Funções Executivas e Idade Relativa como Preditores de Sucesso no Futebol

Matheus Mello[1], Vitor Belloni[1], Fabricio Vasconcellos[2], Jorge Soares[1], Eduardo Ogasawara[1], Lucas Giusti[1,2]

[1]CEFET/RJ - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca

[2]UERJ - Universidade do Estado do Rio de Janeiro

matheus.melo@aluno.cefet-rj.br, vitor.belloni@aluno.cefet-rj.br

fabricio.vasconcellos@uerj.br,jorge.soares@cefet-rj.br

eogasawara@ieee.org, lucas.giusti@aluno.cefet-rj.br

Resumo. *A identificação bem-sucedida de potenciais futuros atletas de elite em uma idade precoce ajuda a fornecer treinamentos de alta qualidade e ambientes de treino para otimizar sua evolução e desenvolvimento. No entanto, uma grande variedade de diferentes habilidades e qualidades são necessárias para ter sucesso em esportes de elite, tornando a identificação de talentos geralmente um problema complexo e multifacetado a se enfrentar. Sendo assim, o presente estudo tem o objetivo de investigar a relação de funções executivas (FE) e idade relativa (IR) com o sucesso de atletas de elite de futebol de base. Especificamente, os testes utilizados foram o Color-Word Interference Test (CWI), Trail Making Test (TMT) e o Teste dos Cinco Pontos (TCP). Foram feitos testes de comparação entre grupos de sucesso e controle, além de um modelo preditivo. Foi possível identificar que os testes de CWI1, CWI3, TMT(b) e TCP3 estiveram relacionados a presença dos atletas em uma plataforma pública confiável de dados (Transfermarkt) de atletas profissionais entre 3 e 4 anos após os testes.*

Abstract. *The successful identification of potentially talented professional athletes at early ages may help to provide optimal training quality and environment to maximize the talent's evolution. However, a variety of skills and qualities are needed to achieve success in elite sports, turning talent identification into a complex and multifaceted problem. Therefore, the present study aims to investigate the relation of executive functions (FE) and relative age (RA) with the success of elite youth soccer athletes. Specifically, the Color-Word Interference Test (CWI), Trail Making Test (TMT), and Design Fluency (DF) were used. Statistical tests were made to compare success and control groups and a predictive machine learning model was also developed. It was possible to identify that CWI1, CWI3, TMT(b), and DF3 were related to the presence of the athletes in a public reliable data platform (Transfermarkt) of professional soccer players 4 to 5 years after the FE tests.*

1. Introdução

Atualmente, organizações esportivas investem recursos substanciais na procura de jogadores com potencial para se destacar. Esses programas de identificação visam detectar

jogadores talentosos que demonstram forte desempenho em esportes específicos, apresentando preditivos de sucesso na carreira futura do atleta [Bergkamp et al., 2019]. A identificação bem-sucedida de potenciais futuros atletas de elite em uma idade precoce ajudaria a fornecer treinamentos de alta qualidade e ambientes de treino para otimizar sua evolução e desenvolvimento.

No entanto, uma grande variedade de diferentes habilidades e qualidades são necessárias para ter sucesso em esportes de elite, tornando a identificação de talentos geralmente um problema complexo e multifacetado a se enfrentar [Jauhiainen et al., 2019]. Devido à subjetividade, as trajetórias de desenvolvimento dos jogadores raramente são lineares, pois as suas habilidades cognitivas e motoras estão interligadas e se desenvolvem por meio de interações dinâmicas com o ambiente de desempenho individual do atleta [Bergkamp et al., 2019].

Programas de identificação de talentos (TID) são projetados para identificar jovens atletas com potencial de sucesso em esportes de elite sênior. Nos últimos anos, os programas TID têm crescido em popularidade e são vistos como caminhos críticos para maximizar o potencial dos atletas para alcançar o sucesso [Johnston et al., 2018].

