futebol somado ao treinamento de futsal. As atividades de formação foram analisadas por três períodos de idade: 1) infância (6-12 anos); 2) início da adolescência (13-15 anos) e 3) Somatório de toda a formação esportiva do atleta até o sub-18. Para análise estatística foi verificada a distribuição dos dados através do teste de Kolmogorov-Smirnov e utilizado o teste de Mann Whitney para comparação entre os sexos.

4. Resultados e Discussão

Os resultados da tabela 1 destacam algumas “defasagens” na vivência do futebol feminino comparado ao futebol masculino no ensino estruturado (competições e treinamento de futebol) durante a infância (6-12 anos) e o início da adolescência (13-15 anos). Percebe-se que durante a infância atletas de futebol feminino acumularam em média 20% da quantidade de horas em competições comparadas aos homens e durante a adolescência apenas 29% em média. Já em relação ao treinamento de futebol essa quantidade de horas em treinamento acumulada pelas mulheres ao longo desses períodos de formação varia de 48% em média durante a infância e 57% em média durante a adolescência comparadas aos homens.

Tabela 1. Número de horas acumuladas (média ± desvio padrão) em diferentes atividades entre jogadores de futebol masculino e feminino sub-18.

Váriaveis	Masculino (n = 82) Média ± DP (h)	Feminino (n = 98) Média ± DP (h)	% de horas do feminino comparado ao masculino	p
Infância (6-12 anos)				
Competição Futebol	552 ± 852	113 ± 169	20%	< 0,001*
Treinamento Futebol	1002 ± 961	485 ± 582	48%	< 0,001*
Recreação futebol	1612 ± 1708	735 ± 1219	46%	< 0,001*
Treinamento Futsal	457 ± 548	458 ± 663	100%	0,991
Treinamento Futebol + Futsal	1462 ± 1245	933 ± 1020	64%	0,003*
Início da Adolescência (13-15 anos)				
Competição Futebol	398 ± 517	117 ± 135	29%	< 0,001*
Treinamento Futebol	911 ± 559	517 ± 379	57%	< 0,001*
Recreação futebol	418 ± 563	332 ± 588	79%	0,324
Treinamento Futsal	76 ± 142	252 ± 315	332%	< 0,001*
Treinamento Futebol + Futsal	979 ± 567	764 ± 503	78%	0,008*
Total acumulado até o sub-18				
Competição Futebol	1031 ± 1412	304 ± 308	29%	< 0,001*
Treinamento Futebol	2125 ± 1507	1368 ± 1007	64%	< 0,001*
Recreação futebol	2217 ± 2260	1179 ± 1867	53%	0,001*
Treinamento Futsal	566 ± 675	791 ± 1025	140%	0,002*
Treinamento Futebol + Futsal	2689 ± 1798	2144 ± 1599	80%	0,033*

Essa “defasagem” na formação esportiva em termos de vivência de treino e competição tende a implicar negativamente no desempenho individual das atletas e coletivo das equipes quando comparamos atletas de uma mesma faixa etária de sexos diferentes. Isso porque essa diferença de vivências esportivas impacta no desenvolvimento tático, técnico, físico e psicológico individual de atletas de futebol feminino, quando comparado ao masculino

(TEOLDO; MACHADO; CARDOSO, 2023). Portanto, esse é um dos principais fatores que leva à necessidade de adaptação de um currículo de formação do futebol masculino de base para o futebol feminino.

Por outro lado, verificamos que o treinamento no futsal vem exercendo um papel importante na formação estruturada de atletas de futebol feminino no Brasil. Percebemos que durante a infância, o tempo de treino médio acumulado no futsal é o mesmo entre o futebol feminino e masculino, enquanto no início da adolescência meninas acumulam em média cerca de três vezes mais horas no futsal quando comparadas aos homens. Esse tempo maior de envolvimento das mulheres no treinamento de futsal, provavelmente é fruto de um sistema de formação no futebol feminino ainda pouco desenvolvido, que não consegue absorver atletas no início da adolescência interessadas na prática regular da modalidade, devido a falta de clubes que possuam categorias nessas idades (ex.: sub-13 e sub-15). Os resultados sugerem que o treinamento no futsal vem sendo uma forma alternativa das mulheres se manterem ativas, através da prática de um esporte similar. Entretanto, é fundamental viabilizar melhores condições para o futebol feminino, através da criação de competições e categorias específicas para essas idades (13-15 anos), uma vez que nessas idades a progressão de ampliação do espaço e número de jogadores que o futebol proporciona é fundamental para o desenvolvimento tático, técnico, físico e psicológico das atletas, que não pode ser substituído unicamente pelo treinamento de futsal (TEOLDO; MACHADO; CARDOSO, 2023).

Dessa forma, o futsal pode ser utilizado como um aliado na formação esportiva de atletas de futebol feminino durante a infância, contribuindo para uma formação denominada de diversificação especializada, que tem sido associada a maior propensão de profissionalização de atletas de futebol feminino ao longo da sua formação (MACHADO; GONZÁLEZ-VÍLLORA; TEOLDO, 2022). Entretanto, é necessário que esteja claro as potencialidades do futsal de desenvolvimento de habilidades para o futebol, mas também as limitações em relação ao que pode ser desenvolvido. É importante destacar a sua limitada contribuição para o ensino de algumas habilidades técnicas mais específicas do futebol (ex.: domínio com a parte interna/externa do pé, realização de passes e chutes de longa distância) e principalmente de aspectos mais coletivos do jogo, como táticas coletivas (5x5 até 11x11) e intersetoriais. Nesse contexto, destacamos que a partir do início da adolescência (por volta dos 13 anos de idade) o jogo coletivo ganha cada vez mais importância para o ensino do futebol e para o desempenho coletivo das equipes e individuais dos jogadores (TEOLDO; GUILHERME; GARGANTA, 2021).

Dessa forma, é importante ressaltar que no atual cenário de estruturação do futebol, durante o início da adolescência a participação no treinamento de futsal ainda exerce um papel importante de formação para as atletas do futebol feminino, especialmente porque apresentam uma “defasagem” de treinamento estruturado no futebol durante a infância quando comparadas aos homens. Entretanto, a partir do início da adolescência, por ser uma fase que o aspecto coletivo do jogo ganha mais importância para o desempenho, o treinamento do futsal deve ser utilizado como uma formação complementar das atletas, mas a quantidade de treinamento estruturado deve ser predominantemente no futebol, por permitir o treinamento mais específico em estruturas, número de jogadoras e espaços maiores.

Considerando a menor vivência acumulada de atletas do futebol feminino em diferentes contextos de formação no futebol (recreativo, treinamento e competitivo) quando comparadas à atletas masculinos, é esperado que atletas femininas de mesma idade cronológica possuam menor competência técnicas e táticas do que atletas masculinos. Para suprir essa “defasagem” de formação sugere-se a implantação de *programas de desenvolvimento individual* para todas

as categorias de futebol de base feminino, buscando uma abordagem mais individualizada de melhora em função da necessidade de cada atleta.

5. Conclusão

A partir das informações apresentadas acima concluímos que é importante adaptar o currículo de formação de base para o futebol feminino e implementar mudanças estruturais como o aumento de competições e categorias etárias no início da adolescência (sub-13 e sub-15). Esses achados permitem discutir baseado em dados empíricos adaptações práticas que podem ser implementadas para atender às necessidades específicas das jogadoras de futebol e promover a igualdade de oportunidades no esporte.

6. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Este estudo também foi financiado pela Lei de incentivo ao Esporte do Governo de Minas Gerais, pelo Programa Academia & Futebol do Ministério da Cidadania, através da Secretaria Nacional de Futebol e Defesa dos Direitos do Torcedor, pela FAPEMIG, CNPQ, FUNARBE e pela PPG, CCB e Reitoria da Universidade Federal de Viçosa.

7. Referências

- FORD, P. R.; YATES, I.; WILLIAMS, A. M. An analysis of practice activities and instructional behaviours used by youth soccer coaches during practice: Exploring the link between science and application. *Journal of Sports Sciences*, v. 28, n. 5, p. 483–495, 2010.
- GÜLLICH, A. Many roads lead to Rome – Developmental paths to Olympic gold in men's field hockey. *European Journal of Sport Science*, v. 14, n. 8, p. 763–771, 2014.
- HAUGAASEN, M.; TOERING, T.; JORDET, G. From childhood to senior professional football: elite youth players' engagement in non-football activities. *Journal of Sports Sciences*, v. 32, n. 20, SI, p. 1940–1949, dez. 2014.
- MACHADO, G.; GONZÁLEZ-VÍLLORA, S.; TEOLDO, I. Contribution of deliberate practice, play, and futsal to the acquisition of decision-making skills in Brazilian professional female soccer players. *International Journal of Sport and Exercise Psychology*, 2022.
- MACHADO, G.; GONZÁLEZ-VÍLLORA, S.; TEOLDO, I. The relationship between deliberate practice, play, and futsal in childhood and adolescence and the development of different decision-making skills in professional female soccer players. *Psychology of Sport & Exercise*, v. 68, n. June, p. 102470, 2023.
- TEOLDO, I. et al. Talent map of female soccer: How does the birthplace and birthdate impact the participation of soccer players in Brazilian Serie A1 Championship? *Journal of Human Sport & Exercise*, p. 13, 2023.
- TEOLDO, I.; CARDOSO, F. Talent map: how demographic rate, human development index and birthdate can be decisive for the identification and development of soccer players in Brazil. *Science and Medicine in Football*, v. 5, n. 4, p. 293–300, 2021.
- TEOLDO, I.; GUILHERME, J.; GARGANTA, J. Para um futebol jogado com ideias: Concepção, treinamento e avaliação do desempenho tático de jogadores e equipes. 2. ed. Curitiba: Appris, 2021.

TEOLDO, I.; MACHADO, G.; CARDOSO, F. Caderno metodológico para a formação de jogadores mais inteligentes e criativos para o jogo: Case UFV Soccer Academy. Curitiba: Appris, 2023.

WILLIAMS, A. M.; FORD, P. R.; DRUST, B. Talent identification and development in soccer since the millennium. *Journal of Sports Sciences*, v. 38, n. 11–12, p. 1199–1210, 2020.

Rankeando Jogadores de Futebol: Criação de um Índice de Performance Baseado em Aprendizado de Máquina

José Américo Pereira Antunes^{1,2,3}

¹ Banco Central do Brasil

² COPPEAD Instituto de Pós-Graduação e Pesquisa em Administração, Universidade Federal do Rio de Janeiro (UFRJ)

MEGE Mestrado em Economia e Gestão Empresarial, Universidade Cândido Mendes (UCAM) – Rio de Janeiro – RJ - Brasil

tesouraria.rj@gmail.com

Abstract. This study develops a classification model that uses unsupervised and supervised learning to rank football players according to their performance and position on the pitch. The model works in two phases: training and classification. The training phase involves the 5 main European leagues, while the classification phase includes any other leagues of interest, at the user's discretion. The model automates the initial phase of the scouting process. In addition to the classification model object of this study, a Python app that automates queries was developed and made publicly available.

