

## TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL UTILIZADAS PARA ANÁLISE DA APTIDÃO CARDIORRESPIRATÓRIA: UMA REVISÃO

Rodrigo Cunha de Mello Pedreiro<sup>1,2,3</sup>, Fernando Torres Coimbra de Sá Balbina<sup>1</sup>, Fernando José Habib Moraes<sup>1</sup>

**Resumo:** A aptidão cardiorrespiratória (ACR) pode ser considerada um dos sinalizadores de saúde, estando diretamente ligada a capacidade respiratória e cardiovascular. Este parâmetro permite o estudo do impacto causado por fatores de risco como: obesidade, hipertensão arterial, tabagismo e dislipidemias. Portanto, o processamento rápido e eficiente dos dados provenientes do ACR é de grande importância para a área médica. Ferramentas modernas para análise de dados têm sido desenvolvidas, a fim de caracterizar de forma eficiente conjuntos de dados. Com a ajuda de códigos computacionais que tem como base a teoria do aprendizado automático ou aprendizado computacional, o próprio programa, através do reconhecimento de padrões, desenvolve a capacidade de "aprender". É o caso das técnicas que envolvem "Machine Learning", onde algoritmos complexos avaliam diversas variáveis, utilizando suas próprias respostas para melhora de desempenho. Utilizando os critérios e recomendações dos Itens de Relatórios Preferenciais para Revisões Sistemáticas, o objetivo desse trabalho é avaliar as técnicas de inteligência artificial, denominadas "Machine Learning", que são utilizadas para análises da ACR, revisando os artigos da literatura encontrada no banco de dados PubMed/Medline.

**Palavras-chave:** VO<sub>2</sub>, Machine Learning, consumo de oxigênio

## ARTIFICIAL INTELLIGENCE TECHNIQUES USED FOR CARDIORRESPIRATORY FITNESS ANALYSIS: A REVIEW

**Abstract:** Cardiorespiratory fitness (ACR) can be considered one of the health indicators, being directly linked to respiratory and cardiovascular capacity. This parameter allows the study of the impact caused by risk factors such as obesity, hypertension, smoking and dyslipidemia. Therefore, fast and efficient processing of data from the ACR is important to the medical field. Modern tools for data analysis have been developed to efficiently characterize data sets. With the help of computer codes based on the theory of automatic learning or computer learning, the program itself, through pattern recognition, develops the ability to "learn". This is the case with techniques that involve "Machine Learning", where complex algorithms evaluate several variables, using their own responses to improve performance.

<sup>1</sup>Doutorando, Universidade Anhembi Morumbi, UAM **contato:** rodrigocmp1@gmail.com

<sup>2</sup>Professor, Universidade Estácio de Sá, UNESA

<sup>3</sup>Professor, Faculdade Santo Antônio de Pádua, FASAP

Using the criteria and recommendations of the Preferred Reporting Items for Systematic Reviews, the objective of this paper is to evaluate artificial intelligence techniques, called "Machine Learning", which are used for ACR analysis, reviewing the literature articles found in PubMed / Medline database.

Keys words: VO<sub>2</sub>, Machine Learning, oxygen consumption

## INTRODUÇÃO

Já é bastante evidenciada na literatura a importância da aptidão cardiorrespiratória (ACR) para a saúde, ela é um excelente indicador da capacidade dos sistemas cardiovascular e respiratório de fornecer o oxigênio, assim como a utilização desse pelos músculos durante uma atividade física. (BOUCHARD, SHEPHARD & STEPHENS, 1994; LAUKKANEN, KURL & SALONEN, 2002).