Além do futebol, a identificação de talentos é muito influente para a medição do sucesso de jovens jogadores em outros esportes, como no caso do basquete, por exemplo. No entanto, o basquete é jogado como um esporte de equipe com várias funções e responsabilidades, portanto, uma infinidade de fatores desempenha um papel na decisão do resultado do jogo. Os gerentes procuram jogadores que podem combinar capacidades atléticas superiores com outros fatores [Silva et al., 2020]. Um dos fatores decisivos para ganhar um jogo no basquete é ter menos erros do que a equipe adversária. Para obter este objetivo, os jogadores devem perder menos a bola e fazer mais decisões corretas com e sem a bola [Silva et al., 2020].

No futebol, a identificação de talentos é uma das mais importantes tarefas dos treinadores das seleções juvenis de futebol e essa identificação, em termos de idade cronológica, é muito baixa [Mustafovic et al., 2020]. Uma das razões pelas quais os clubes estão tentando identificar jogadores em tais a idade precoce é o medo de que a seleção seja feita tarde demais. Também há duas razões: 1) desenvolvimento de um jogador 2) medo daquele jogador talentoso ser selecionado por outros clubes [Mustafovic et al., 2020]. A identificação e desenvolvimento de talentos no futebol juvenil masculino são processos complexos e multidimensionais. A Associação Inglesa de Futebol "4 *Corner Model*" é frequentemente adotado por academias para apoiar treinadores e praticantes para implementar uma multidisciplinaridade de abordagem para sua identificação de talentos e estrutura de desenvolvimento. Essa abordagem holística é caracterizada por técnico/tático, psicológico, atributos sociais e físicos, que são fundamentais para o recrutamento da academia e desenvolvimento individual de atleta de longo prazo [Kelly et al., 2020].

A capacidade de obter informações além de uma análise subjetiva permite que os treinadores reconheçam potenciais jogadores de elite que podem se encaixar no estilo de jogar em seu clube. Pode ser que certos clubes identifiquem o tamanho, força e velocidade como fundamentais para sucesso em seu clube e podem querer apenas selecionar indivíduos com esses pré-requisitos físicos de atuação. Por outro lado, outros clubes podem se esforçar para adotar um estilo de jogo mais "criativo" com maior ênfase na habilidade e

técnica. O Ajax FC, clube de sucesso na Holanda, por exemplo, famoso por sua linha de produção de jovens talentos, incentiva os treinadores a usar a sigla TIPS (Técnica, Inteligência, Personalidade e Velocidade) dentro de sua prática de identificação de talentos [Unnithan et al., 2012].

No futebol, uma grande variedade de características físicas e habilidades técnicas são necessárias para o sucesso. Além disso, as habilidades e características psicológicas também desempenham um papel importante no nível de elite. Portanto, a identificação de talentos no esporte deve ser baseada em um conjunto versátil de variáveis [Jauhiainen et al., 2019]. Dentro desse conjunto de variáveis, a idade relativa (IR) e as FE dos jogadores demonstram suma importância. A IR se refere à diferença de idades entre as crianças da mesma faixa etária que resulta de suas diferentes datas de nascimento ao longo do ano. Assim, analisando em um cenário escolar, por exemplo, se uma classe da primeira série for composta por crianças nascidas de acordo com o "ano escolar" referente, de 1º de setembro a 31 de agosto, então as crianças com datas de nascimento em setembro terão aproximadamente uma vantagem de IR de um ano sobre as crianças nascidas em agosto do ano seguinte [Barnsley et al., 1992].

A maioria das pesquisas relacionadas a IR no futebol juvenil revelaram que o mês que o atleta faz aniversário tem um impacto significante na identificação de talentos e seu desenvolvimento [Kelly et al., 2020]. Já as funções executivas (FE) gerenciam outras funções cognitivas mais básicas e envolvem funções como inibição de comportamento, atenção e memória de trabalho, sendo importantes para o desempenho bem-sucedido em vários esportes, porque essas funções facilitam a adaptação a novas situações ou mudanças, a atenção e lembrança de estratégias de jogo [Verburgh et al., 2014].

