

Sport Analytics A Systematic Review

Jair Bogo¹, Janine F. Oliveira¹, Elivelton R. Lima²,
Rafael Ferreira Mello², Ana Paula C. Furtado^{1,2}

¹Centro de Estudos e Sistemas Avançados do Recife (CESAR SCHOOL)
AV. Cais do Apolo, 77 – Recife, PE – Brasil

²Universidade Federal Rural de Pernambuco (UFRPE)
Recife, PE – Brasil

{jbo, jfo}@cesar.school,

{elivelton.rodriques, rafel.mello, anapaula.furtado}@ufrpe.br

Abstract. *Data Science is a field related to Information Technology with the highest development in the last few years. It is being used in a wide range of sectors, strongly impacting the competitiveness of companies and how people interact with products and services. When applied in the context of sports, known as Sports Analytics, it is used to extract patterns, define/improve tactics/training sections, and predict results. This article presents a summary of the recent literature of this field, especially explaining how the knowledge is acquired and applied in this field.*

Resumo. *Ciência de Dados é uma das áreas da Tecnologia de Informação em maior desenvolvimento, sendo usada nos mais diversos ramos, impactando fortemente a competitividade das empresas e a forma como as pessoas interagem. O seu uso nos esportes é designado como Sport Analytics, sendo usado para entender padrões, definir táticas/treinamentos e prever resultados. Este artigo procura resumir as pesquisas que estão sendo realizadas neste campo, bem como entender o conhecimento gerado e como ele está sendo aplicado.*

1. Introdução

Com o advento da Internet das Coisas (IoT) e a popularização de dispositivos móveis, os quais trazem embutidos a capacidade de utilização do *Global Position System* (GPS), que capturam e geram dados em grandes quantidades, aliados à facilidade de armazenamento propiciada pelo ambiente de nuvem, novas oportunidades surgiram para a indústria e para a ciência de dados [Value and Team 2015].

Com o enorme volume de dados gerados por estes dispositivos, a tradicional separação entre a matemática estatística e análise de dados não mais atendeu a necessidade imposta pelo mercado [Dhar 2013]. Devido a essa necessidade, o conceito de Ciência de Dados popularizou-se focando principalmente na produção, coleta e análise de dados, amparado pelo crescimento e barateamento do poder computacional disponível [Hayashi 1998].

Ciência de dados está presente nas mais diversas áreas de atuação, primeiramente impulsionado pela capacidade de entender eventos passados com a utilização de algoritmos e características de aprendizado de máquina. Porém, as aplicações mais relevantes que dão destaque a esse conhecimento/técnica é que podem, também, prever eventos futuros, comportamentos em massa e categorizar os mais diversos assuntos e possíveis ações, trazendo maior competitividade às empresas que utilizam essas técnicas [Dhar 2013].

Para os esportes devemos considerar que as empresas ligadas ao setor movimentam cerca de 2,6% do PIB europeu [UE 2018], tornando-se uma área importante para uso de Ciência de Dados, que quando aplicada aos esportes recebeu uma definição própria. Para [Costa et al. 2017] *Sport Analytics* é a forma "como a análise de uma coleção de dados históricos que, se realizada adequadamente, pode trazer vantagem competitiva para um time ou atleta".

O interesse em análise de dados esportivos não se restringe apenas ao mercado ou empresas, a academia por sua vez, possui diversas pesquisas relacionadas ao assunto. Segundo [Cokins et al. 2017], definiu-se uma taxonomia para classificação das pesquisas em análise de dados nos esportes, destacando-se a divisão em três grupos principais: grupo 1: dados usados com foco em competições; grupo 2: trata o esporte como recreação e grupo 3: análise para subsidiar as casas de apostas.

A motivação deste trabalho é compreender como está atualmente o conhecimento relacionado com *Sport Analytics* e seu impacto na sociedade, focando nos atletas e nas equipes. Esse trabalho está organizado da seguinte maneira: inicialmente uma apresentação dos conceitos básicos relacionados com o tema. Na próxima seção, os métodos, procedimentos e protocolos usados na revisão sistemática são apresentados. Na seção seguinte, os resultados e análise dos dados de forma detalhada. Finalmente, na última seção, expomos algumas conclusões e propostas para trabalhos futuros.