Resumo. Este estudo desenvolve um modelo de classificação que utiliza aprendizado não-supervisionado e supervisionado para ranquear jogadores de futebol de acordo com seu desempenho e posição em campo. O modelo trabalha em duas fases: treinamento e classificação. A fase de treinamento envolve as 5 principais ligas europeias, enquanto a fase de classificação inclui as demais ligas de interesse, a critério do usuário. O modelo automatiza a fase inicial do processo de scouting. Em complemento ao modelo de classificação objeto deste estudo, uma ferramenta em Python que automatiza as consultas foi desenvolvida e disponibilizada publicamente.

1. Problema

A tecnologia pode alavancar o desempenho dos clubes e permitir que façam mais com menos. Brentford FC é um bom exemplo. Clube de futebol intensamente orientado por dados, o Brentford FC alcançou a promoção à Premier League em 2021. O sucesso do clube chamou atenção para esta nova abordagem de gestão do futebol e colocou em dúvida a forma convencional de fazê-lo.

A tecnologia permite, entre outros, que os clubes sejam mais eficientes na descoberta de talentos. Há mais de uma centena de ligas profissionais e dezenas de milhares de jogadores de futebol em atividade. Nesse oceano de possibilidades, faz-se necessária uma etapa preliminar de mineração e identificação de talentos para posterior aprofundamento. Nesse contexto, proponho a seguinte pergunta de pesquisa:

Como desenvolver um índice de performance que permita ranquear jogadores segundo o desempenho em campo?

2. Objetivo Geral

Este estudo desenvolve um modelo de classificação destinado a ranquear jogadores de futebol (Pappalardo *et al*, 2019) e orientar a prospecção de talentos para o clube. O experimento usa dados do Wyscout das cinco principais ligas europeias para treinar um algoritmo de aprendizado de máquina e construir um modelo de classificação, que é então aplicado a um conjunto de ligas internacionais qualquer para ranquear jogadores.

3. Metodologia

Na fase de treinamento, os jogadores das cinco principais ligas europeias recebem ratings atribuídos por meio de aprendizado não supervisionado, enquanto na fase de classificação os jogadores recebem ratings atribuídos por meio de aprendizado supervisionado. O passo-a-passo encontra-se detalhado nas subseções.

3.1. Posições e Perfis/Funções Correspondentes em Campo

Os jogadores são classificados de acordo com a posição mais comum em campo, conforme atribuído pelo Wyscout e detalhado na figura 1.

EXTREMO LWF 17/18/19	ATACANTE CF 20/21/22	EXTREMO RWF 17/18/19
EXTREMO LAMF LW 17/18/19	MEIA AMF 15/16	EXTREMO RAMF RW 17/18/19
	2º VOLANTE LCMF 12/13/14	
	RCMF 12/13/14	
LATERAL LWB LB 3/4/5	1º VOLANTE LDMF DMF RDMF 9/10/11	LATERAL RWB RB 3/4/5
	ZAGUEIRO LCB CB RCB 6/7/8	
	GOLEIRO GK 1/2	

Figura 1 – Posições mais jogadas pelo jogador, de acordo com a terminologia adotada pela Wyscout.

As nove posições dão origem a vinte e dois perfis ou funções em campo, conforme tabela 1.

Tabela 1 – Posições e Perfis/Funções correspondentes

Posições	Perfis/Funções em campo
1/2 – Goleiro	1 – Goleiro Clássico 2 – Goleiro Líbero
3/5 – Lateral/Ala	3 – Lateral Defensivo 4 – Lateral Ofensivo 5 – Lateral Equilibrado
6/8 – Zagueiro	6 – Zagueiro Clássico 7 – Zagueiro Construtor 8 – Zagueiro Equilibrado
9/11 – Primeiro Volante	9 – Primeiro Volante Clássico 10 - Primeiro Volante Construtor 11 - Primeiro Volante Equilibrado
12/14 - Segundo Volante	12 – Segundo Volante Box to Box 13 - Segundo Volante Organizador 14 - Segundo Volante Equilibrado
15/16 – Meia	15 – Meia Organizador 16 – Meia Atacante
17/19 - Extremos	17 – Extremos Organizador 18 – Extremo Tático 19 – Extremo Agudo
20/22 – Atacante	20 – Atacante Referência 21 – Atacante Móvel 22 – Segundo Atacante

3.2. Matriz de *Features*, Base de Treinamento e Base de Classificação

Os passos seguintes no desenvolvimento do modelo envolvem:

- i) Determinação das *features*: cada um dos vinte e dois perfis/funções derivados das nove posições recebe um conjunto de *features*, baixadas do Wyscout (Wyscout, 2022), relacionadas ao perfil e com capacidade estatística para discriminar boas e más performances do perfil sob análise. Abaixo, exemplo para o perfil Atacante Referência.

Tabela 2 - Mapeamento de *features* segundo a posição/perfil do jogador em campo

Features	Atacante Referência
Duelos_Aéreos_Ganhos	✓
Duelos_Ofensivos_Ganhos	✓
xG	✓
Conversão_Gols	✓
Conversão_xG	✓
Ameaça_Ofensiva	✓
xA	✓
Deep_Completions	✓
Passes_Chave	✓
Passes_ÁreaPênalti_Certos	✓

- ii) Definição da base de treinamento: todos os jogadores das cinco principais ligas europeias (Inglaterra, Alemanha, Espanha, Itália e França) com pelo menos 900 minutos jogados, em cinco temporadas completas 2018/19, 2019/20, 2020/21, 2021/22, 2022/23.

iii) Rotulagem da base de treinamento: utilização de algoritmo de aprendizado não-supervisionado por meio da técnica de redução de dimensionalidade, denominada Análise de Componente Principal (Pearson, 1901; Hotelling, 1933), para determinação do Índice de Performance dos jogadores (Al-Asadi, 2018; García-Aliaga et al, 2021). A técnica para gerar ratings (rótulos) na base de treinamento segue Svirydzenka (2016), com a diferença de que são utilizados os dois componentes principais mais importantes e não apenas um.

iv) Definição da base de classificação: Coletei dados de mais de 40 mil jogadores, com mais de 900 minutos jogados, de 64 ligas, em 4 janelas de tempo distintas (temporadas fechadas de 2019/20, 2020/21, 2021/22, 2022/23; ou as equivalentes 2020, 2021, 2022 e 2023 até 26.09.2023).

v) Aplicação do modelo de classificação e ranqueamento de jogadores na liga em que atuam, realizado por meio de aprendizado de máquina supervisionado, utilizando o algoritmo de regressão Random Forest (Leo Breiman, 2001; Almulla and Alam, 2020). O rating final é um número contínuo, entre 0 e 1. O produto dessa análise será o ranqueamento por liga de todos os jogadores com mais de 900 minutos na temporada.

4. Resultados

O modelo permite ordenar todos os jogadores em atividade em sua liga, segundo a posição/perfil em campo, por temporada. A partir dessas informações, os analistas podem rapidamente elaborar *shortlists* de prospectos nas diversas posições/funções. Segue exemplo com os 5 melhores meia-atacantes do Brasileirão/23, com mais de 900 minutos em campo até a rodada 24.

Tabela 3 – 5 Meias Atacantes mais bem ranqueados na Série A na temporada 2023

Atleta	Equipe	Jogos	Minutos	Rating
F. Cristaldo	Grêmio	20	1466	1
Vitor Bueno	Athletico Paranaense	19	1285	0,799
Cauly	Bahia	20	1822	0,790
G. De Arrascaeta	Flamengo	15	1058	0,777
Raphael Veiga	Palmeiras	18	1561	0,699

Em complemento ao modelo de classificação objeto deste estudo, desenvolvi uma ferramenta em *Python* que automatiza as consultas e a disponibilizei na plataforma de compartilhamento *Streamlit* e que pode ser acessada por meio do seguinte *link*:

<https://jamerico1898-softplayerz-soft-playerz-vx07dg.streamlit.app/>

5. Conclusão

Este estudo desenvolve um modelo de classificação, que utiliza aprendizado de máquina supervisionado e não supervisionado para classificar jogadores de futebol de acordo com seu desempenho e posição em campo. Visa discriminar bons e maus jogadores por meio de análise quantitativa e pode ser usado como o primeiro passo na descoberta de talentos e prospecção de jogadores, automatizando a fase inicial do processo de *scouting*. Reduz substancialmente o universo de prospectos a ser aprofundado e deve ser complementado pela igualmente importante análise do histórico de lesões, habilidades emocionais, comprometimento profissional e ajustes para diferenças entre ligas.

Referências

- Al-Asadi, M. (2018). "Decision support system for a football team management by using machine learning techniques", doi: 10.13140/RG.2.2.17433.75367.
- Almulla, J., Alam, T. (2020). "Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League", in *IEEE Access*, vol. 8, pp. 213695-213705, doi: 10.1109/ACCESS.2020.3038601.
- García-Aliaga, A., Marquina, M., Coterón, J., Rodríguez-González, A., Luengo-Sánchez, S. (2021). "In-game behaviour analysis of football players using machine learning techniques based on player statistics", in *International Journal of Sports Science & Coaching*, 16(1), 148-157. doi:10.1177/1747954120959762
- Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components." *Journal of Educational Psychology*, 24(6), 417-441.
- Leo Breiman. (2001). "Random Forests." *Machine Learning*, 45(1), 5-32.
- Pappalardo, L., Cintia, P., Rossi, A. *et al.* (2019). "A public data set of spatio-temporal match events in soccer competitions", In *Sci Data* v 6, 236. 10.1038/s41597-019-0247-7
- Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space." *Philosophical Magazine*, Series 6, Vol. 2, No. 11, pp. 559-572
- Svyrydzenka, K. (2016). "Introducing a New Broad-based Index of Financial Development", in International Monetary Fund Working Paper, WP/16/5
- Wyscout (2022). https://dataglossary.wyscout.com/player_metrics/, acessado em 10 de abril de 2022.

IMPORTÂNCIA DA AVALIAÇÃO TRIDIMENSIONAL DO CHUTE E DA CORRIDA EM ATLETA PROFISSIONAL- ESTUDO DE CASO

**Liria Akie Okai-Nóbrega¹, Lucas Valentim de Freitas¹, Vanessa Lara Araújo¹,
Diego da Silva Carvalho¹, Sabrina Penna Cintra¹, Priscila Albuquerque Araújo¹,
Renatha Carvalho¹, Aline Castro¹, Renata Frederico¹, Sergio Teixeira Fonseca¹,
Thales Rezende de Souza¹**

¹Programa de Pós-Graduação em Ciências da Reabilitação; Escola de Educação Física,
Fisioterapia e Terapia Ocupacional; Universidade Federal de Minas Gerais.