De acordo com a literatura, existem poucos artigos que trabalham métodos de inteligência artificial (IA) utilizando os parâmetros de IR e FE a fim de identificar talentos no futebol. Dessa maneira, a contribuição científica deste trabalho foca em: 1) identificação de variáveis importantes para o sucesso através de análises estatísticas; 2) Reprodução de modelos de aprendizado de máquina para prever o sucesso dos atletas.

2. Métodos

Os dados alvo deste estudo foram coletados nas categorias de base de três clubes de elite do Rio de Janeiro para acompanhamento do desenvolvimento de jovens jogadores de futebol em busca de suas FE, tendo sido aprovado pelo comitê de ética da Universidade do Estado do Rio de Janeiro com o número de registro CRD42017077640.

As informações de sucesso dos jogadores foram retiradas do Transfermarkt, um site de estatísticas de futebol que é conhecido principalmente por publicar avaliações monetárias, apelidadas de "valores de mercado", para uma amostra muito grande de jogadores profissionais de futebol. Esses valores de mercado são consultados regularmente tanto por pesquisadores na literatura de economia ou administração do esporte quanto na imprensa popular [Peeters, 2018].

Um total de 199 jogadores de futebol juniores com idade variando entre 16 e 19 anos participaram dos eventos de teste específicos entre os anos de 2016 e 2017. Alguns dos jogadores testados estão atualmente buscando uma carreira no futebol internacional ou já atuam por um clube internacional, variando entre as categorias de base *Sub-20* (Jogadores com 20 anos ou abaixo), *Sub-23* (Jogadores com 23 anos ou abaixo) ou atuantes

no Profissional. Dessa forma, visando o âmbito preditivo deste trabalho, a intenção será de, através de estratégias de modelagem, analisar quais jogadores ainda estariam com seus dados sendo ainda atualizados no site do Transfermarkt atualmente, 4 a 5 anos após os testes realizados com os atletas). A premissa considerada foi de que atletas que não estejam cadastrados no Transfermarkt teriam uma probabilidade significativamente menor que já ter alcançado o nível profissional.

As FEs são habilidades cognitivas que permitem o controle e regulação de pensamentos e ações. Algumas dessas funções são chamadas de funções centrais e são base para todas as outras mais complexar geralmente endereçadas como funções de alto nível. No presente trabalho, cinco testes de FE foram realizados para fins de avaliação. Os testes selecionados são reconhecidos como relacionados às funções centrais de inibição e flexilidade cognitiva, principalmente. A inteligência fluida também tem papel importante, especialmente no Teste dos Cinco Pontos e Teste de Trilhas, que serão descritos nessa seção.[Diamond, 2013; Vestberg et al., 2012]

Para obtenção dos valores de FE, os jogadores foram submetidos a uma sequência de testes cognitivos. Todos os testes foram realizados em papel e caneta. Antes da realização dos sub testes, os avaliados puderam realizar uma tarefa similar menor sem limite de tempo com o intuito de sanar eventuais dúvidas e se familiarizarem com o teste. Após essa etapa de familiarização, iniciava-se o teste. Os testes submetidos foram:

Teste dos cinco pontos (TCP): Consiste no participante usar as mãos com o auxílio de papel e caneta afim de combinar todos os pontos dentro de um quadrado com uma só linha em um período de 60 segundos. Três sub testes foram utilizados com regras baseadas em quais pontos poderiam ser combinados: 1) Somente pontos pretos; 2) Somente pontos brancos; 3) Todos os pontos pretos e brancos, alternadamente. Foi considerado para análise dos dados a soma de todas as figuras corretas e as combinações que não atingem os requerimentos são excluídos. [Vestberg et al., 2012; Huijgen et al., 2015]

Teste de trilhas (TMT(b)): O participante usa as mãos com auxílio de papel e caneta para combinar todas as figuras com uma linha. As figuras possuem ordens diferentes em cada sub teste. No sub teste A, as figuras são ordenadas como números, de 1 até 25. Na versão B, as figuras são ordenadas como números e letras subsequentemente (1, A, 2, B, 3, C). Em ambos os testes, é necessário combinar todas as figuras na ordem correta no menor tempo possível e o tempo gasto para a realização é utilizado como resultado dos sub testes. O tempo do sub teste B foi considerado como resultado do teste [Huijgen et al., 2015].