2. Protocolo Aplicado

Adaptamos o modelo de Mapeamento Sistemático para Engenharia de Software descrito por [Kitchenham and Charters 2007] juntamente com o descrito por [Budgen and Brereton 2006] para definirmos o modelo de revisão sistemática a ser seguido no presente trabalho contendo os seguintes passos: (1) definição do tema com pergunta de pesquisa e perguntas secundárias; (2) definição dos critérios de inclusão e exclusão; (3) definição das strings de busca; (4) pesquisa de estudos com revisão de pares; (5) leitura dos artigos selecionados e (6) análise, que serão detalhadas a seguir.

2.1. Pergunta de pesquisa

O objetivo desta revisão é identificar estudos primários que utilizem Ciência de Dados aplicados aos esportes. Para atingir esse objetivo foi usado a questão a seguir como base da pesquisa:

Como a análise de dados pode melhorar o desempenho de atletas?

Para melhorar o entendimento e guiar a revisão foram elaboradas perguntas secundárias que sustentam o assunto principal:

- Quais ferramentas podem ser utilizadas para coletar dados de atividades dos atletas?

- Como IoT pode ajudar na coleta de dados de atletas?
- Como categorizar os atletas?
- Quais indicadores utilizar para análise de desempenho dos atletas?
- Como dados históricos podem prever o desempenho do atleta?
- Quais são os desafios e oportunidades relacionados à análise de dados e planejamento esportivo?

2.2. Strings de busca de artigos

Nesse estudo foram utilizadas as seguintes bases de divulgação científica com acesso dentro do ambiente acadêmico: ACM Digital Library; IEEE Xplore; ScienceDirect – Elsevier e SCOPUS seguindo orientações de [Kitchenham and Charters 2007]. Foram usadas duas strings de busca, "Sports Analytics" e "Athlete performance" sem a utilização de sinônimos, da seguinte forma: "(("Athlete performance")) OR ("Sports Analytics"))".

A sintaxe pode variar de acordo com a base científica pesquisada, porém constantemente foi usado o padrão de pesquisa mais amplo, ou seja, que trouxesse mais trabalhos. O filtro poderia ser usado no título, ou metadados ou até mesmo no conteúdo do artigo, quando essa opção era disponibilizada. O resultado obtido, após a aplicação das strings de buscas, está descrito na tabela 1.

Tabela 1. Total de artigos selecionados por base

Base	Quantidade de Artigos
ACM Digital Library	41
IEEE	1.872
ScienceDirect – Elsevier	541
SCOPUS	1.564
Total	4.018

2.3. Critérios de inclusão e exclusão

Durante o desenvolvimento deste trabalho, foi usado como base de pesquisa artigos que tratassem do uso de dados aplicados em alguma área do esporte, privilegiando aqueles que aplicassem algum algoritmo de Inteligência Artificial. Para isso os estudos deveriam responder ao menos uma ou mais das perguntas de pesquisa considerando as perguntas secundárias e que fossem publicados entre 01 de janeiro de 2014 e 31 de maio de 2019.

Os critérios de exclusão usados, para a escolha dos trabalhos, foram:

- estudos não listados pelas palavras-chaves (termos de busca);
- que não estejam em formato texto;
- que não demonstrem evidências empíricas;
- com dados parciais, estudo ainda em progresso ou, resumos;
- que não sejam estudos científicos;
- que exibam apenas opiniões pessoais, inclusive de pesquisadores experientes da área;
- que não estejam escritos em inglês;
- não acessíveis na Web (considerando que as bases de pesquisas sejam acessíveis no ambiente acadêmico);
- focados apenas em subsidiar casas de apostas.

