Original Article

Machine learning in sports medicine: A new approach in human exercise

ANDREA SCHULZ GALVÃO[1], MARCELLE KARYELLE MONTALVÃO GOMES[1], NAIRANA CRISTINA SANTOS FREITAS[1], LEONARDO SANTOS MACEDO[1], DANIEL MANCINI OLIVEIRA[1], MÁRCIO VINICIUS ABREU VERLI[1], ROBERTO LOHN NAHON[2], LUIS CARLOS OLIVEIRA GONÇALVES[1,3] ✉, ANÍBAL MONTEIRO MAGALHÃES-NETO[1,3]

[1]Graduate Program in Physical Education (PPGEF). Federal University of Mato Grosso (UFMT). Brazil
[2]Brazilian Society of Orthopedics (SBOT). Brazilian Society of Exercise and Sports Medicine (SBME). Brazil
[3]Graduate Program in Basic and Applied Immunology and Parasitology (PPGIP). Federal University of Mato Grosso (UFMT). Brazil

ABSTRACT

The present study aimed to investigate the possible correlations between the cytokine and adipokine Tumour Necrosis Factor Alpha with parameters of body composition and lipid metabolism in young, high-level athletes after an incremental treadmill test observed in a sample of five individuals, male, high-level running athletes who the difficulty of treating large databases with different individuals, multiple biomarkers, and collection times, in addition to physical parameters and sample characteristics, added to the decrease in new findings induced by the application of statistical tools of univariate analysis, indicate the need to apply exploratory machine learning strategies, generating holistic and integrated analysis of the results. The present study showed a negative correlation between TNF and HDL and a similarity between the same TNF and LDL. These findings do not indicate a cause-and-effect relationship but suggest a possible modulation of the immune system, lipid metabolism, and exercise that requires further investigation.
Keywords: Sport medicine; Data mining; Immunometabolism; Sportomics.

Cite this article as:
Galvão, A. S., Gomes, M. K. M., Freitas, N. C. S., Macedo, L. S., Oliveira, D. M., Verli, M. V. A., Nahon, R. L., Gonçalves, L. C. O., & Magalhães-Neto, A. M. (2023). Machine learning in sports medicine: A new approach in human exercise. *Journal of Human Sport and Exercise, 18*(2), 501-508. https://doi.org/10.14198/jhse.2023.182.19

✉ **Corresponding author.** *Graduate Program in Physical Education (PPGEF). Federal University of Mato Grosso (UFMT). Brazil.*
https://orcid.org/0000-0001-5368-1194
E-mail: luisogoncalves@yahoo.com.br
Submitted for publication December 23, 2022.
Accepted for publication January 30, 2023.
Published April 01, 2023 *(in press* February 10, 2023).
JOURNAL OF HUMAN SPORT & EXERCISE ISSN 1988-5202.

doi:10.14198/jhse.2023.182.19

INTRODUCTION

Advances in technology involving processing, memory, storage, and richness of data, with the rapid dissemination of information over the internet and the formation of large databases, have generated machine learning strategies in different areas, from games such as Poker to video games, marketing, administration and in recent years reaching the field of medical and health sciences (Deo, 2015).

This knowledge quickly reached the field of sports medicine, with contributions in the field of genomics, sportomics, biomechanics, kinesiology, sports training and performance, among others (Cust et al., 2019; Tanisawa et al., 2020; Antink et al., 2021; Gonçalves et al., 2022).

From a multivariate statistical strategy with machine learning, the study's primary objective was to investigate possible relationships between the Tumour Necrosis factor and lipid profile in young running athletes.

MATERIALS AND METHODS

This cross-sectional, descriptive, and observational study was based on a sports strategy with high-level young athletes.

Participants

The studied population comprises high-level track and field athletes, male, from the city of Barra do Garças, Mato Grosso. The convenience sample consisted of five male high-performance track, and field athletes who are members of Barra do Garças – MT, Brazil athletics association.