Baixos níveis da ACR estão diretamente associados com o maior risco de desenvolver doenças cardiovasculares e de morrer prematuramente independentemente de estar associado com outros fatores de riscos mais comuns (BLAIR, *et al.* 1989; BLAIR, *et al.* 1989; STAMATAKIS, *et al.* 2013; LAUKKANEN, *et al.* 2007; KOKKINOS, *et al.* 2008; NES, *et al.* 2014; LAUKKANEN, *et al.* 2004; KODAMA, *et al.* 2009). Além disso, a ACR mostra-se melhor preditor de risco do que quando comparada com outros fatores de risco mais frequentemente estudados, tais como: a obesidade, hipertensão arterial, tabagismo e dislipidemias (NES, *et al.* 2014; CHURCH, *et al.* 2004; KATZMARZYK, CHURCH & BLAIR, 2004). Utilizar a tecnologia e procurar melhores estratégias para avaliar esse importante indicador de saúde é de extrema importância.

Já se sabe que a união da matemática e a ciência computacional podem contribuir para análise dados relacionados à saúde, com isso, podemos citar a *Machine Learning* (ML), que é uma técnica moderna de análise de dados e tem uma capacidade única de aprender e melhorar seu rendimento sem ser programada e sem a necessidade de algum tipo de instrução humana. O principal objetivo dos algoritmos de classificação de ML supervisionados é resumidamente explicar a variável dependente em termos das variáveis independentes (AGGARWAL, 2014; SAKR, *et al.* 2017).

Os algoritmos são ajustados com base na amostra de treinamento e no sinal de erro. Em geral, as técnicas estatísticas convencionais geralmente dependem do processo de teste de hipóteses. Este processo é muito orientado pelo usuário onde o usuário especifica variáveis, forma funcional e tipo de interação. Portanto, a intervenção do usuário pode influenciar os modelos resultantes. Com as técnicas de ML, a principal hipótese é que existe um padrão (e não uma associação) no conjunto de variáveis preditoras que identificarão o resultado. Podendo se tornar uma excelente estratégia de análise (SAKR, et al. 2017). Com isso, se faz necessário buscar na literatura quais são as técnicas mais utilizadas tentando avaliar a ACR e quais têm os melhores desfechos.

O objetivo primário do presente estudo foi avaliar através de uma revisão as técnicas de inteligência artificial, denominadas de *Machine Learning* que mais são utilizadas para análises da ACR, e o objetivo secundário foi avaliar quais são as melhores técnicas a serem utilizadas para avaliar estudos na área da saúde utilizando a ACR como principal variável de estudo.

## **METODOLOGIA**

O método desta pesquisa foi elaborado e executado utilizando os critérios e recomendações dos Itens de Relatórios Preferenciais para Revisões Sistemáticas e Meta-análises (PRISMA) (MOHER, et al., 2009) e o Manual Cochrane para Revisões Sistemáticas de Intervenções (HIGGINS & GREEN, 2011).

### **Informações de coleta**

Foi realizada uma revisão da literatura no banco de dados PubMed/Medline utilizando as seguintes palavras-chave e combinações: “*volume of oxygen*”, “*cardiorespiratory fitness*”, “*vo2*” com “*artificial intelligence*”, “*machine learning*” conduzida entre dezembro de 2018 e fevereiro de 2019. Na pesquisa não foram aplicados filtros.