(okainobrega@gmail.com, lucasvfreitas@hotmail.com,
laraaujo.vanessa@gmail.com, diego.carvalho.fisio@gmail.com,
spenna_fisio@yahoo.com.br, pripsbr@gmail.com, renathacfisio@gmail.com,
alinecastro67@gmail.com, renataaffisioterapia@gmail.com,
sergioteixeirafonseca@gmail.com, thalesrsouza@gmail.com)

Abstract. The three-dimensional motion analysis of kicking and running in professional athletes is crucial to understand movement mechanics and optimize sports performance. 3D analysis performs a quantitative and comprehensive understanding of movement patterns, including aspects such as kinematics, kinetics, and force distribution. This approach helps identify specific body areas (local or global), for improvement, minimizes the risk of injuries, and maximizes movement efficiency and sports gesture. The use of data mining methodologies or decision trees can optimize clinical decision-making

Resumo. A avaliação tridimensional do chute e da corrida em atletas profissionais é crucial para compreender a mecânica do movimento e otimizar o desempenho esportivo. A análise 3D permite uma compreensão quantitativa e mais abrangente dos padrões de movimento, incluindo aspectos como cinemática, cinética e a distribuição de forças. Essa abordagem ajuda a identificar áreas: locais ou globais, de melhorias específicas, minimizar o risco de lesões e maximizar a eficiência do movimento e gesto esportivo. O uso de metodologias de mineração de dados ou árvores de decisão podem otimizar a tomada de decisão clínica.

1. Introdução

A avaliação tridimensional do chute e da corrida em atletas profissionais é crucial para compreender a mecânica do movimento e otimizar o desempenho esportivo [Shan et al, 2005]. A análise tridimensional permite uma compreensão quantitativa e mais abrangente dos padrões de movimento, incluindo aspectos como cinemática, cinética e a distribuição de forças durante o chute [Kellis e Katis, 2007] e a corrida. Essa abordagem detalhada e quantificável, ajuda a identificar áreas, local ou global, de melhorias específicas, minimizar o risco de lesões e maximizar a eficiência do movimento e gesto esportivo.

2. Objetivo

Descrever a biomecânica (cinemática e cinética) do chute do futebol e da corrida de um atleta profissional sub 20 de futebol e a sua relação com as queixas do atleta

3. Metodologia

Participou do estudo um atleta de futebol do sexo masculino de 19 anos, 10 anos de prática, destro para o chute, que atua na posição de meio de campo. Tem histórico de ruptura de ligamento cruzado anterior (LCA) do joelho esquerdo em 2018 com cirurgia de reconstrução, e pubalgia ocorrida há três meses. A queixa atual é de dor nos músculos posteriores da coxa e adutores do quadril esquerdo, que ocorre somente em treinos de alta intensidade. Testes clínicos também foram realizados para contribuir com a interpretação dos achados biomecânicos. A avaliação tridimensional do movimento foi realizada com o sistema de análise do movimento com um total de 17 câmeras da Qualysis, para o rastreamento de marcadores passivos nos segmentos cabeça, membros superiores, tronco, pelve, e membros inferiores, além de plataformas de força e, na corrida em esteira instrumentada (Figura 1).

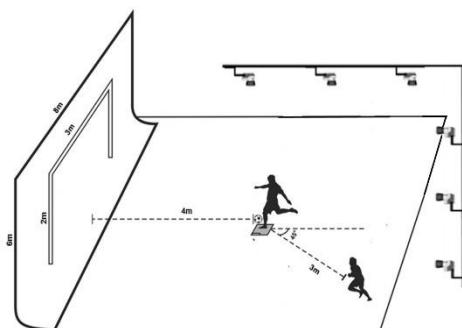


Figura 1- Esquema ilustrativo do setup de coleta do chute. Foram 9 câmeras e 3 plataformas de força no setup de coleta para chute e 8 câmeras no set up de coleta da corrida (associada a uma esteira instrumentada)

Além disso, foram extraídas as variáveis momento nos planos sagital e frontal e, potência articular no plano sagital, para as articulações do quadril, joelho e tornozelo. A posição de referência (i.e., zero grau) para cada articulação analisada foi determinada quando o sistema de coordenadas local do segmento proximal estava paralelo ao sistema de coordenadas local do segmento distal. O atleta realizou três chutes válidos com o peito do pé na sua força máxima e em seguida correu na velocidade auto selecionada e máxima (que conseguisse manter por um minuto) na esteira instrumentada. Os resultados deste estudo de caso foram comparados a média de dez atletas profissionais e com a literatura.

4. Resultados

O atleta apresentou pico de velocidade média do chute de 94,32 km/h, dentro dos parâmetros de referência para atletas profissionais. Durante todo o ciclo do chute, que vai desde o balanço posterior até a desaceleração da perna, o atleta apresenta menor exploração da amplitude de movimento da articulação lombopélvica, nos planos sagital e transverso, que parece ser compensado por maior exploração da amplitude de movimento da articulação toracolombar, nos planos sagital e transverso. Em relação aos torques articulares, há menor participação dos músculos oblíquos abdominais nas fases de

balanço posterior e preparação da perna, indicando menor contribuição desses músculos no armazenamento de energia, que ocorre nessas fases.

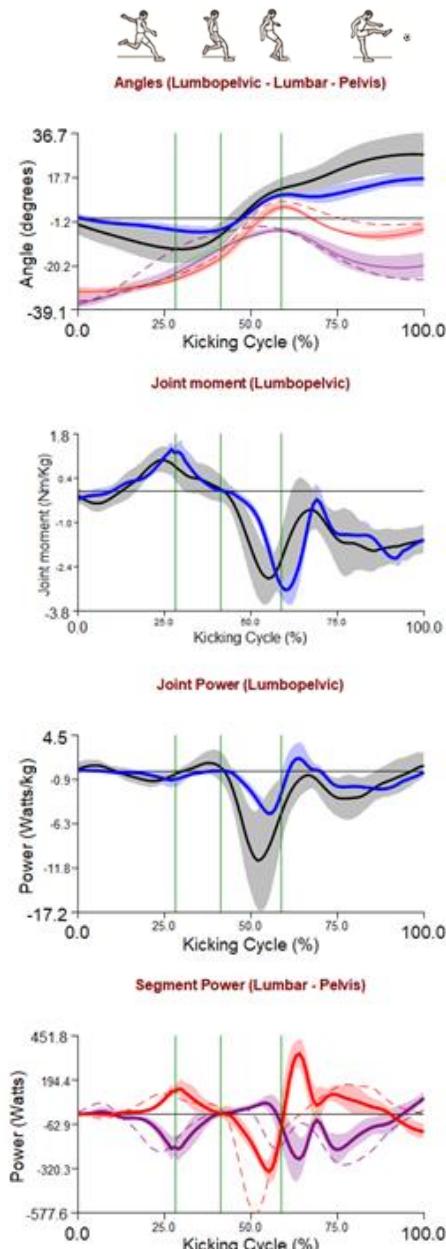


Figura 2- Curvas médias e de desvio padrão do segmento lombopélvico (flexão positivo e extensão negativo): em preto o grupo de atletas; em azul o atleta em avaliação; em roxo o segmento proximal (lombar) e em vermelho o segmento distal (pelve). O eixo horizontal representa a porcentagem do ciclo de chute (0-100%). As linhas verticais (em verde) representam respectivamente: pico de extensão do quadril, pico de flexão do joelho e contato com a bola. No gráfico inferior, a potência segmentar média do atleta (linha contínua) e do grupo de atletas (linha tracejada) para melhor visualização. (-baseado em Carvalho, 2021).

Esse padrão faz com que as forças descendentes atuando no púbis direito (do membro de chute), causadas pelos flexores e adutores do quadril direito, sejam predominantes em

relação às forças ascendentes causadas pelo oblíquo interno direito, o que sobrecarrega a sínfise púbica. (Figura 2).

Os resultados referentes ao trabalho articular revelaram menor contribuição do tronco que parece ser decorrente da menor contribuição da articulação lombo pélvica e pode ser uma possível preservação da articulação, pelo histórico de lesões. Durante a corrida, na fase de apoio, em ambas as velocidades, observou-se uma menor flexão do membro inferior esquerdo (flexão do joelho e quadril) em comparação com o direito associada a menor momento extensor do joelho esquerdo, menor absorção/geração de energia no joelho esquerdo e menor absorção de energia no quadril esquerdo. Além disso, no tornozelo, observou-se menor momento flexor plantar e menor absorção de energia do lado esquerdo. Essas alterações biomecânicas da corrida podem estar relacionadas à fraqueza de extensores de joelho, extensores de quadril e flexores plantares à esquerda confirmada nos testes clínicos. Esses achados demonstram uma redução da capacidade de absorver as forças de impacto com o solo com o membro inferior esquerdo e podem justificar o aumento do primeiro pico da força de reação de solo vertical no lado esquerdo. Esses resultados podem estar relacionados ao quadro de dor em posterior de coxa à esquerda e à história de cirurgia de reconstrução do LCA. Além disso, na corrida, em velocidade máxima, observou-se maior rotação externa do membro inferior esquerdo (quadril, joelho e tibia), com consequente menor pronação do pé esquerdo. Essas alterações biomecânicas podem estar relacionadas à maior rigidez dos rotadores externos do quadril do lado esquerdo, o que pode gerar sobrecarga nos músculos adutores do quadril esquerdo e o histórico de pubalgia.

5. Conclusão

Os resultados mostram a importância da avaliação tridimensional do movimento juntamente com a relação de testes clínicos para o entendimento das sobrecargas no sistema musculoesquelético. Portanto, a avaliação biomecânica da corrida e do chute no futebol com interpretação dos achados por meio de raciocínio clínico pode ser útil na tomada de decisão para tratamento e prevenção de lesões de atletas do futebol. Novas pesquisas considerando uma amostra maior devem ser realizadas para demonstrar a viabilidade em grande escala deste estudo além do uso de metodologias de mineração de dados ou árvores de decisão pode otimizar a tomada de decisão clínica.

6. Referências Bibliográficas

- SHAN, Gongbing; WESTERHOFF, Peter. (2005) Full-body kinematic characteristics of the maximal instep soccer kick by male soccer players and parameters related to kick quality. *Sports Biomechanics*, [S.L.], v. 4, n. 1, p. 59-72, jan . Informa UK Limited. <http://dx.doi.org/10.1080/14763140508522852>.
- Kellis E, Katis A. Biomechanical characteristics and determinants of instep soccer kick (2007). *J Sports Sci Med*. Jun 1;6(2):154-65. PMID: 24149324; PMCID: PMC3786235.
- CARVALHO, Diego da Silva; OCARINO, Juliana Melo; CRUZ, Aline de Castro; BARSANTE, Leonardo Drumond; TEIXEIRA, Breno Gonçalves; RESENDE, Renan Alves; FONSECA, Sérgio Teixeira; SOUZA, Thales Rezende. (2021) The trunk is exploited for energy transfers of maximal instep soccer kick: a power flow study. *Journal Of Biomechanics*, [S.L.], v. 121, p. 110425, maio Elsevier BV. <http://dx.doi.org/10.1016/j.jbiomech.2021.110425>.
- Neumann, D. A. (2009). *Kinesiology of the Musculoskeletal System: Foundation for Rehabilitation*. New York, USA: Mosby.