Experimental design

As an essential warm-up, the athletes performed dynamic and static stretching in all body segments, lasting 5 minutes. Then, using a gas analyser, the treadmill exercise protocol was progressively increased until exhaustion. The warm-up that preceded the test consisted of light running, which lasted 15 minutes. Afterward, the athletes ran for an average distance of 3 km in total, characterized by a gradual load increase in stages until voluntary exhaustion. Thus, the initial speed was 10 km/h with an increment of 1 km every 400 meters.

Cardiovascular assessment during the stress test was performed using an exercise electrocardiogram (ErgoPC Elite) with monitoring at lead points (D1, D2, D3, AVR, AVL, AVF, MC5, V1, V2, V3, V4, V5, and V6). The metabolic assessment was performed using a computerized analyser (MetaLyzer 3B), a face mask (V2 mask Small), and software for capturing and demonstrating data and storing and processing all cardiorespiratory and metabolic variables evaluated.

Statistical analysis

Initially, descriptive statistics were performed on the data, with measurements of position (mean, median, mode, and percentiles) and dispersion (amplitude, variance, standard deviation, and standard error).

Afterward, the univariate analysis of these data was performed using the Shapiro-Wilk normality test (because the sample was smaller than 30 individuals). The equal variance test would be applied if the Shapiro-Wilk test presented a result indicating normal distribution ($p > .05$). For results with $p > .05$, the paired T-Student test would follow; if $p \leq .05$, the paired T-Student test would follow the non-parametric Mann-

Witney test. If the Shapiro-Wilk test presented a result indicating non-normal distribution ($p \leq .05$), the non-parametric Mann-Witney test would be applied directly.

Still, in the phase of the univariate analysis, the analysis of repeated measures ANOVA One Way dependent was performed because they were the same individuals in different conditions and moments.

At this stage, it was observed that using traditional univariate statistics, only one variable showed a statistical difference between collection times (TNF). This led to applying an exploratory machine-learning strategy among the variables.

In this phase, in order to seek a bivariate measure between the data, because the observations contain quantitative values, the Pearson and Spearman correlation tests were applied, with the Spearman correlation being used for a visual analysis using the heat map strategy and the Pearson test as an initial measure for the following machine learning analyses.

As exploratory models of machine learning: CLUSTER - Classical Clustering (Agglomerative Hierarchical Method) and Nearest neighbour (single linkage); ORDINATION – Principal component Analysis (PCA) and Correspondence Analysis (CA).

The minding map with the statistical procedure is shown in Figure 1.

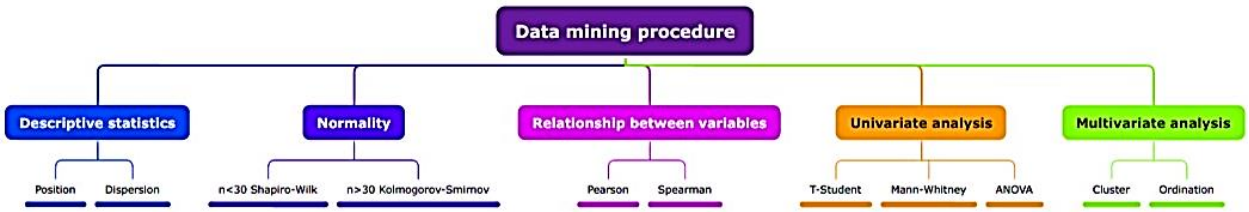

Figure 1. Minding map about statistics.

The Z score was previously applied because the observations contained different measurement units:

$$Z = \frac{Raw\ score - average}{Standard\ deviation}$$

SigmaPlot 14.5 (Academic Perpetual License - Single User – ESD Systat® USA), Past 4.03 (Free version for Windows) were used to carry out the different statistical tests and produce the graphs, and Mindomo premium signature (Expert Software applications SRL – Romania) was used to create mind map.

Ethical approval

This study was conducted according to the guidelines laid down in the Declaration of Helsinki, and the experiment met the requirements of research using human subjects (National Health Council, 2012). This study was approved by the Ethics and Research Committee (number 2,230,073) of the Federal University of Mato Grosso (UFMT) and was registered at clinicaltrials.gov (NCT 03522883). The individuals were informed that they could withdraw from the study at any time. Written informed consent was obtained from each subject, who was instructed on the nature of the research and the procedures involved.