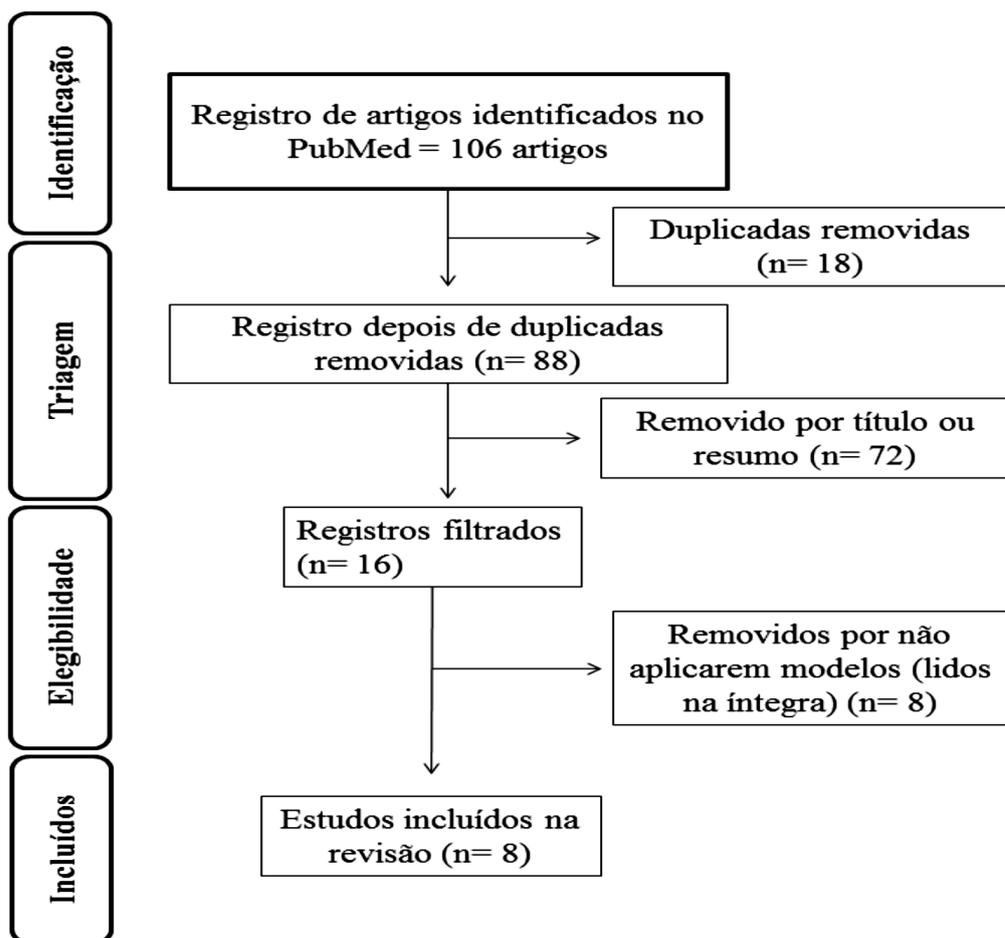
**Estratégia de pesquisa**

Não foram selecionados para análise estudos de caso e estudos de revisões, além disso, foram selecionadas apenas pesquisas dos últimos 10 anos (2009 até 2019).

**RESULTADOS**

O resultado total identificado na busca foi de 106 artigos no PubMed/Medline. Após o processo de remoção de duplicatas (n= 18), permaneceu um total de 88 artigos. 72 artigos foram removidos após a leitura dos seus títulos e/ou resumos, ficando 16 artigos, e estes foram lidos na íntegra, onde tivemos a exclusão de mais 8, restando 8 artigos no final, que foram incluídos na revisão (Figura 1)

**Estratégia de pesquisa.**



**Figura 1-** Seleção dos estudos

Os objetivos dos estudos selecionados, assim como o tamanho da amostra, os testes utilizados para a avaliação da ACR (VO<sub>2</sub>), as principais técnicas de ML e seus melhores resultados encontrados na aplicação destes, estão descritas na Tabela 1.

Tabela 1- Características dos estudos, principais ML utilizados e melhores técnicas utilizadas.