AS ESTRATÉGIAS BIOMECÂNICAS DE TRONCO E MEMBRO INFERIOR PARA O CHUTE MÁXIMO DO FUTEBOL SÃO COMPLEMENTARES?

Lucas Valentim de Freitas^{1,2}, Diego da Silva Carvalho^{1,2}, Sabrina Penna Cintra^{1,2}, Gabriela Menezes Blaso³, Brenda Vilas Boas Gomes³, Pedro Henrique de Carvalho Mello³, Sérgio Teixeira da Fonseca¹⁻³, Juliana de Melo Ocarino¹⁻³, Priscila Albuquerque de Araújo^{1,2}, Liria Akie Okai Albuquerque Nobrega^{1,2}, Thales Rezende de Souza¹⁻³

¹Programa de Pós-Graduação em Ciência da Reabilitação, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil

²Laboratório de Análise do Movimento, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil

³Departamento de Fisioterapia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil

Introdução

O chute com o dorso do pé é a principal ação ofensiva no futebol. É um movimento complexo, repetitivo e que envolve transmissão descendente de energia cinética entre segmentos. Por isso, o entendimento dos mecanismos que envolvem o chute é de interesse de profissionais do esporte e pesquisadores, visando melhora de performance e compreensão de possíveis mecanismos de lesão. Trabalhos recentes têm destacado a contribuição dos segmentos superiores do corpo no desempenho do chute. Considerando aspectos cinemáticos, jogadores experientes, comparados a novatos, apresentam maior amplitude de movimento dos segmentos superiores do corpo, com maior extensão e rotação axial do tronco antes do impacto com a bola, que resulta em um melhor desempenho do chute máximo (i.e., maior velocidade da bola). Considerando aspectos cinéticos, os torques articulares do tronco têm um papel importante no fluxo descendente de energia cinética para o membro inferior (MI) de chute. Torques flexores do tronco contribuem para o deslocamento anterior do MI de chute, assim como torques rotadores de tronco contribuem para a rápida extensão de joelho desse mesmo membro. Dessa forma, as estratégias cinemática e cinética do tronco devem ser consideradas para se compreender a realização do chute do futebol, assim como as estratégias do membro inferior tradicionalmente consideradas. As estratégias biomecânicas (cinemáticas e cinéticas) do tronco e membro inferior podem ser complementares, uma vez que ambos contribuem para o deslocamento anterior do membro de chute. Entretanto, ainda não se sabe se existem associações entre as estratégias biomecânicas do tronco e do MI durante o chute do futebol.

Objetivo

Investigar associações entre as estratégias cinemática e cinética do tronco e do MI para a execução do chute máximo do futebol.

Metodologia

A cinemática e cinética articulares foram medidas em 50 jogadores amadores de futebol (idade: $23,7 \pm 4,61$ anos; altura: $1,77 \pm 0,05$ m; massa corporal: $75,40 \pm$

9,88 kg; IMC: $24,17 \pm 2,86$ kg/m²). Todos os participantes eram considerados habilidosos no chute, com 45 participantes (90%) com > 5 anos, três (6%) com 2 a 5 anos e dois (4%) com 1 a 2 anos de tempo de prática, frequência média de treino de 1.94 ± 1.22 vezes/semana e velocidade média da bola de 20.0 ± 2.29 m/s. A Figura 1 apresenta a configuração do laboratório e descreve aspectos da medição. Os participantes eram instruídos a chutar a bola com o dorso do pé, com o máximo de força possível e na direção de um alvo retangular desenhado sobre a lona. Os participantes executaram até que quatro chutes fossem considerados válidos para análise. Para evitar fadiga, um limite de 15 chutes foi definido.

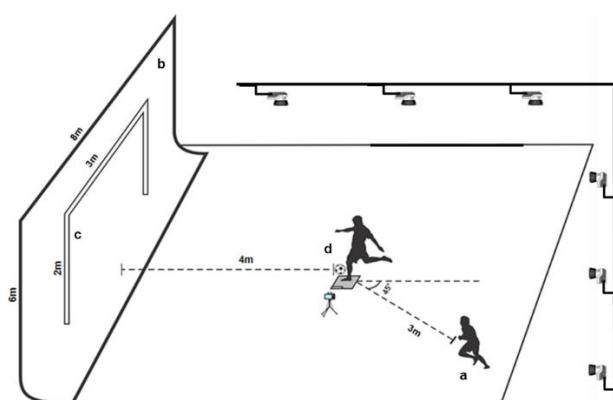


Figura 1. Visão geral da configuração do laboratório.a) posição inicial do participante, cerca de três metros de distância da bola, a um ângulo de 45º em relação à linha de direção do chute; b) lona plástica (tamanho: 8 m x 6 m), fixada no teto do laboratório, a quatro metros de distância da bola, para absorver o impacto da bola; c) alvo retangular desenhado na tela nas medidas oficiais de um gol de futsal (altura: 2 m; largura: 3 m; espessura: 8 cm); d) posicionamento da bola de futebol na lateral da plataforma de força (retângulo cinza representado na figura) de forma que o pé de apoio (MI esquerdo) pouse na plataforma. Adaptado de Carvalho *et al.* (2021).

Para as estratégias biomecânicas do tronco, foram computados picos de extensão e rotação à esquerda (estratégia cinemática) e de torques flexor e rotador à direita (estratégia cinética). Para as estratégias biomecânicas do MI, foram computados os picos de extensão do quadril e flexão do joelho (estratégia cinemática) e de torques flexor do quadril e extensor do joelho (estratégia cinética). A Análise de Correlação Canônica (ACC) foi utilizada para investigar e medir associações entre um grupo de variáveis independentes (variáveis do tronco) e um grupo de variáveis dependentes (variáveis do MI), em que combinações dessas variáveis representam estratégias do tronco e do MI, respectivamente. Além disso a análise permite identificar mais de uma associação entre as estratégias. Uma correlação canônica foi realizada para investigar associações entre as estratégias cinemáticas de tronco e MI. Uma segunda correlação canônica foi realizada para investigar associações entre as estratégias cinéticas de tronco e MI. A Figura 2 apresenta um fluxograma das etapas sugeridas para interpretação da ACC.

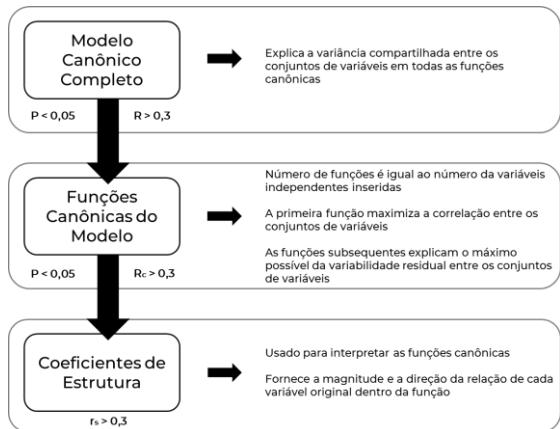


Figura 2. Fluxograma das etapas para interpretação da Análise de Correlação Canônica. Adaptado de Parrington *et al.* (2015).

Resultados

Os dados descritivos são apresentados na Tabela 1.

Tabela 1. Descrição das variáveis originais cinemáticas e cinéticas

Variáveis	Média ± DP	Mín.	Máx.	IC95%	CV (%)
Pico de Extensão do Tronco (°)	25,37 ± 7,50	- 2,06	40,35	23,24 – 27,51	24,8
Pico de Rotação do Tronco (°)	26,28 ± 7,47	9,71	41,13	24,16 – 28,41	28,4
Pico de Extensão do Quadril (°)	22,37 ± 8,60	3,81	45,04	19,93 – 24,82	38,5
Pico de Flexão do Joelho (°)	109,00 ± 8,15	91,53	123,31	106,68 - 111,32	7,5
Pico de Torque Flexor do Tronco (N.m/kg)	2,04 ± 0,77	0,44	4,10	1,82 – 2,26	38,0
Pico de Torque Rotador do Tronco (N.m/kg)	2,99 ± 0,55	1,92	4,56	2,83 – 3,14	18,4
Pico de Torque Flexor do Quadril (N.m/kg)	3,91 ± 0,65	2,74	5,60	3,73 – 4,09	16,6
Pico de Torque Extensor do Joelho (N.m/kg)	1,48 ± 0,37	0,74	2,37	1,37 – 1,58	25,2

Notas: Desvio Padrão (DP); Mínimo (Min); Máximo (Max); Intervalo de Confiança de 95% (IC95%); Coeficiente de Variação (CV).

(°): graus;

(N.m/kg): Newton metro por quilograma

As funções canônicas identificadas para as estratégias cinemáticas não foram significativas ($p > 0,05$), indicando que não há associação entre as estratégias cinemáticas de tronco e MI (Tabela 2). As funções canônicas identificadas para as estratégias cinéticas do tronco e MI foram significativas ($p \leq 0,01$) e interpretáveis ($R_c^2 \geq 0,13$). O modelo canônico completo explicou 59% da variância compartilhada entre os conjuntos de variáveis.

Tabela 2. Correlações canônicas entre as estratégias biomecânicas de tronco e MI de chute

Estratégias Cinemáticas					
	R _c	R _c ²	Wilks	F	p
1	,240	,057	,925	0,919	,457
2	,138	,019	,981	0,907	,346

Estratégias Cinéticas					
	R _c	R _c ²	Wilks	F	P
1	,730	,533	,405	13,124	<,001
2	,364	,132	,868	7,169	,010

R_c: Correlação canônica; R_c²: Coeficiente de determinação canônica; Wilks: Estatística de Wilks Lambda (teste de significância do R_c); F: Estatística F (teste de variância); p: significância estatística entre as variáveis canônicas.

Uma primeira associação ($R_c=0.730$, $p<0,001$), considerada excelente, mostrou que quanto maiores os picos de torques de flexão e rotação do tronco, maiores os picos dos torques flexor do quadril e extensor do joelho (Figura 3). Esse padrão foi encontrado em 39 (78%) participantes. Uma segunda relação ($R_c=0.364$, $p=0,010$), mais fraca e presente em 11 (22%) participantes, é caracterizada por menor pico de torque de flexão do tronco associado a maior pico de torque de extensão do joelho e menor pico de torque de flexão do quadril (e vice-versa).