RESULTS

From the calculation of Pearson's correlation coefficient, indicating values between -1 and 1, with a negative value for negative correlations and a positive one for positive correlations, it was possible to observe between correlations with a value of $p < .05$ (represented by squares), negative correlation between TNF and HDL, and positive correlations between the lipogram. components (Figure 2).

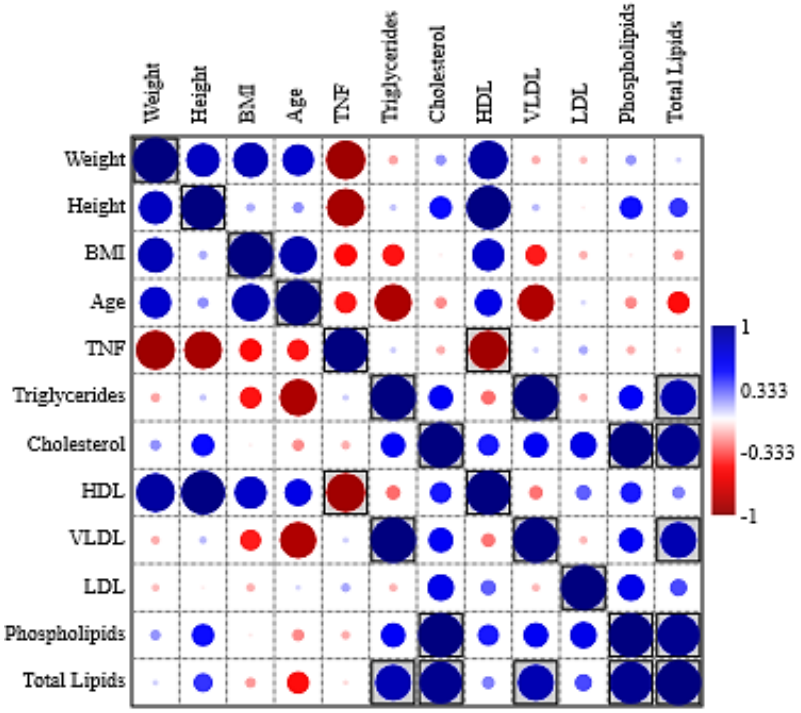

Figure 2. Pearson correlation between variables.

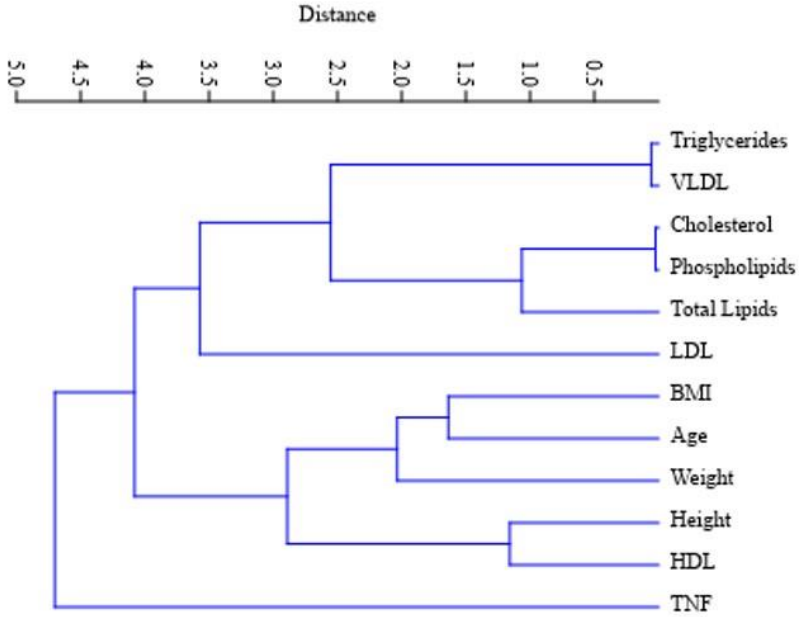

Figure 3. Hierarchical clustering with algorithm single linkage and Euclidean similarity index.