Color-Word Interference Test (CWI3): O teste consiste em três sub testes. No primeiro sub teste, uma lista de palavras com nomes de cores é apresentada com cor de fonte preta e o participante deve falar em voz alta todos os nomes apresentados no monitor no menor tempo possível. No segundo sub teste, letras X são apresentadas em diferentes cores e os participantes devem nomear as cores das fontes. Finalmente, no terceiro sub teste, os nomes das cores são mostrados novamente, mas em cores não correspondentes de fontes (palavra azul escrita na cor vermelha), e a instrução é nomear a cor da fonte de cada palavra no menor tempo possível, ignorando a palavra escrita. O resultado foi considerado como o tempo necessário para responder o terceiro sub teste [Vestberg et al., 2012].

Idade Relativa. O parâmetro de IR a ser levado como critério no trabalho foi principalmente com a variável *Mês*. A partir da data de nascimento coletada e analisada de cada jogador na base de dados, foi selecionado o mês respectivo que cada atleta nasceu e atribuído sua categoria respectiva. A definição de categoria é baseada no calendário iniciado em janeiro e finalizado em dezembro, demonstrando uma vantagem para os jogadores ao início do mês, graças à lógica sobre IR e suas vantagens graças à ela [Barnsley et al., 1992].

Teste de Mann Whitney. No campo das ciências comportamentais, o Mann-Whitney U é um dos testes não paramétricos comumente mais usados. Este teste foi independente elaborado por Mann and Whitney [1947] e Wilcoxon [1945]. Este método é, portanto, frequentemente chamado de teste Mann-Whitney ou o teste de soma de classificações de Wilcoxon [Nachar et al., 2008]. O teste de Mann-Whitney U inicialmente implica o cálculo de uma estatística U para cada grupo. Essas estatísticas têm uma distribuição conhecida sob a hipótese nula identificado por Mann e Whitney (1947). No presente trabalho, o teste de Mann Whitney foi usado para identificar variáveis que mostrem diferença significativa entre os grupos de sucesso considerados. O valor alpha foi determinado em 0.05 para todos os testes.

Modelagem Random Forests. Um dos nossos objetivos é investigar com que qualidade podemos detectar potenciais futuros jogadores de elite entre o grande grupo de jogadores com base nas informações de teste coletadas. Dessa maneira, utilizamos o classificador preditivo *Random Forests* [Breiman, 2001]. Este é um classificador que consiste em uma coleção de classificadores estruturados em árvore {h(**x**, Θ_k), k= 1, ...} onde $\{\Theta_k\}$ são vetores aleatórios distribuídos de forma idêntica e cada árvore lança um voto unitário para a classe mais popular na entrada **x**. Assim, para a k-ésima árvore, é gerado o vetor aleatório Θ_k, independentemente dos respectivos vetores anteriormente criados $\Theta_1, ..., \Theta_k - 1$, porém com a mesma distribuição. Além disso, a árvore "cresce"utilizando um método de treinamento e Θ_k, resultando no classificador h(**x**, Θ_k), onde **x** é o vetor de entrada. Consequentemente na etapa final, após um vasto número de árvores geradas, a classe mais popular é designada de acordo com o sistema de votação entre a grande mazela de árvores aleatórias que votam na classe, devido a isso o famoso nome de Random Forests [Breiman, 2001]. Para este projeto, a etapa de treinamento do modelo foi baseada em 75% treino e 25% teste.