Referências	N	Objetivo do estudo	Teste de ACR	Modelos de análise	Melhores resultados
Alghamdi et al. (2017)	325 55	Avaliar o desempenho relativo de vários ML para prever diabetes utilizando informações de ACR	Teste de esteira, utilizando protocolo padrão de Bruce	Decision tree; Naive Bayes; Logistic regression; Logistic model tree; Random forest	Random forest e Naive Bayes
Al Mallah et al. (2017)	342 12	Comparar 10 anos de mortalidade para todas as causas usando regressão logística e ML através de dados de ACR	Teste de esteira, utilizando protocolo padrão de Bruce	K-nearest neighbor; Statical logistic regression	K-nearest neighbor
Dumond et al. (2017)	14	Melhorar a acurácia das estimativas de volumes respiratórios através de PRM e facilitar o uso dessa técnica	Testes com magnetômetros em repouso, em pé, no computador, na esteira e cicloergômetro	Multiple linear regression; Artificial neural network	Multiple linear regression
John et al. (2013)	277	Comparar técnicas de modelagens simples e complexa para estimar a classificação ventilatória utilizando ActiGraphTM	Testes de caminhada e corrida em esteira, esportes, atividades domésticas e ocupacionais	Multiple regression and activity count cut-point analyses; Random forest	Multiple regression e Random forest
Kolus et al. (2014)	35	Avaliar dois modelos para análise do VO2 partindo da FC	Step test segundo Meyer e Flenghi (1995)	Analytical Method; Neuro-fuzzy systems	Neuro-fuzzy systems
Nakamura et al. (2016)	23	Estimar o VO2 usando sensores de movimentos (6) em diferentes intensidades	Teste em esteira ergométrica	Neural network; Linear regression	Neural network
Sakr et al. (2017)	342 12	Apresentar uma avaliação e comparação de como as técnicas de ML podem ser aplicadas em estudos médicos utilizando a ACR	Teste de esteira, utilizando protocolo padrão de Bruce	Decision tree; Support vector machine; Artificial neural network; Naive Bayesian classifier; Bayesian network; K-nearest neighbor; Random forest	Bayesian network, Naive Bayesian classifier, Decision tree e principalmente Random forest
Sakr et al. (2018)	230 95	Avaliar e comparar o desempenho de diferentes técnicas de ML na predição de hipertensão a partir de dados de ACR	Teste de esteira, utilizando protocolo padrão de Bruce	LogitBoost; Naive Bayesian classifier; Locally weighted naïve bayes; Artificial neural network; Support vector machine; Random tree forest	Random tree fores

## DISCUSSÕES

De uma maneira geral, conseguimos identificar bons resultados utilizando ML como forma de análise de dados em saúde, principalmente quando a variável em estudo é a ACR. Dentre as técnicas mais utilizadas, podemos perceber se destacaram por sua qualidade das seguintes ML: Naive Bayes (ALGHAMDI et al. 2017), K-nearest neighbor (AL MALLAH et al. 2017), Multiple linear regression (DUMOND et al. 2017), Multiple regression (JOHN et al. 2013), Neuro-fuzzy systems (KOLUS et al. 2014), Neural network (NAKAMURA et al. 2016), Bayesian network, Naive Bayesian classifier, Decision tree (SAKR et al. 2017) e principalmente Random forest (ALGHAMDI et al. 2017; JOHN et al. 2013; SAKR et al. 2017; SAKR et al. 2018) que são uma combinação de preditores de “árvores”, de tal forma que cada árvore (variável) depende dos valores de um vetor aleatório amostrado de forma independente e com a mesma distribuição para todas as árvores (variáveis) na floresta (no conjunto de dados) (BREIMAN, 2001).

A classificação de dados é a base do ML. Através da ciência de dados, várias ferramentas foram criadas a fim de classificar dados, e o algoritmo de Floresta Aleatória (Random Forest) se mostrou estar entre as melhores (VERIKAS et al., 2011). Para entender o porquê disso, é necessário entender o fundamento desse algoritmo.

Árvores de Decisão constituem a Floresta Aleatória. Cada árvore, baseada em suas regras de decisão "se então", tem seus nodos divididos em "galhos", que novamente se subdividem, baseados em novas regras de decisão "se então", gerando novos galhos. Sendo assim, cada regra está classificando os dados baseada em uma característica, criando classes de dados. O objetivo então é encontrar uma Árvore de Decisão capaz de classificar os dados de forma que, no fim, existam classes de dados na mais variada forma possível, ou seja, segregados de maneira mais rigorosa aceitável pelas suas características (CRIMINISI et al., 2012; DE'ATH & FABRICIUS, 2000).