2.4. Aplicação do protocolo

A primeira fase consistiu no acesso às bases de pesquisa disponíveis na internet dentro do ambiente acadêmico para aplicar as strings de busca de acordo com a sintaxe apropriada em cada base científica. O resultado dessa pesquisa foi exportado para arquivos do tipo csv com o objetivo de possibilitar maior mobilidade e manipulação dos dados para geração de estatísticas. No caso da *ScienceDirect* a exportação foi feita no formato bib e convertido posteriormente para csv totalizando 4.018 artigos selecionados.

No segundo passo foi feita a leitura dos títulos baseados nos critérios de inclusão e exclusão buscando sempre respostas às questões de pesquisa. As leituras foram realizadas por duas pessoas em cada artigo, marcando com 'S' como aprovado ou 'N' como não aprovado para o próximo passo. Foram selecionados apenas os artigos onde os dois pesquisadores tinham marcado com 'S', totalizando 191 artigos para serem analisados na próxima fase.

Em seguida foi aplicado o mesmo padrão para a leitura do resumo dos artigos, com revisão por pares, resultando em um total de 73 artigos selecionados pelos pares. No quarto passo foi realizada a leitura da introdução e conclusão, também, com revisão por pares selecionando finalmente 24 artigos que compuseram o resultado deste trabalho. Na tabela 2 mostramos a quantidade de trabalhos selecionados por fase e a taxa de concordância entre os avaliadores, onde foi definido como concordância quando ambos os pares marcaram o mesmo artigo como 'S' ou 'N'.

Tabela 2. Quantidade de artigos por fase de aplicação do protocolo

Fase	Quantidade de Artigos	Taxa de Concordância
Buscas nas bases	4.018	-
Análise dos títulos	191	81,04%
Análise do resumo	73	62,03%
Análise da introdução/conclusão	24	60,27%

2.5. Leitura dos artigos e aplicação dos critérios de qualidade

Após a seleção dos pares aplicado nas fases descritas anteriormente, uma leitura criteriosa dos estudos selecionados foi realizada e aplicado os critérios de qualidade baseados nos modelos descritos por [Kitchenham and Charters 2007] e [Budgen and Brereton 2006], assim foram elencados oito critérios, julgados suficientes para classificar os estudos selecionados através dos filtros de busca.

- Q1: O estudo examina como *Sport Analytics* pode melhorar a vida dos atletas?;
- Q2: O estudo é baseado em pesquisa - não apenas em opiniões de especialistas?;
- Q3: Os objetivos do estudo estão claramente definidos?;
- Q4: O contexto do estudo foi adequadamente descrito?;
- Q5: Os métodos para coleta de dados foram corretamente usados e descritos?;
- Q6: O projeto de pesquisa foi adequado para atingir os objetivos da pesquisa?;
- Q7: Os resultados da pesquisa foram adequadamente validados?;
- Q8: O estudo contribui para a pesquisa ou para as necessidades diárias dos cidadãos de alguma forma?;

O primeiro critério foi proposto com o objetivo de avaliar se o estudo estava focado no bem estar ou evolução dos atletas. Já os critérios Q2 a Q7 medem a forma da pesquisa, o quanto se aproxima de estudos que podem ser reproduzíveis e seguem os padrões de escrita de exposição de resultados de pesquisas científicas. A pergunta Q8 possui a finalidade de medir se o estudo contribui para as pesquisas ou se tem aplicabilidade para a sociedade em geral ou aos atletas em especial.

O resultado da aplicação dos critérios de qualidade estão mostrados na tabela 3. Esta etapa, dentro do protocolo adotado na pesquisa, não elimina nenhum trabalho, mas sugere quais são os artigos que serão mais usados na discussão dos resultados.