3. Resultados

3.1. Comparação com Mann Whitney

Os resultados dos testes de comparação estão indicados na Tabela 1. Considerando a complexidade dos testes, é possível observar que os níveis mais complexos mostraram diferença significativa entre a classe de atletas encontrados quando comparados ao restante. Esses resultados corroboram com estudos passados, onde testes mais complexos também mostraram relações significativas com o sucesso [Vestberg et al., 2012; Huijgen et al., 2015].

Sendo assim, esses resultados foram utilizados para selecionar as variáveis do modelo preditivo descrito na próxima subseção. Especificamente, foram treinados dois modelos, um com todas as variáveis e outro somente os resultados dos testes mais complexos (TCP3, CWI3 e TMT(b)).

Tabela 1. Resultados dos testes de Mann Whitney U. * p < 0.05; ** p < 0.01

Modelo	u	p
CWI1	1576.0	0.003**
CWI2	2079.0	0.190
CWI3	1702.5	0.012*
TMT(a)	2092.0	0.204
TMT(b)	1943.0	0.086
TCP1	1991.0	0.116
TCP2	1858.5	0.047*
TCP3	1700.5	0.012*
Mês	1204.0	0.239

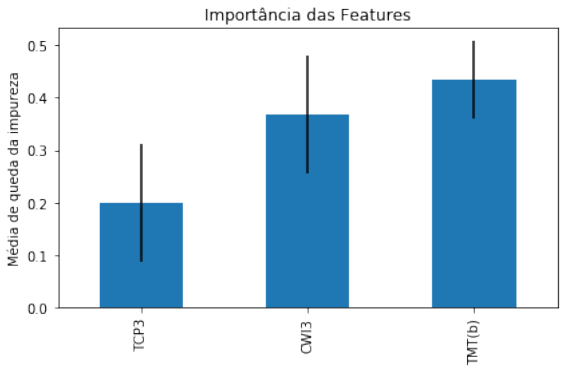
Em relação à IR (Mês), não houve significância na comparação entre grupos. Esse resultado é contrário a alguns outros estudos que mostraram alguma relação da IR no sucesso de atletas de futebol e outros esportes [Mulazimoglu, 2014; Cobley et al., 2008].

3.2. Modelagem com Random Forests

Foram desenvolvidos dois modelos de aprendizado de máquina com o objetivo de prever se o atleta estaria cadastrado na plataforma Transfermarkt entre 3 e 4 anos após os testes de FE. Os resultados de desempenho preditivo estão descritos na Tabela 2.

Tabela 2. Desempenho preditivo dos modelos

Modelo	AUC
Todas as FE e Mês	57%
Somente CWI3, TMT(b) e DF3	68%



(c)

Figura 1. Importância das features medida pela queda média na impureza.

É possível observar que o desempenho do modelo melhora de maneira importante após a remoção de variáveis que não foram significativas nos testes de comparação. Especificamente, o ganho de desempenho é de 10% na AUC.

Por fim, as importâncias das variáveis independentes foram calculadas e consolidadas no gráfico representado na Figura 1. É possível observar que testes de FE principalmente baseados no controle inibitório (TMT(b) e ST3), mostraram valores maiores de

importância para o desempenho do modelo. Isso indica o desenvolvimento ótimo dessa FE pode ser importante para o sucesso de atletas de futebol.

4. Conclusões

O presente trabalho investigou a influência de FEs e IR no sucesso de atletas de elite de futebol de base. Foi possível identificar que os testes de CWI1, CWI3, TMT(b) e TCP3 estiveram relacionados a presença dos atletas em uma plataforma pública confiável de dados (Transfermarkt) de atletas profissionais entre 3 e 4 anos após os testes.