A Floresta Aleatória alcança esse objetivo utilizando a resposta de várias árvores de decisão, fundamentada no conceito de que "um grupo de modelos não relacionados obterá resultados melhores do que modelos individuais". Cada árvore tem as características utilizadas para análise dos dados aleatorizadas (Randomizing), assim como os dados que elas

estão analisando (Bagging) e, as próprias árvores são reajustadas quanto aos pesos das características (Boosting). Portanto, cada "árvore" integrante da "floresta" gera um resultado, e o melhor resultado obtido servirá de modelo para a próxima iteração.

Os aspectos abordados conferem a este algoritmo grande poder de classificação, fato evidenciado pela maior ocorrência observada nos artigos que foram objeto desse estudo.

### Limitações do estudo

Este estudo se tornou limitado por não ter acesso a outros bancos de dados como: Web of Science e Scopus. O problema do baixo número de artigos selecionados poderia ser resolvido com acesso dessas plataformas. Mais trabalhos tentando investigar a aplicação de ML na saúde principalmente avaliando a variável ACR se fazem necessário.

### CONCLUSÃO:

Apesar já do amplo conhecimento sobre as ciências da computação, ainda existem poucos estudos que utilizam ML como técnica para análise de dados em saúde, esta técnica pode trazer resultados mais reais (com menos interferência humana) de estudos que utilizam e investigam a ACR. *Random forest* aparece inicialmente como a melhor técnica de ML para análise desta variável.

### REFERÊNCIAS

AGGARWAL CC. **Data classification: algorithms and applications**. CRC Press; 2014. <https://www.crcpress.com/Data-Classification-Algorithms-and-Applications/Aggarwal/p/book/9781466586741>.

ALGHAMDI M, AL-MALLAH M, KETEYIAN S, BRAWNER C, EHRMAN J, SAKR, S (2017) **Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) project**. PLoS ONE 12(7):e0179805.

AL-MALLAH et al. **Using Machine Learning to Define the Association between Cardiorespiratory Fitness and All-Cause Mortality (from the Henry Ford Exercise Testing Project)**. Am J Cardiol 2017;120:2078–2084

BLAIR SN, KOHL HW 3RD, PAFFENBARGER RS JR, CLARK DG, COOPER KH, GIBBONS LW. **Physical fitness and all-cause mortality: a prospective study of healthy men and women**. JAMA. 1989; 262:2395–401.

BLAIR SN, KAMPERT J, KOHL HW 3rd, et al. **Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women**. JAMA. 1996; 276:205–10.

BOUCHARD C, SHEPHARD RJ, STEPHENS T. **Physical activity, fitness, and health: International proceedings and consensus statement**. Champaign: Human Kinetics; 1994.

BREIMAN, L. **Machine Learning** (2001) 45: 5. Springer.

CHURCH TS, CHENG YJ, EARNEST CP, et al. **Exercise capacity and body composition as predictors of mortality among men with diabetes**. Diabetes Care. 2004;27:83–8.

CHURCH T, KAMPERT JB, GIBBONS LW, BARLOW CE, BLAIR SN. **Usefulness of cardiorespiratory fitness as a predictor of all-cause and cardiovascular disease mortality in men with systemic hypertension**. Am J Cardiol. 2001;88(6):651–6.

CRIMINISI, A., SHOTTON, J., KONUKOGLU, E. **Foundations and Trends R in Computer Graphics and Vision**. Vol. 7, Nos. 2–3 (2011) 81–227 c 2012.

DE'ATH, G, FABRICIUS, K. E. **Classification and regression trees**. Ecology, Vol. 81, No. 11 (2000).