Tabela 3. Análise qualitativa dos artigos

Referência	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total
[Ruiz et al. 2017]	0	1	1	1	1	1	1	0	6
[Aoki et al. 2017]	0	1	1	1	1	1	1	0	6
[Op De Beéck et al. 2018]	1	1	1	1	1	1	1	1	8
[Naglah et al. 2018]	1	1	1	1	1	1	0	1	7
[Moatamed et al. 2017]	1	1	0	1	1	0	1	1	6
[Kasiri-Bidhendi et al. 2015]	0	1	1	1	1	1	1	0	6
[Fernandez et al. 2016]	1	1	0	1	0	0	0	1	4
[Altini and Amft 2018]	1	1	1	1	1	1	1	1	8
[Kempe et al. 2018]	0	1	1	1	0	1	0	0	4
[Rossi et al. 2017]	1	1	1	1	1	1	0	1	7
[Anand et al. 2018]	0	1	1	1	1	1	0		5
[Cintia et al. 2015]	0	1	1	1	1	1	1	0	6
[Pariath et al. 2018]	0	1	1	1	1	1	1	0	6
[Gu et al. 2019]	0	1	1	1	0	1	1	0	5
[Young et al. 2019]	0	1	1	1	1	1	1	0	6
[de Leeuw et al. 2018]	1	1	1	1	1	1	1	0	7
[Nagarajan and Li 2018]	0	1	1	1	1	1	1	1	7
[Hamdad et al. 2018]	0	1	1	1	1	1	1	1	7
[Dubbs 2018]	0	1	1	1	0	0	0	0	3
[Kautz et al. 2017]	1	1	1	1	1	1	1	1	8
[de Smet et al. 2017]	1	1	0	1	0	1	1	1	6
[Adam 2016]	0	1	1	0	0	0	0	0	2
[O'Donoghue et al. 2015]	0	1	1	1	0	1	0	1	5
[Jelinek et al. 2014]	0	1	1	1	1	1	1	0	6
Total	9	24	21	23	17	20	16	11	

3. Resultados: Análise e Discussão

Para melhorar o desempenho de atletas usando a análise de dados pode-se utilizar diversas abordagens, dependendo dos objetivos e do esporte, e existem diversos trabalhos relacionados ao tema. Da adaptação dos modelos propostos por [Kitchenham and Charters 2007] e [Budgen and Brereton 2006] a apresentação de resultado e análise serão efetuados de forma conjunta pelos principais tópicos verificados du-

rante a leitura dos artigos. Tendo como referência a pergunta que norteia essa pesquisa, será dado ênfase no tema *Sport Analytics* utilizando os seguinte grupos de análise: objetivo do trabalho, esporte utilizado como base e algoritmos de IA utilizados.

3.1. Por objetivo de trabalho

Durante a fase de leitura, foi realizado uma análise crítica do objetivo principal de cada trabalho escolhido. Nos casos em que o trabalho tivesse mais de um objetivo, foi proposto focar no objetivo principal, segundo o posicionamento dos autores e interpretação após a leitura. Foi dividido em três grupos principais: foco no atleta, foco na equipe e previsão de resultados. A quantidade de artigos por grupo está descrita na tabela 4.

Tabela 4. Distribuição dos objetivos do estudo em todos os artigos analisados

Objetivo do Estudo	Quantidade de Artigos
Foco no atleta	7
Foco na equipe	8
Previsão de resultados	9

3.1.1. Foco no atleta

Nesta categoria foram classificados trabalhos que, mesmo não tendo atletas como objeto central de estudo, tinham como objetivos principais:

- melhorar as condições físicas dos atletas;
- prevenir lesões por sobrecarga de esforço;
- aumentar o desempenho, seja por modelos de treinamento ou aplicação de novas táticas em competições.

Quantitativamente foram encontrados cinco artigos que estavam relacionados aos dois primeiros itens e dois relacionados à aplicação de estratégias.

Para prevenção de lesões, destacamos a ubiquidade dos aparelhos de IoT com a capacidade de coleta de dados necessários para a análise de carga de trabalho de cada atleta (batimento cardíaco, sensores de força, acelerômetro, entre outros) e sua facilidade de transmissão, que possibilita o surgimento de muitas aplicações na área esportiva e de saúde [Moatamed et al. 2017].

O uso de dados coletados para medição de fadiga através da obtenção do *Rating of Perceived Exertion* (RPE) e sua utilização como fundamentação de seus trabalhos foi usado por [Op De Beéck et al. 2018], [Naglah et al. 2018], [Moatamed et al. 2017] e [Rossi et al. 2017], demonstrando a relevância desta abordagem pela quantidade de trabalhos selecionados. Enquanto [Kautz et al. 2017] baseou-se na coleta de movimentos através de vídeos, e baseado na repetição de movimentos prever a ocorrência de lesões.