Em relação às FE, foi possível observar que as fases mais complexas dos testes parecem ter potencial preditivo para a presença dos atletas no Transfermarkt anos após os testes. O pouco tempo de acompanhamento após os testes (4 a 5 anos) pode ser destacado como uma limitação. Futuros estudos devem investigar mais anos de acompahamento, bem como outras FE como planejamento e fluência verbal.

Referências

R.H. Barnsley, A.H. Thompson, and P. Legault. Family Planning: Football Style. The Relative Age Effect in Football. *International Review for the Sociology of Sport*, 27(1):77–87, 1992.

T.L.G. Bergkamp, A.S.M. Niessen, R.J.R. den Hartigh, W.G.P. Frencken, and R.R. Meijer. Methodological Issues in Soccer Talent Identification Research. *Sports Medicine*, 49(9):1317–1335, 2019.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Stephen P. Cobley, Joerg Schorer, and Joseph Baker. Relative age effects in professional german soccer: A historical analysis. *Journal of Sports Sciences*, 26(14):1531–1538, 2008. doi: 10.1080/02640410802298250. URL https://doi.org/10.1080/02640410802298250. PMID: 19040189.

Adele Diamond. Executive functions. *Annual review of psychology*, 64:135–168, 2013.

B.C.H. Huijgen, S. Leemhuis, N.M. Kok, L. Verburgh, J. Oosterlaan, M.T. Elferink-Gemser, and C. Visscher. Cognitive functions in elite and sub-elite youth soccer players aged 13 to 17 years. *PLoS ONE*, 10(12), 2015.

S. Jauhiainen, S. Aÿrämö, H. Forsman, and J.-P. Kauppi. Talent identification in soccer using a one-class support vector machine. *International Journal of Computer Science in Sport*, 18(3): 125–136, 2019.

K. Johnston, N. Wattie, J. Schorer, and J. Baker. Talent Identification in Sport: A Systematic Review. *Sports Medicine*, 48(1):97–109, 2018.

A. Kelly, M.R. Wilson, D.T. Jackson, and C.A. Williams. Technical testing and match analysis statistics as part of the talent development process in an English football academy. *International Journal of Performance Analysis in Sport*, 20(6):1035–1051, 2020.

Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

Olcay Mulazimoglu. The relative age effect (rae) in youth and professional soccer players in turkey. *The Anthropologist*, 18(2):391–398, 2014. doi: 10.1080/09720073.2014.11891557. URL https://doi.org/10.1080/09720073.2014.11891557.

Emir Mustafovic, Denis Causevic, Nedim Covic, Mirza Ibrahimovic, Haris Alic, Ensar Abazovic, and Semir Masic. Talent Identification in Youth Football: A Systematic Review. *Journal of Anthropology of Sport and Physical Education*, 4(4):37–43, 2020.

Nadim Nachar et al. The mann-whitney u: A test for assessing whether two independent samples come from the same distribution. *Tutorials in quantitative Methods for Psychology*, 4(1):13–20, 2008.

Thomas Peeters. Testing the wisdom of crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, 34(1):17–29, 2018. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2017.08.002. URL https://www.sciencedirect.com/science/article/pii/S0169207017300754.

A.F. Silva, D. Conte, and F.M. Clemente. Decision-making in youth team-sports players: A systematic review. *International Journal of Environmental Research and Public Health*, 17 (11), 2020.

V. Unnithan, J. White, A. Georgiou, J. Iga, and B. Drust. Talent identification in youth soccer. *Journal of Sports Sciences*, 30(15):1719–1726, 2012.

L. Verburgh, E.J.A. Scherder, P.A.M. Van Lange, and J. Oosterlaan. Executive functioning in highly talented soccer players. *PLoS ONE*, 9(3), 2014.

T. Vestberg, R. Gustafson, L. Maurex, M. Ingvar, and P. Petrovic. Executive functions predict the success of top-soccer players. *PLoS ONE*, 7(4), 2012.

F Wilcoxon. Individual comparisons by ranking methods. biom. bull., 1, 80–83. 1945.