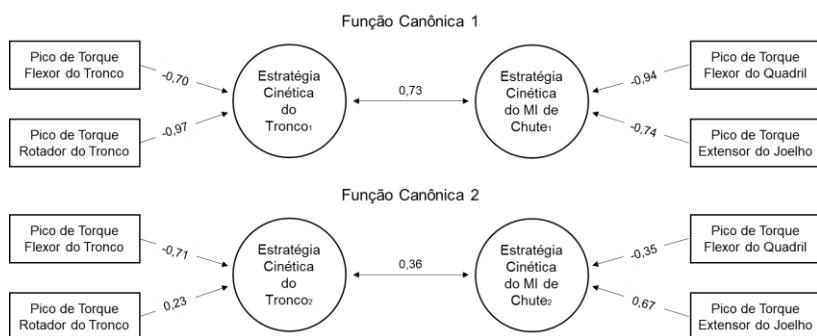


Figura 3. Modelo de correlação canônica entre as estratégias cinéticas do tronco e MI de chute

Conclusão

Os resultados deste estudo indicam que as estratégias cinemáticas de tronco e MI não são complementares na população estudada. Por outro lado, as estratégias cinéticas de tronco e MI, para a realização do chute máximo, são complementares e podem apresentar dois padrões de complementaridade. A relação mais forte e presente na maior parte dos participantes é caracterizada por maiores torques de flexão e rotação do tronco associados a maiores torques de flexão do quadril e extensão do joelho, que parece ser usada para potencializar o chute em indivíduos experientes. Uma segunda relação, mais fraca e presente em um número menor de participantes, é caracterizada por menores torques de flexão do tronco e do quadril acompanhados de maiores torques de extensão do joelho (e vice-versa), que parece ser uma relação compensatória para lidar com variações dos torques articulares. Assim, as estratégias cinéticas do tronco, compreendidas a partir dos torques articulares, são complementares às estratégias cinéticas do MI; porém, em dois padrões diferentes. Os padrões encontrados apontam para a existência de diferentes formas para realizar o chute máximo com o dorso do pé no futebol. Estratégias complementares com potencial para contribuir com o maior desempenho do chute e estratégias complementares compensatórias com potencial de gerar sobrecarga a estruturas.

Referências

- AUGUSTUS, S.; HUDSON, P. E.; HARVEY, N.; SMITH, N. Whole-body energy transfer strategies during football instep kicking: implications for training practices. **Sports Biomechanics**, 2021.
- CARVALHO, D. D. S.; OCARINO, J. M.; CRUZ, A. D. C.; BARSANTE, L. D. et al. The trunk is exploited for energy transfers of maximal instep soccer kick: A power flow study. **Journal of Biomechanics**, 121, p. 110425-110425, 2021/5// 2021.
- FONSECA, S. T.; SOUZA, T. R.; OCARINO, J. D. M.; GONÇALVES, G. P. et al. Applied biomechanics of soccer. In: **Athletic and Sports Issues in Musculoskeletal Rehabilitation**: Saint Louis: Saunders Elservier, 2011. p. 315-329.
- FULLENKAMP, A. M.; CAMPBELL, B. M.; LAURENT, C. M.; LANE, A. P. The Contribution of Trunk Axial Kinematics to Poststrike Ball Velocity During Maximal Instep Soccer Kicking. **Journal of Applied Biomechanics**, 31, n. 5, p. 370-376, 2015/10// 2015.
- KELLIS, E.; KATIS, A. Biomechanical characteristics and determinants of instep soccer kick. **Journal of sports science & medicine**, 6, n. 2, p. 154-165, 2007.
- LEES, A.; NOLAN, L. The biomechanics of soccer: A review. **Journal of Sports Sciences**, 16, n. 3, p. 211-234, 1998/1// 1998.
- NAITO, K.; FUKUI, Y.; MARUYAMA, T. Multijoint kinetic chain analysis of knee extension during the soccer instep kick. **Human Movement Science**, 29, n. 2, p. 259-276, 2010.
- PARRINGTON, L.; BALL, K.; MACMAHON, C. Kinematics of a striking task: accuracy and speed-accuracy considerations. **Journal of Sports Sciences**, 33, n. 4, p. 346-357, 2015/2// 2015.
- SHAN, G.; WESTERHOFF, P. Soccer: Full-body kinematic characteristics of the maximal instep Soccer kick by male soccer players and parameters related to kick quality. **Sports Biomechanics**, 4, n. 1, p. 59-72, 2005.
- SHERRY, A.; HENSON, R. K. Conducting and Interpreting Canonical Correlation Analysis in Personality Research: A User-Friendly Primer. **Journal of Personality Assessment**, 84, n. 1, p. 37-48, 2005/2// 2005.
- TABACHNICK, B. G.; FIDELL, L. S. Using Multivariate Statistics. **Journal of Applied Psychology**, 87, n. 4, p. 611-628, 2017.

Efeito do Treinamento Resistido de Músculos do Tronco e do Membro Superior na Cinemática do Chute do Futebol

Miguel Arcanjo de Assis¹, Thiago Ribeiro Teles Santos², Sergio Teixeira Fonseca¹, André Gustavo Pereira de Andrade³, Priscila Albuquerque Araújo¹, Liria Akie Okai-Nóbrega¹, Thales Rezende de Souza¹, Renan Alves Resende¹, Juliana Melo Ocarino¹

¹ Programa de Pós-Graduação em Ciências da Reabilitação; Escola de Educação Física, Fisioterapia e Terapia Ocupacional; Universidade Federal de Minas Gerais.

² Programa de Pós-Graduação em Fisioterapia; Faculdade de Educação Física e Fisioterapia; Universidade Federal de Uberlândia.

³ Programa de Pós-Graduação em Ciências do Esporte; Escola de Educação Física, Fisioterapia e Terapia Ocupacional; Universidade Federal de Minas Gerais.

julianaocarino@gmail.com

1. Introdução

O chute é um dos gestos mais característicos do futebol. A maior parte da literatura relacionada a esse gesto esportivo já demonstrou a contribuição dos segmentos do membro inferior (e.g., coxa) para o desempenho do chute. Esses achados oferecem suporte para que o treino de futebol incorpore exercícios resistidos de músculos do membro inferior (e.g., quadríceps). Além desses segmentos, os movimentos de tronco e membro superior são essenciais para a realização do chute. Por exemplo, estudos já indicaram que jogadores mais habilidosos apresentam maior extensão e rotação de tronco que jogadores menos experientes. Outro estudo já indicou que a energia é transferida do tronco para a pelve nas fases iniciais do chute e, em seguida, para o membro inferior que irá chutar a bola. Essa energia transferida do tronco para a pelve é relacionada à velocidade do pé ao realizar o contato com a bola. Esses achados corroboram o entendimento que o sistema musculoesquelético integra múltiplos segmentos corporais por meio de conexões miofasciais entre componentes próximos e distantes. Nesse sentido, o fortalecimento de músculos que compõem conexões miofasciais do membro superior e tronco poderia modificar o padrão cinemático de chute, otimizando os movimentos de tronco e quadril.

2. Objetivo

Examinar os efeitos do exercício resistido de músculos do membro superior e tronco anterior na cinemática do chute com o dorso do pé do futebol.

3. Metodologia

Vinte e seis participantes do sexo masculino foram aleatoriamente alocados em dois grupos: 13 no grupo de treinamento (GT) e 13 no grupo controle (GC). Os critérios de inclusão foram idade entre 18 e 30 anos, praticar futebol no máximo 1x por semana e apresentar índice de massa corporal $<30 \text{ kg/m}^2$. Todos os participantes não poderiam ter realizado programa de fortalecimento ou apresentar dor ou história de lesão nos últimos seis meses. Os participantes que não conseguiram realizar o chute com o dorso do pé e

que não participaram de no mínimo 80% das sessões de treinamento foram excluídos. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais (ETIC 0676.0.203.000-11).

3.1. Análise do Movimento

A cinemática tridimensional do tronco, pelve e segmentos do membro inferior dominante durante o chute com o dorso do pé foi capturada pelo sistema a 100 Hz. Após cinco chutes submáximos utilizados para familiarização, foram registrados pelo menos cinco chutes máximos válidos. Os chutes foram classificados como válidos se a bola atingisse um alvo perpendicular (2,5 x 4m), localizado 5m a frente de onde estava a bola de futebol. Os dados cinemáticos foram filtrados com filtro passa-baixa Butterworth de quarta ordem com frequência de corte de 10 Hz. Os dados de joelho e quadril no plano sagital, assim como do tronco no plano sagital e transverso foram extraídos e analisados do início da fase de *backswing* (i.e., retirada do pé do membro inferior de chute) até o final da fase *leg acceleration* (i.e., contato do pé com a bola). A série temporal foi normalizada em 101 pontos (0% a 100%).

3.2. Programa de treinamento

O GT realizou exercício resistido supervisionado e progressivo com objetivo de hipertrofiar os seguintes músculos: flexores e extensores do cotovelo, peitoral maior, oblíquos abdominais e reto abdominal. Esses músculos foram escolhidos uma vez que participam de linhas de transmissão de força. O treinamento ocorreu 3x por semana durante oito semanas, sendo que nas primeiras quatro semanas, foram realizados exercícios resistidos isolados para cada grupo muscular. Na quinta semana, foi adicionado um exercício que recrutava músculos do braço e tronco simultaneamente. Na sétima semana, foi adicionado um exercício que recrutava músculos do braço e tronco, assim como simulava o chute. Além disso, o GT praticou o chute com o dorso do pé em uma quadra, 3x por semana durante as oito semanas. O GC realizou somente essa prática de chute por oito semanas.

3.3. Análise estatística

Statistical Parametric Mapping (SPM) t independente comparou as curvas angulares de quadril entre grupos antes e após a intervenção. SPM t dependente comparou as diferenças em quadril antes e após a intervenção em cada grupo. A versão não paramétrica desses testes foi utilizada para os dados de joelho e tronco, uma vez que o teste D'Agostino-Pearson K^2 revelou uma distribuição não normal desses dados. O tamanho de efeito Hedge g foi calculado. O nível de significância foi de 0,05 para todas as análises.

4. Resultados

Uma diferença significativa foi revelada para a cinemática de quadril no GT entre 14% a 25% do ciclo de chute ($g = 0,316-0,321$). Nesse intervalo, o GT apresentou maior extensão média de quadril após o programa de treinamento que na avaliação pré-intervenção. Não foram identificadas outras diferenças.

5. Conclusões

O programa de intervenção aumentou a extensão do quadril em um intervalo do chute que corresponde à fase de *backswing*. Ou seja, os participantes adotaram um

padrão que aumentava o deslocamento do membro inferior de chute para trás. Essa modificação de padrão pode refletir uma otimização do fluxo de energia mecânica entre membro superior/tronco com o membro inferior. Além disso, outros estudos já observaram maior extensão de quadril durante o *backswing* em atletas experientes comparado aos não experientes. Nessa perspectiva, os resultados indicam que o treinamento de músculos de segmentos não locais (i.e., segmentos corporais que não os de membro inferior) podem contribuir com o chute com o dorso do pé e tem o potencial de modificar a cinemática do quadril. Assim, o programa de treinamento pode ter resultado em maior efetividade no recrutamento muscular e na transferência de tensão entre os componentes miofasciais contínuos do membro superior e tronco.

Este estudo foi publicado no *Journal of Applied Biomechanics*: Assis MA, Santos TRT, Fonseca ST, Andrade AGP, Araújo PA, Souza TR, Resende RA, Ocarino JM. Effects of Resistance Training of Upper Limb and Trunk Muscles on Soccer Instep Kick Kinematics. J Appl Biomech. 2023 Jul 24;39(4):254-263.