Estratégia foi outro enfoque dados nos trabalhos e usado por [de Leeuw et al. 2018] para apoio aos corredores, neste caso para distâncias de 10 km, meia maratona e maratona, estudando o impacto do *pacing* desenvolvido durante as competições. Já [Altini and Amft 2018], também em um trabalho direcionados aos corredores, deu destaque para dados obtido através de IoT, como dados corporais, volume e estilo de treino, e propôs um modelo para verificar o impacto do volume de treino no resultado dos atletas.

3.1.2. Foco na equipe

Diversas possibilidades foram destacadas por [Pariath et al. 2018], [Ruiz et al. 2017] e [Kempe et al. 2018] no uso de *Sport Analytics*: auxiliar equipes, treinadores, gerentes de esporte na contratação de atletas; verificação de qualidades e defeitos de equipes adversárias direcionando tática de treinamentos e atuações durante as competições; validar se um determinado atleta corresponde ao salário pago ou verificar qual atleta deveria ser contratado para aumentar a qualidade e eficácia do time.

Já os autores [Jelinek et al. 2014], [Nagarajan and Li 2018] e [Pariath et al. 2018] focam na capacidade de cada atleta de contribuir para o desempenho do time, como acerto de passes no futebol ou quantidades de assistências e rebotes no caso do basquete, enquanto [Ruiz et al. 2017] e [Kempe et al. 2018] estudam qual o desempenho geral das equipes.

Os trabalhos de [Anand et al. 2018] e [Kasiri-Bidhendi et al. 2015] estudam esportes individuais, como tênis (classificação de jogadas) e boxe (identificação de golpes) respectivamente, porém focados no esporte em si, e não na melhoria ou aperfeiçoamento do atleta como objetivo principal.

3.1.3. Previsão de resultados

Modelos para prever diretamente resultados foram estudados por [Fernandez et al. 2016], [Gu et al. 2019], [Young et al. 2019] e [Dubbs 2018] levando em consideração o treino das equipes (quantidade, modelos), bem como do desempenho individual do atleta. Uma variação proposta [Hamdad et al. 2018] foi avaliar o desempenho das equipes com base nos salários de cada jogador.

Já [de Smet et al. 2017] utilizou dados de corridas realizadas para propor uma metodologia que auxilia os atletas e organizadores de eventos a prever os resultados com base em eventos passados, levando em consideração dados geográficos de treinamento e das provas.

No futebol, [Aoki et al. 2017] enfocou o quanto o resultado das partidas podem ser devido à qualidade das equipes ou fatores de sorte. Por sua vez [Adam 2016] desenvolveu um modelo para previsão da quantidade de gols em partidas de futebol. Um trabalho bastante interessante foi desenvolvido por [O'Donoghue et al. 2015] usando dados do futebol gaélico para detecção de anomalias em dados esportivos.

3.2. Por esporte

A quantidade de trabalhos pelos esporte mais estudados pode ser visualizada na figura 1, onde deve-se considerar que alguns estudos focaram em mais de um esporte, como por exemplo [Aoki et al. 2017] e [Dubbs 2018].

Os esportes coletivos são os que tiveram a maior quantidade de estudos selecionados, representando um total de 17 trabalhos, ou aproximadamente 71%, enquanto os esportes individuais representaram aproximadamente 29%, ou 7 trabalhos. Nominalmente, o futebol teve a maior participação dentre os esportes com um total de dez estudos selecionados, demonstrado na figura 1.