DUMOND et al. **Estimation of respiratory volume from thoracoabdominal breathing distances: comparison of two models of machine learning**. Eur J Appl Physiol. Online 2017

GREEN J. P. S. **Cochrane handbook for systematic reviews of interventions**: John Wiley & Sons; 2011.

JOHN et al. **Simple to complex modeling of breathing volume using a motion Sensor**. Sci Total Environ. Author manuscript; available in PMC 2014 June 01.

KATZMARZYK PT, Church TS, Blair SN. **Cardiorespiratory fitness attenuates the effects of the metabolic syndrome on all-cause and cardiovascular disease mortality in men**. Arch Intern Med. 2004; 164(10):1092–7.

KODAMA S, SAITO K, TANAKA S, et al. **Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis.** JAMA. 2009; 301(19):2024–35.

KOKKINOS P, MYERS J, KOKKINOS JP, PITTARAS A, NARAYAN P, MANOLIS A, KARASIK P, GREENBERG M, PAPADEMETRIOU V, SINGH S. **Exercise capacity and mortality in black and white men.** Circulation 2008;117:614–622.

KOLUS et al. **Estimating oxygen consumption from heart rate using adaptive neuro-fuzzy inference system and analytical approaches.** Applied Ergonomics xxx (2014) 1e9

LAUKKANEN JA, KURL S, SALONEN JT. **Cardiorespiratory fitness and physical activity as risk predictors of future atherosclerotic cardiovascular diseases.** Curr Atheroscler Rep 2002;4:468–476.

LAUKKANEN JA, RAURAMAA R, SALONEN JT, KURL S. **The predictive value of cardiorespiratory fitness combined with coronary risk evaluation and the risk of cardiovascular and all-cause death.** J Intern Med 2007;262:263–272.

LAUKKANEN JA, KURL S, SALONEN R, RAURAMAA R, SALONEN JT. **The predictive value of cardiorespiratory fitness for cardiovascular events in men with various risk profiles: a prospective population based cohort study.** Eur Heart J. 2004;25(16):1428–37.

MOHER D, LIBERATI A, TETZLAFF J, ALTMAN DG, Group P. **Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.** Journal of clinical epidemiology. 2009; 62(10):1006–12. <https://doi.org/10.1016/j.jclinepi.2009.06.005> PMID: 19631508. 29. Higgins

NAKAMURA et al. **Applying Neural Network to VO<sub>2</sub> Estimation using 6-axis Motion Sensing Data.** 978-1-4577-0220-4/16/\$31.00 ©2016 IEEE

NES, B. M., L. J. VATTEN, J. NAUMAN, I. JANSZKY, and U. WISLKFF. **A Simple Nonexercise Model of Cardiorespiratory Fitness Predicts Long-Term Mortality.** Med. Sci. Sports Exerc., Vol. 46, No. 6, pp. 1159–1165, 2014.

STAMATAKIS E, HAMER M, O'DONOVAN G, BATTY GD, AND KIVIMAKI M. **A non-exercise testing method for estimating cardiorespiratory fitness: associations with all-cause and cardiovascular mortality in a pooled analysis of eight population-based cohorts.** European Heart Journal (2013) 34, 750–758

SAKR et al. **BMC Medical Informatics and Decision Making** (2017) 17:174  
DOI 10.1186/s12911-017-0566-6

Revista Conhecendo Online: Ciências da Saúde

Abril de 2020, v6, n1

ISSN: 2359-5256 (Online)

SAKR S, ELSHAWI R, AHMED A, QURESHI WT, BRAWNER C, KETEYIAN S, et al. (2018) **Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford Exercise Testing (FIT) Project.** PLoS ONE 13(4): e0195344.

VERIKAS, A., GELZINIS, A., BACAUSKIENE, M. **Mining data with random forests: A survey and results of new tests.** Pattern Recognition 44 (2011) 330–349.