Estudo sobre sprints em jogadores da Série B: Momentos e características

Pedro Monteiro¹, Lucas Vinicius¹, Filipe Lourenço¹ e Pedro Meneses¹

¹Ceará Sporting Club (CSC) – Caixa Postal 60425-813 – Fortaleza – CE – Brasil

fisiologia@cearasc.com.br

Abstract. The aim of this study was to examine sprints at speeds above 30 km/h and their relationship with match outcomes. The analysis involved comparing the number and distance of sprints between halves of the game and correlating sprint trajectories with wins, draws, or losses. Paired samples t-tests and non-parametric ANOVA were used, which did not reveal statistically significant differences between halves of the game or in relation to the match outcomes.

Resumo. O objetivo deste estudo foi examinar sprints em velocidades acima de 30 km/h e sua relação com os resultados das partidas. A análise envolveu comparação da quantidade e distância dos sprints entre os tempos de jogo e a correlação das trajetórias dos sprints com vitórias, empates ou derrotas. Utilizou-se teste t para amostras emparelhadas e ANOVA não-paramétrica, que não revelaram diferenças estatisticamente significativas entre os tempos de jogo ou em relação ao resultado das partidas.

1. Introdução

O futebol evoluiu ao longo das décadas em aspectos táticos, técnicos e físicos. A integração da tecnologia, especificamente o uso de GPS, tem permitido um entendimento mais profundo das exigências físicas dos atletas em campo, evidenciando a importância de velocidades e intensidades elevadas.

A distância total percorrida pelos atletas diminuiu ligeiramente, mas houve um aumento considerável nas demandas de intensidade (Barnes et al., 2014). A distância percorrida em velocidades entre 19,8 km/h e 25,2 km/h aumentou 29%, enquanto a distância em velocidades acima de 25,2 km/h aumentou 50%. Essas ações representam de 7% a 11% e de 1% a 3% do total percorrido durante uma partida, respectivamente (Lago-Peñas, 2022).

A literatura atual apresenta uma lacuna no consenso sobre as zonas de velocidade no futebol (Gualtieri, 2023). Geralmente, velocidades acima de 19,8 km/h são classificadas como alta intensidade e acima de 25,2 km/h como sprints (Gualtieri, 2020). Estudos recentes estão se concentrando em sprints acima de 30 km/h, dada a sua intensidade física extrema (Oliva-Lozano et al., 2023). Estudar sprints nesta zona de velocidade oferece uma perspectiva sobre os momentos de pico de performance

2. Objetivos

Este estudo tem como objetivo examinar ações de sprint em velocidades acima 30 km/h, identificando seus padrões e características. Busca-se quantificar a frequência e a distância desses sprints ao longo dos dois tempos de jogo, assim como analisar as trajetórias dos sprints, avaliando sua direção (para frente ou para trás) e a sua correlação com os resultados das partidas.

3. Metodologia

Para realização deste trabalho foram utilizados dados secundários fornecidos pelo departamento de fisiologia do exercício de um clube da segunda divisão do campeonato Brasileiro de futebol. Foram processados dados de GPS de atletas coletados em jogos oficiais na temporada 2023.

Durante cada sessão de treino ou jogo, os atletas foram monitorados por meio do sistema de GPS (Catapult® Vector S7) inseridos em um colete entre as escápulas dos atletas. Os dados de velocidade foram coletados a cada 0,1 segundo, oferecendo precisão na captura de movimentos.

Analisou-se um recorte de 11 partidas do mesmo treinador no Campeonato Brasileiro da Série B, mantendo assim o mesmo modelo de jogo. Desses jogos, a equipe obteve vitória em quatro, sofreu derrotas em outras quatro ocasiões, e registrou três empates.

Os dados foram exportados em formato CSV e analisados com Power BI para contar a frequência e a distância dos sprints ao longo dos dois tempos de jogo. Os dados brutos foram analisados com Python para verificar a trajetória dos sprints (para frente ou para trás), no qual foram examinados 181 eventos de vídeo, dos quais 22 foram descartados por estarem duplicados.

O software estatístico Jamovi foi utilizado para análise descritiva e inferencial dos dados. A normalidade dos dados foi verificada pelo teste de Shapiro-Wilk. O teste t para amostras emparelhadas foi aplicado para comparar o número de ações em sprint entre os dois tempos. Para investigar diferenças nas médias das distâncias de sprints, segmentadas por resultado do jogo (vitória, derrota, empate), utilizou-se a ANOVA a um fator não-paramétrico.

4. Resultados e Discussão

Na comparação das ações de sprint nos jogos (tabela 1), verificou-se uma média de 6,91 sprints no primeiro tempo e 5,91 no segundo, indicando uma diminuição na frequência dos sprints conforme o jogo avança. A variabilidade das ações se mostrou similar entre ambos os tempos, com uma leve elevação no primeiro. No entanto, o teste t para amostras emparelhadas ($t(10) = 0.526$, $p = 0.611$) não demonstrou diferenças estatisticamente significativas na média de sprints entre os dois tempos de jogo.

Tabela 1. Frequência de sprints acima de 30 km/h por partida.

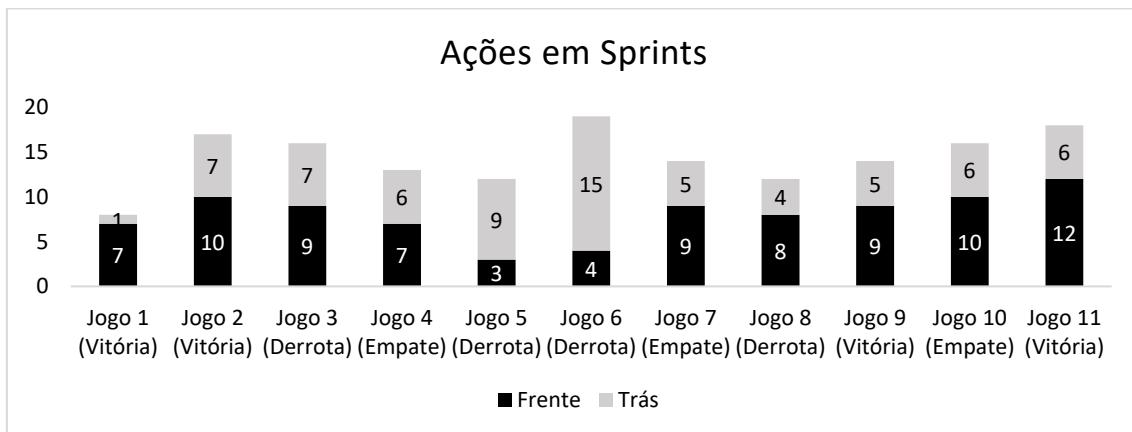
Jogo	1º Tempo	2º Tempo	Total	Sprints por posição					
				ZAG	LAT	VOL	MC	ATA	CA
Jogo 1	5	3	8	2	0	4	2	0	0
Jogo 2	9	6	15	6	0	7	0	0	2
Jogo 3	3	12	15	4	5	4	0	0	2
Jogo 4	5	6	11	3	3	3	0	0	2
Jogo 5	4	3	7	1	0	1	0	1	4
Jogo 6	16	3	19	6	1	6	0	1	5
Jogo 7	9	4	13	4	4	2	3	0	0
Jogo 8	4	6	10	5	1	2	2	0	0
Jogo 9	7	4	11	7	2	2	0	0	0
Jogo 10	10	5	15	4	3	2	2	2	2
Jogo 11	4	13	17	4	2	6	1	1	3

ZAG = Zagueiro; LAT = Lateral; VOL = Volante; MC = Meio-campo; ATA = Atacante; CA = Centroavante.

Em relação às distâncias em sprint, observou-se uma média de 130 metros no primeiro tempo e 122 metros no segundo, indicando uma possível redução na distância de sprint conforme o jogo progride. Apesar de uma maior dispersão no segundo tempo, o teste t para amostras emparelhadas ($t(10) = 0.168$, $p = 0.870$) não mostrou diferença estatisticamente significativa nas distâncias percorridas entre os dois tempos.

Analisou-se 159 eventos de sprint com o auxílio de algoritmos previamente descritos, cujos resultados podem ser visualizados no gráfico 1.

Gráfico 1. Direção das ações em sprints acima de 30 km/h.



A média de sprints para frente foi maior em jogos vencidos e menor em jogos perdidos, sugerindo um potencial relação entre a agressividade ofensiva e os resultados positivos. Mas essa tendência não é estatisticamente significativa, o teste de Kruskal-Wallis indicou que não existem diferenças estatisticamente significativas na distribuição dos sprints para frente ($\chi^2 = 4.710$, $p = 0.095$) ou para trás ($\chi^2 = 0.341$, $p = 0.843$) entre vitórias, derrotas e empates. Este achado sugere que, com base nos dados disponíveis, a frequência dos sprints em ambas as direções não está associada de maneira conclusiva aos resultados dos jogos.

5. Conclusão

Concluímos que não há diferença estatisticamente significativa nas ações e distâncias de sprints entre o primeiro e o segundo tempo, e embora haja uma tendência de maiores sprints para frente em vitórias e menos em derrotas, essa diferença não é significativa. A análise dos vídeos sugere que os sprints mais intensos ocorrem tipicamente durante as transições do jogo. Esses picos de alta intensidade são frequentemente alinhados com momentos cruciais, como contra-ataques e esforços defensivos imediatos. Este entendimento sublinha a relevância dos sprints em momentos chave, e novas pesquisas são necessárias para explorar o contexto tático por trás desses padrões.

6. Referências

BARNES, Chris et al. The evolution of physical and technical performance parameters in the English Premier League. **International journal of sports medicine**, p. 1095-1100, 2014.

GUALTIERI, Antonio et al. Workload monitoring in top-level soccer players during congested fixture periods. **International journal of sports medicine**, v. 41, n. 10, p. 677-681, 2020.

GUALTIERI, Antonio et al. High-speed running and sprinting in professional adult soccer: current thresholds definition, match demands and training strategies. A systematic review. **Frontiers in Sports and Active Living**, v. 5, p. 1116293, 2023.

LAGO-PENAS, Carlos et al. Evolution of physical and technical parameters in the Spanish LaLiga 2012-2019. **Science and Medicine in Football**, v. 7, n. 1, p. 41-46, 2023.

OLIVA-LOZANO, José M.; FORTES, Víctor; MUYOR, José M. When and how do elite soccer players sprint in match play? A longitudinal study in a professional soccer league. **Research in Sports Medicine**, v. 31, n. 1, p. 1-12, 2023.