Para [Aoki et al. 2017] vários fatores favorecem a utilização de esportes que tenham confederações ou ligas bem definidas: primeiramente porque ligas formam sistemas

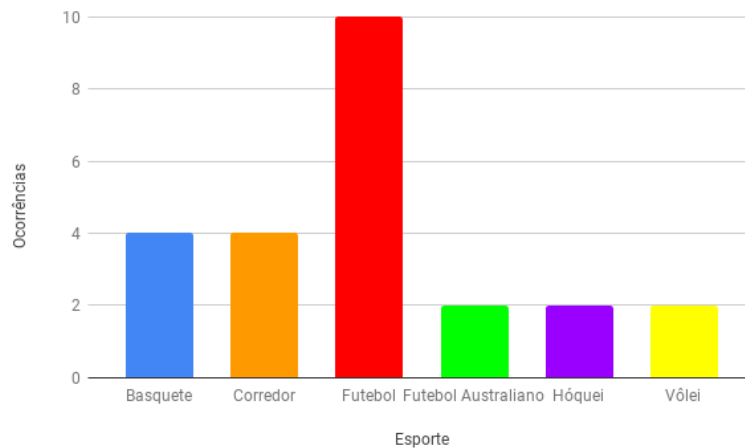


Figura 1. Quantidade de estudos por esporte

relativamente isolados; grande quantidade de dados disponíveis e grande popularidade que alimenta um negócio multimilionário.

Para o futebol foi verificado que os objetivos estavam equilibrados entre Previsão de Resultados estudados por [Aoki et al. 2017], [Fernandez et al. 2016], [Dubbs 2018] e [Adam 2016] e Foco na Equipe [Ruiz et al. 2017], [Kempe et al. 2018], [Cintia et al. 2015] e [Pariath et al. 2018], enquanto apenas dois tinham maior interesse em melhorias individuais dos atletas [Naglah et al. 2018] e [Rossi et al. 2017] corroborando o citado por [Aoki et al. 2017] como sendo um esporte com grande foco financeiro.

O esporte individual com maior destaque foi a corrida, com 4 estudos, sendo a maioria destes trabalhos focados nos atletas [Op De Beéck et al. 2018], [Altini and Amft 2018] e [de Leeuw et al. 2018] com prevenção de lesões e estratégias de corridas e um trabalho de [de Smet et al. 2017] focado em previsão de resultados.

3.3. Por algoritmo

Sport Analytics está associado ao uso intensivo de algoritmos de IA para tratamento de grandes bases de dados, sendo assim são duas as pré condições para o uso de *Sport Analytics* como conhecemos hoje: dados e uso de algoritmos de IA. Nos trabalhos analisados o maior destaque está para a utilização de algoritmos como Regressão Linear em 11 estudos, SVM em 8 e seguidos de KNN e k-means com 5 utilizações.

Os três esportes que mais tiveram variação na quantidade de algoritmos de IA e métodos estatísticos utilizados foram Basquete, Corredores e Futebol como mostrado na tabela 5.

4. Conclusão

O principal objetivo desse trabalho foi fazer uma pesquisa e análise em área relacionadas com a desempenho de atletas e uso de dados no esporte, entendendo como isso impacta os atletas e sua adoção pelas equipes.

Para atingir esta objetivo foi utilizada uma revisão sistemática que abrangeu inicialmente mais de 4.000 artigos, que com a aplicação de um protocolo explanado no

Tabela 5. Distribuição de algoritmos de IA/métodos estatísticos

Esporte	Algoritmo	Quantidade
Basquete	Bayesian	1
	K-means	1
	Linear Regression	2
	Logistical Regression	1
	Monte Carlo	1
	Naive Bayes	1
	Regression	1
Total - Basquete		8
Corredor	Gradient Boosted Regression Trees	1
	K-means	1
	Linear Regression	2
	MLR	1
	Neural Network	1
	Regression	1
	Subgroup Discovery	1
Total - Corredor		8
Futebol	Bayesian	1
	Decision Tree	1
	Generalized Additive Models	1
	K-means	1
	KNN	3
	Linear Regression	5
	Logistical Regression	1
	Monte Carlo	1
	Neural Network	1
	PCA	1
	Random Forest	4
	Regression	1
SVM	3	
Total - Futebol		24

corpo do artigo e revisão por pares chegamos a um total de 24 artigos para discutirmos tópicos relacionados com prevenção de lesões, desempenho de atletas, entre outros e entendimento de como algoritmos de IA podem ser usados em *Sport Analytics*.