Utilizando o OpenPose para análise cinemática tridimensional do salto de goleiros de futebol no campo: efeito da lateralidade no pênalti

Rafael Luiz Martins Monteiro^{1,2}, Paulo Roberto Pereira Santiago^{1,2}

¹Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo (USP) – São Paulo – SP – Brasil

² Escola de Educação Física e Esporte de Ribeirão Preto, Universidade de São Paulo (USP) – São Paulo – SP – Brasil

1. Introdução

Compreender os aspectos cinemáticos do salto do goleiro no futebol é importante para o desenvolvimento e treinamento dos atletas. A análise da velocidade do salto e do padrão de movimento do goleiro permite descrever quantitativamente o salto e, com isso, realizar a avaliação do desenvolvimento do atleta. A identificação do efeito da lateralidade no salto pode evidenciar assimetrias, bem como aspectos limitantes ou virtudes individuais de desempenho. O desenvolvimento de técnicas de visão computacional possibilitou a identificação automática de pontos anatômicos em vídeos. O uso destas técnicas para cinemática do salto de goleiros de futebol representa uma inovação ao permitir que análises mais ecológicas sejam realizadas diretamente em campo, diferentemente dos demais estudos da área que foram realizados em laboratório (IBRAHIM et al., 2019a, 2019b, 2020a, 2020b; MONTEIRO et al., 2022; SPRATFORD; MELLIFONT; BURKETT, 2009).

2. Objetivo

Analisar, dentro de campo, o efeito da lateralidade na cinemática do salto de jovens goleiros de futebol no pênalti.

3. Material e Métodos

O estudo contou com a participação de oito goleiros da categoria de base de um clube de futebol da segunda divisão estadual de São Paulo com médias de idade de $16,64 \pm 2,06$ anos, massa de $80,2 \pm 10,99$ kg, estatura de $1,83 \pm 0,1$ m, experiência atuando na posição de $6,25 \pm 3,45$ anos e frequência de treino de $4,25 \pm 1,39$ dias/semana. Cada goleiro realizou 20 tentativas de defesa de pênalti no campo, com bolas lançadas por uma máquina. Os atletas executaram o mesmo número de saltos para cada lado. As bolas foram lançadas em direção aos cantos do gol, sempre buscando atingir a metade central da trave. O lado que a bola seria lançada era desconhecido para os atletas. A lateralidade dos membros inferiores foi determinada pelo Inventário de Preferência Lateral Global (IPLAG) (MARIM et al., 2011). As tentativas de defesa de pênalti foram gravadas por três câmeras (GoPro HERO 3+ Black Edition, operando em 120 Hz) (Figura 1).

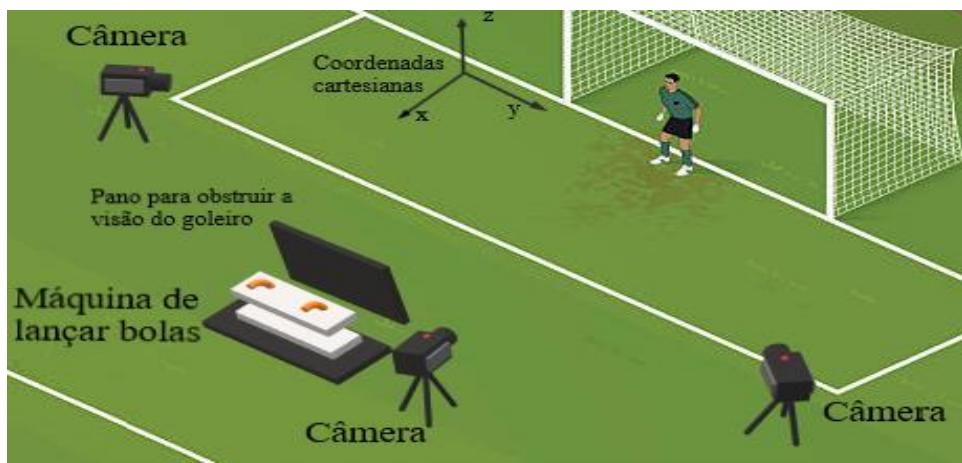


Figura 1. Setup experimental da coleta de dados

As identificações dos centros articulares nos vídeos foram obtidas por meio do algoritmo de pose humana baseado em redes neurais convolucionais (OpenPose) (CAO et al., 2019), para posterior reconstrução utilizando o método o método 3D-DLT (*direct linear transformation*). Quando necessário, as coordenadas 2D obtidas pelo OpenPose foram corrigidas manualmente no software Dvideo (v. 1.0.0.1) (FIGUEROA; LEITE; BARROS, 2003). Os dados foram suavizados com o filtro digital *Butterworth* de 4^a ordem com frequência de corte obtida pela análise de resíduos e ajustada em 7 Hz. A reconstrução 3D foi realizada com os vídeos de somente duas câmeras. Para a calibração foram utilizadas as dimensões padrão do gol, e os erros para os eixos foram: X (anteroposterior) = 5,4 cm, Y (médio-lateral) = 1,5 cm e Z (vertical) = 1,7 cm. Para análise de dados o centro de massa corporal (CM) dos goleiros foi calculado, e os saltos foram normalizados temporalmente (0 - 100%), com início na retirada do pé ipsilateral ao lado do salto do solo e fim no momento em que o CM atinge a velocidade pico. Do total de 154 saltos, 30 foram descartados, restando 124 saltos validados. As tentativas foram descartadas devido ao local de lançamento da bola não ter sido ideal para que os goleiros realizassem os saltos corretamente. Duas situações de descarte prevaleceram: (1) quando a bola foi lançada muito perto do centro do gol; (2) quando a bola foi lançada muito alta, o que modifica o padrão de movimento do goleiro. As variáveis analisadas foram: curvas do ângulo de flexão/extensão de joelho ipsilateral ao salto, velocidade do CM, tempo para atingir a velocidade pico, distância do passo frontal para impulsão, ângulo de saída frontal e distância entre as pernas na postura de preparação para o salto. Para análise estatística foram realizados o teste t de *Student* para amostras pareadas comum e utilizando o *Statistical Parametric Mapping* para comparação das curvas de velocidade e ângulo do joelho. Em todos os casos o nível de significância foi de $p < 0,05$. O d de Cohen foi utilizado para medir o tamanho do efeito obtido. Todos os procedimentos foram realizados com a linguagem de programação Python 3.

4. Resultados

Seis goleiros foram classificados com preferência lateral do membro inferior direita e dois esquerda. Foram encontrados maiores valores nos saltos para o lado da perna não dominante nas variáveis: velocidade do centro de massa (88,4 – 100% do ciclo de impulsão; $p = 0,018$), curva do ângulo de flexão/extensão de joelho ipsilateral ao salto (41,3 – 62,6% do ciclo de impulsão; $p = 0,005$) (Figura 2) e distância do passo frontal

$(0,343 \pm 0,18 / 0,439 \pm 0,24 \text{ m}; p = 0,009; d = 0,45)$ (Tabela 1). Não foram encontradas diferenças entre as variáveis tempo para velocidade resultante pico, ângulo de saída frontal e distância entre as pernas na postura de preparação para o salto.

Efeito da lateralidade

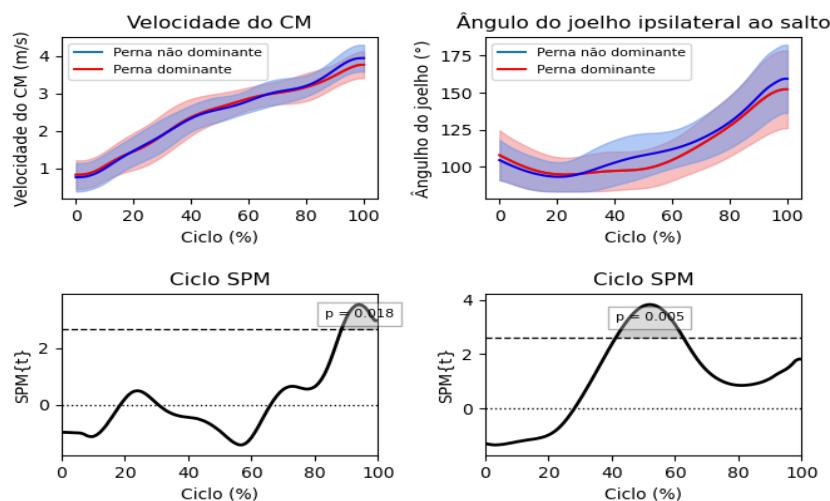


Figura 2. Velocidade do centro de massa e ângulo do joelho ipsilateral ao salto entre as tentativas para o lado da perna dominante e não dominante (em cima) e ciclo SPM (embaixo) normalizados na série temporal (0-100%). Os gráficos superiores representam a média e desvio padrão (área sombreada) da velocidade do centro de massa e do ângulo do joelho ipsilateral ao salto, enquanto os gráficos inferiores representam o SPM_t. A área sombreada nos gráficos inferiores indica significância ($p < 0,05$). SPM indica mapeamento paramétrico estatístico; SPM_t, t representa uma função do ciclo; CM indica centro de massa.

Variáveis	Perna dominante	Perna não dominante	Valor p
Tempo para velocidade resultante pico (s)	$0,538 \pm 0,08$	$0,546 \pm 0,086$	0,583
Distância do passo frontal (m)	$0,343 \pm 0,181$	$0,439 \pm 0,239$	0,009*
Ângulo de saída frontal (°)	$12,998 \pm 6,686$	$15,262 \pm 6,607$	0,051

Tabela 1 – Média e desvio padrão das variáveis analisadas nos saltos para o lado da perna dominante e não dominante. (* $p < 0,05$. t de Student pareado).

5. Conclusão

Os saltos para o lado da perna não dominante apresentaram maior velocidade do centro de massa quando comparados ao lado da perna dominante. O ângulo do joelho ipsilateral ao salto e a distância do passo frontal podem ter contribuído para essa observação. Identificar a relação da lateralidade com a performance cinemática do salto permite que treinadores identifiquem assimetrias nos atletas, o que pode representar um lado mais fraco no salto do goleiro. Aumentar a velocidade do salto lateral dos goleiros de futebol permite que eles tenham mais tempo para identificar o lado em que a bola

será chutada no pênalti, aumentando sua chance de saltar para o lado correto e realizar a defesa (SAVELSBERGH et al., 2002, 2005; WANG; JI; ZHOU, 2019). A análise cinemática do salto do goleiro de futebol atrelada ao uso de técnicas visão computacional pode ser uma importante ferramenta para: (1) aprofundar o conhecimento quanto ao salto de goleiros; (2) acompanhamento da performance do salto de atletas e (3) aperfeiçoamento do treinamento e padrão de movimento.

Referências

- CAO, Z. et al. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 43, n. 1, p. 172–186, 2019.
- FIGUEROA, P. J.; LEITE, N. J.; BARROS, R. M. L. A flexible software for tracking of markers used in human motion analysis. **Computer Methods and Programs in Biomedicine**, v. 72, p. 155–165, 2003.
- IBRAHIM, R. et al. The Effect of Preparatory Posture on Goalkeeper's Diving Save Performance in Football. **Frontiers in Sports and Active Living**, v. 1, n. August, p. 1–10, 2019a.
- IBRAHIM, R. et al. Kinematic and kinetic analysis of the goalkeeper's diving save in football. **Journal of Sports Sciences**, v. 37, n. 3, p. 313–321, 2019b.
- IBRAHIM, R. et al. Angular Velocity, Moment, and Power Analysis of the Ankle, Knee, and Hip Joints in the Goalkeeper's Diving Save in Football. **Frontiers in Sports and Active Living**, v. 2, n. February, p. 1–9, 2020a.
- IBRAHIM, R. et al. Angular Velocity, Moment, and Power Analysis of the Ankle, Knee, and Hip Joints in the Goalkeeper's Diving Save in Football. **Frontiers in Sports and Active Living**, v. 2, 28 fev. 2020b.
- MONTEIRO, R. L. M. et al. Penalty feet positioning rule modification and laterality effect on soccer goalkeepers' diving kinematics. **Scientific Reports**, v. 12, n. 1, 1 dez. 2022.
- MARIM, E. D. A.; LAFASSE, R.; OKAZAKI, V. H. Inventário de preferência lateral global (IPLAG). **Brazilian Journal of Motor Behavior**, v. 6, n. 3, 2011.
- SAVELSBERGH, G. J. P. et al. Visual search, anticipation and expertise in soccer goalkeepers. **Journal of Sports Sciences**, v. 20, n. 3, p. 279–287, 2002.
- SAVELSBERGH, G. J. P. et al. Anticipation and visual search behaviour in expert soccer goalkeepers. **Ergonomics**, v. 48, n. 11–14, p. 1686–1697, 2005.
- SPRATFORD, W.; MELLIFONT, R.; BURKETT, B. The influence of dive direction on the movement characteristics for elite football goalkeepers. **Sports Biomechanics**, v. 8, n. 3, p. 235–244, 2009.
- WANG, Y.; JI, Q.; ZHOU, C. Effect of prior cues on action anticipation in soccer goalkeepers. **Psychology of Sport and Exercise**, v. 43, n. 399, p. 137–143, 2019.

Efeitos imediatos de uma órtese elástica no tronco na biomecânica e no desempenho do chute em praticantes de futebol: um estudo de viabilidade

Sabrina P. Cintra^{1,3}, Sergio T. Fonseca³, Diego C. Souza^{1,3}, Lucas V. Freitas^{1,3}, Gabriela M. Blaso³, Luísa P. Antunes³, Fabrício A. Magalhães², Priscila A. Araújo², Liria A. Nobrega², Thales R. Souza³

¹Programa de Pós-Graduação em Ciências da Reabilitação - Universidade Federal de Minas Gerais (UFMG),

²Laboratório de Análise do Movimento - UFMG,

³Departamento de Fisioterapia – UFMG

Introdução: Considerando a participação do tronco no chute do futebol, uma órtese elástica, conectando o tronco aos membros inferiores, pode melhorar o desempenho nessa tarefa. Essa órtese pode ter efeitos imediatos na mecânica do chute, porém, há maior potencial de efeitos após um período de aprendizado, em médio e longo prazos. Antes da realização de estudos longitudinais, é necessário que seja testada a viabilidade de se investigar os efeitos da órtese na biomecânica e no desempenho do chute.

Objetivo: Esse estudo investigou se o uso imediato de uma órtese elástica: permitiria coletar a cinemática de tronco e pelve; geraria mudanças biomecânicas no tronco de acordo com o mecanismo de ação proposto; restringiria amplitudes de movimento do tronco; e prejudicaria a biomecânica do membro inferior e o desempenho do chute.

Métodos: A órtese utilizada constitui-se de quatro faixas elásticas cruzadas no tronco e quadril e fixadas por ancoragens inelásticas (figura 1). Participaram do estudo 26 homens (24 ± 13 anos; tempo de experiência 13 ± 6 anos) e foram coletados sete chutes com força máxima e dorso do pé, em quatro condições: controle, tensão confortável, “ótima” e alta tensão do elástico. Ângulos e momentos das articulações tóraco-lombar, lombo-pélvica, e do membro inferior de chute, foram medidas no plano sagital. As mesmas variáveis foram calculadas para as duas articulações do tronco no plano transverso. Amplitudes de movimento articular do tronco e membro inferior de chute e a velocidade de partida da bola foram computadas. Análises de variância (ANOVAs) de medidas repetidas e testes-t pareados, do método de Mapeamento Estatístico Paramétrico, foram utilizados

para comparação de séries temporais. ANOVAs de medidas repetidas foram usadas para comparar as amplitudes dos movimentos articulares. As comparações foram entre as condições experimentais e a condição controle. O nível de significância foi estabelecido a 0,05.



Figura 1. Composição e posicionamento da órtese elástica no corpo do participante, juntamente com os marcadores passivos para coleta de dados cinemáticos.

Resultados: A coleta de dados cinemáticos dos segmentos do tronco e pelve foi viável com o uso da órtese. A condição de alta tensão aumentou significativamente o ângulo de extensão da articulação tóraco-lombar ($p=0,027$) (figura 2). Nas condições de tensões “ótima” e alta, a órtese aumentou a rotação da articulação tóracolombar para o lado do membro de apoio ($p<0,01$) e aumentou o torque rotador de tronco para o mesmo lado ($p=0,045$) (figura 2). Para os demais ângulos, amplitudes e momentos articulares, e velocidade da bola, não houve efeito significativo.

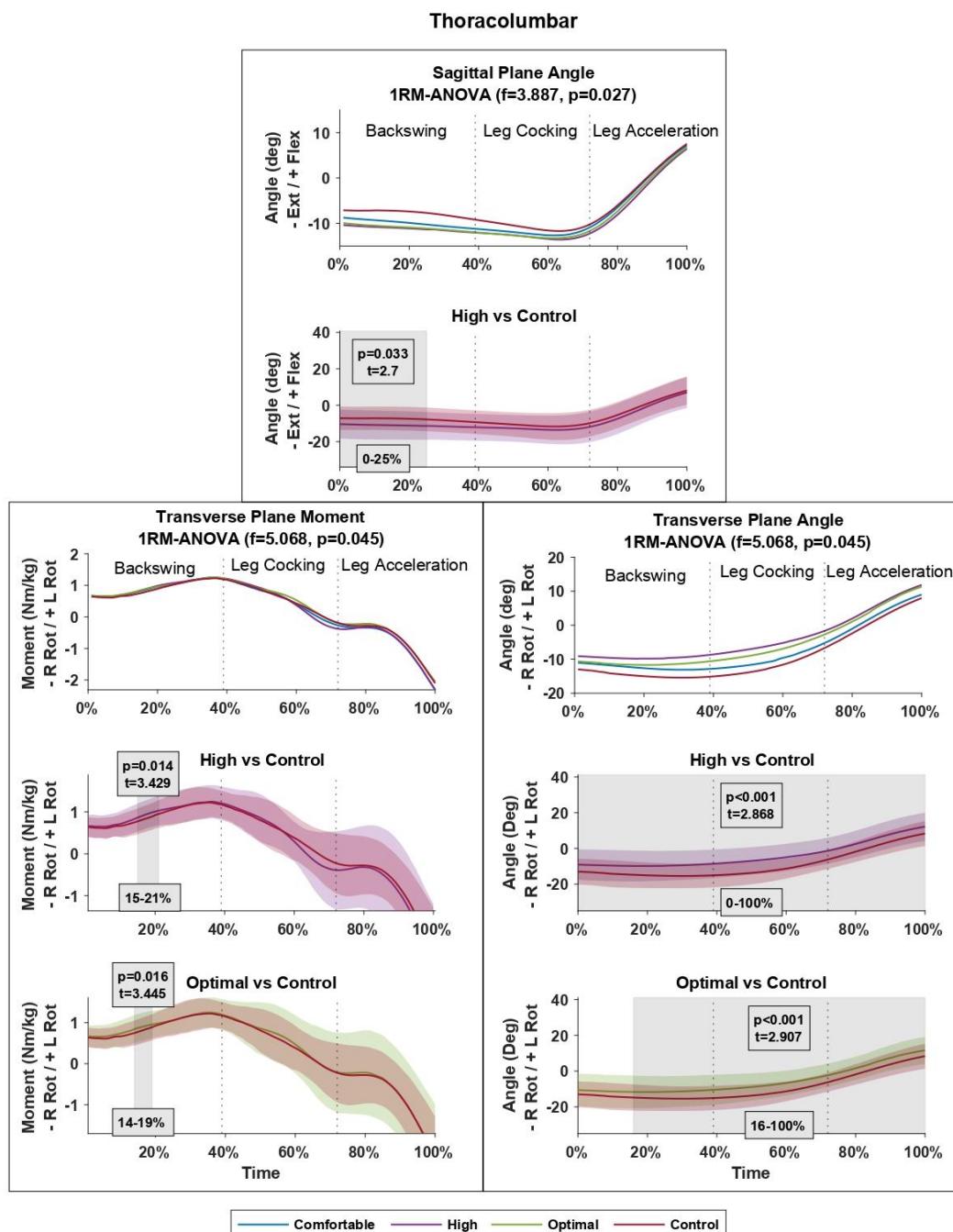


Figura 2. Curvas do ângulo toracolombar nos planos sagital e transverso, indicando a comparação entre as quatro condições de teste (SPM-ANOVAs) e comparação entre pares (SPM-teste-t), com média e desvio padrão. A região sombreada em cinza nos gráficos mostra o trecho da curva onde houve diferença significativa entre as condições de teste.

Conclusão: O uso imediato da órtese elástica no tronco (a) não inviabilizou a coleta de dados laboratoriais, (b) produziu efeitos coerentes com o mecanismo de ação proposto, apesar de terem

ocorrido apenas para a articulação tóraco-lombar (e não para a lombo-pélvica), (c) não restringiu as amplitudes de movimento articular, e (d) não levou a mudanças inesperadas ou indesejadas na biomecânica do membro de chute e no desempenho do chute. Assim, pode-se considerar que o estudo dos efeitos da órtese elástica é viável e que os efeitos a médio e longo prazos, após período de aprendizado, devem ser investigados.

Palavras-chave: Biomecânica. Chute de futebol. Cinemática do tronco. Cinética do tronco. Órtese. Faixa elástica.

Referências

1. Bregman DJ, Harlaar J, Meskers CG, de Groot V. Spring-like Ankle Foot Orthoses reduce the energy cost of walking by taking over ankle work. *Gait Posture*, 35(1):148-53, 2012.
2. Collins, SH; Wiggin, MB; Sawicki, GS. Reducing the energy cost of human walking using an unpowered exoskeleton. *Nature*, v. 522, n. 7555, p. 212-5, Jun 2015
3. Robertson DG, Winter DA. Mechanical energy generation, absorption and transfer amongst segments during walking. *J Biomech*. 1980;13(10):845-54.



II Football Analytics Modelling & Experience

17 de novembro de 2023

Belo Horizonte, Brasil

Compêndio de Artigos

Realização

SALab - Sports Analytics Lab

Organização

Departamento de Ciência da Computação - DCC - UFMG