Durante a análise ficou evidente que o futebol é o esporte onde o uso de *Sport Analytics* é mais utilizado, mas com destaque também para esportes como basquete e corredores. Já prevenção de lesões, como objetivo principal dos estudos, é um dos itens que merece destaque com cinco artigos dedicados a este tema.

Como trabalhos futuros, a intenção é conduzir estudo focado na utilização de *Sport Analytics* para corredores amadores e outros esportes individuais que não sejam enquadrados em ligas bem definidas de forma a difundir novas técnicas que possam trazer melhoria na capacidade esportivas desses atletas.

Para o futebol, a maioria dos trabalhos concentrou-se em campeonatos de pontos corridos, onde a técnica das equipes sobrepõe-se à fatores de sorte. Trabalhos destinados a prever resultados e estratégias em campeonatos disputados em partidas eliminatórias, como por exemplo a Libertadores da América e Liga dos Campeões, precisam ser elaborados para preencher essa lacuna. O mesmo se aplica para campeonatos de basquete, vôlei e hóquei, onde temos carência de trabalhos acadêmicos realizados onde fatores aleatórios podem sobrepor a preparação das equipes.

Referências

- Adam, A. (2016). Generalised linear model for football matches prediction. *CEUR Workshop Proceedings*, 1842.
- Altini, M. and Amft, O. (2018). Estimating Running Performance Combining Non-invasive Physiological Measurements and Training Patterns in Free-Living. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018-July:2845–2848.
- Anand, A., Sharma, M., Srivastava, R., Kaligounder, L., and Prakash, D. (2018). Wearable motion sensor based analysis of swing sports. *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, 2018-January:261–267.
- Aoki, R. Y. S., Assuncao, R. M., and de Melo, P. O. S. (2017). Luck is Hard to Beat: The Difficulty of Sports Prediction. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17*, pages 1367–1376, New York, NY, USA. ACM.
- Budgen, D. and Brereton, P. (2006). Performing systematic literature reviews in software engineering. In *Proceeding of the 28th international conference on Software engineering - ICSE '06, ICSE '06*, page 1051, New York, New York, USA. ACM Press.
- Cintia, P., Giannotti, F., Pappalardo, L., Pedreschi, D., and Malvaldi, M. (2015). The harsh rule of the goals: Data-driven performance indicators for football teams. *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015*.
- Cokins, B. G., Degrange, W., Chambal, S., and Walker, R. (2017). Sports analytics taxonomy, V1.0. *ORMS-Today, informs*, 43(3):1–9.
- Costa, I., Pires, C., and Marinho, L. (2017). *Capítulo 2 Sports Analytics: Mudando o Jogo*, chapter 2, pages 30–62. Sociedade Brasileira de Computação - SBC.
- de Leeuw, A.-W., Meerhoff, L. A., and Knobbe, A. (2018). Effects of Pacing Properties on Performance in Long-Distance Running. *Big Data*, 6(4):248–261.
- de Smet, D., Verleysen, M., and Francaux, M. (2017). Running Race Times Prediction and Runner Performances Comparison using a Matrix Factorization Approach. In *Proceedings of the 5th International Congress on Sport Sciences Research and Technology Support*, number icSPORTS, pages 96–101. SCITEPRESS - Science and Technology Publications.
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12):64–73.

- Dubbs, A. (2018). Statistics-free sports prediction. *Model Assisted Statistics and Applications*, 13(2):173–181.
- Fernandez, J., Medina, D., Gomez, A., Arias, M., and Gavaldà, R. (2016). From Training to Match Performance: A Predictive and Explanatory Study on Novel Tracking Data. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, number October, pages 136–143. IEEE.
- Gu, W., Foster, K., Shang, J., and Wei, L. (2019). A game-predicting expert system using big data and machine learning. *Expert Systems with Applications*, 130:293–305.
- Hamdad, L., Benatchba, K., Belkham, F., and Cherairi, N. (2018). Basketball Analytics. *Data Mining for Acquiring Performances*. pages 13–24. Springer, Cham.
- Hayashi, C. (1998). What is Data Science ? Fundamental Concepts and a Heuristic Example. In *What is Data Science ? Fundamental Concepts and a Heuristic Example*, pages 40–51. Springer, Tokyo.
- Jelinek, H. F., Kelarev, A., Robinson, D. J., Stranieri, A., and Cornforth, D. J. (2014). Using meta-regression data mining to improve predictions of performance based on heart rate dynamics for Australian football. *Applied Soft Computing Journal*, 14(PART A):81–87.
- Kasiri-Bidhendi, S., Fookes, C., Morgan, S., Martin, D. T., and Sridharan, S. (2015). Combat sports analytics: Boxing punch classification using overhead depthimager. *Proceedings - International Conference on Image Processing, ICIP*, 2015-December:4545–4549.
- Kautz, T., Groh, B. H., Hannink, J., Jensen, U., Strubberg, H., and Eskofier, B. M. (2017). Activity recognition in beach volleyball using a Deep Convolutional Neural Network: Leveraging the potential of Deep Learning in sports. *Data Mining and Knowledge Discovery*, 31(6):1678–1705.
- Kempe, M., Goes, F. R., and Lemmink, K. A. (2018). Smart data scouting in professional soccer: Evaluating passing performance based on position tracking data. *Proceedings - IEEE 14th International Conference on eScience, e-Science 2018*, pages 409–410.
- Kitchenham, B. and Charters, S. (2007). issue: EBSE 2007-001. *Technical report*, 2(3).
- Moatamed, B., Darabi, S., Gwak, M., Kachuee, M., Metoyer, C., Linn, M., and Sarrafzadeh, M. (2017). Sport analytics platform for athletic readiness assessment. *2017 IEEE Healthcare Innovations and Point of Care Technologies, HI-POCT 2017*, 2017-December:156–159.
- Nagarajan, R. and Li, L. (2018). Optimizing NBA player selection strategies based on salary and statistics analysis. *Proceedings - 2017 IEEE 15th International Conference on Dependable*, 2018-January:1076–1083.
- Naglah, A., Khalifa, F., Mahmoud, A., Ghazal, M., Jones, P., Murray, T., Elmaghraby, A. S., and El-baz, A. (2018). Athlete-Customized Injury Prediction using Training Load Statistical Records and Machine Learning. In *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, number d, pages 459–464. IEEE.

- O'Donoghue, J., Roantree, M., Cullen, B., Moyna, N., Sullivan, C. O., and McCarren, A. (2015). Anomaly and event detection for unsupervised athlete performance data. *CEUR Workshop Proceedings*, 1458(October):205–217.
- Op De Beéck, T., Meert, W., Schütte, K., Vanwanseele, B., and Davis, J. (2018). Fatigue Prediction in Outdoor Runners Via Machine Learning and Sensor Fusion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, pages 606–615, New York, NY, USA. ACM.
- Pariath, R., Shah, S., Surve, A., and Mittal, J. (2018). Player Performance Prediction in Football Game. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, number Iceca, pages 1148–1153. IEEE.
- Rossi, A., Perri, E., Trecroci, A., Savino, M., Alberti, G., and Iaia, F. M. (2017). GPS data reflect players' internal load in soccer. *IEEE International Conference on Data Mining Workshops, ICDMW*, 2017-November:890–893.
- Ruiz, H., Power, P., Wei, X., and Lucey, P. (2017). "The Leicester City Fairytale?": Utilizing New Soccer Analytics Tools to Compare Performance in the 15/16 & 16/17 EPL Seasons. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1991–2000.
- UE (2018). *Study on the Economic Impact of Sport through Sport Satellite Accounts*. Number April.
- Value, I. and Team, E. (2015). Sports Analytics & Risk monitoring based on Hana Platform. *International SoC Design Conference (ISOCC)*, pages 221–222.
- Young, C. M., Luo, W., Gastin, P., Tran, J., and Dwyer, D. B. (2019). The relationship between match performance indicators and outcome in Australian Football. *Journal of Science and Medicine in Sport*, 22(4):467–471.