The next step was the application of the Z score to standardize the values of each observation, followed by the calculation of the Euclidean index of similarity to define the degree of similarity between the behaviour of the study variables, these results being presented initially by hierarchical clustering and with using the single linkage algorithm (Figure 3), and then by Neighbour-joining clustering (Figure 4).

From the hierarchical clustering, it was possible to observe two large groups of similarity, the first formed by the lipid profile biomarkers and the second with the components of body composition and HDL.

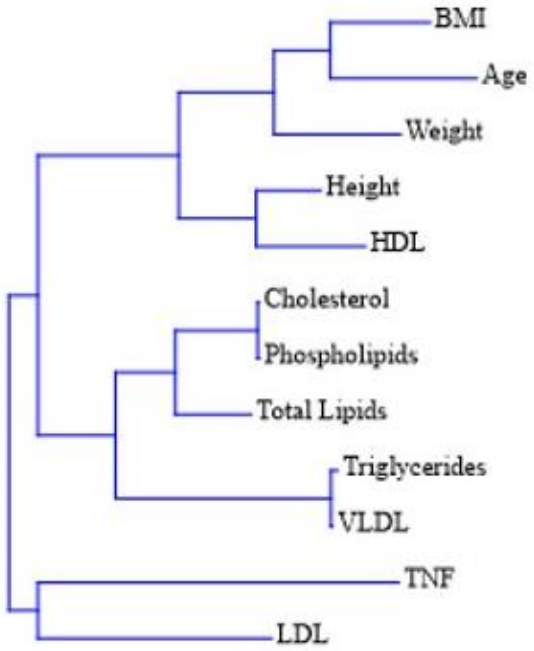

Figure 4. Neighbour joining clustering about Euclidean similarity index.

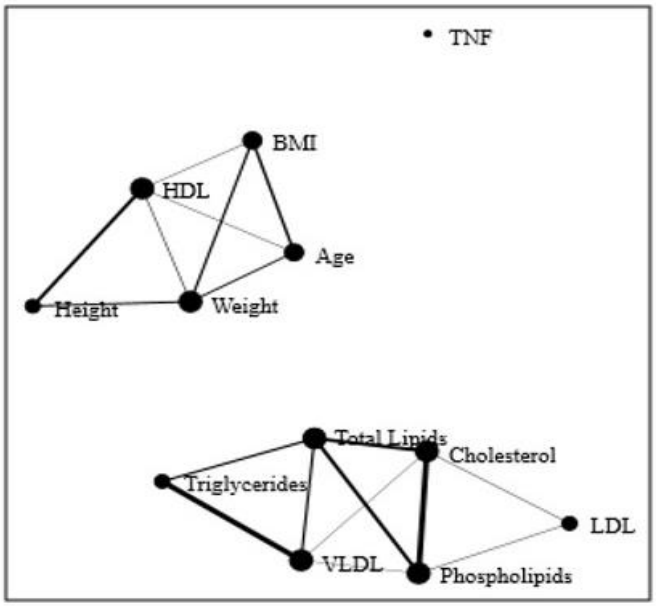

Figure 5. Network plot with Fruchterman-Reingold Algorithm after calculating Euclidean similarity index.

On the other hand, neighbour-joining clustering showed an intriguing similarity between TNF and LDL. Adding together the information on the correlation between TNF and HDL (Figure 2) and the similarity between the same TNF and LDL (Figure 4), a possible relationship between these plasma lipoproteins and Tumour Necrosis Factor becomes clear.

Finally, the Network plot with the Fruchterman-Reingold Algorithm (Figure 5) corroborated the findings of the Hierarchical clustering with algorithm single-linkage and Euclidean similarity index (Figure 3), giving robustness to the data.

DISCUSSION

Sportomicss studies aim to reproduce the natural conditions of different sports, training methods, and forms of exercise to seek an observation between the immunometabolic relationships of the method in question (Gonçalves et al., 2012; Gonçalves et al., 2022). These provide, in addition to reference values for the technical commission of the sport, also parameters for health professionals who work for public health and the well-being of practitioners (Gonçalves et al., 2020).

In the present sports study with a holistic analysis strategy using multivariate statistics as an exploratory machine learning model, it was possible to identify a similarity between the behaviour of TNF and plasma lipoproteins HDL and LDL under the effect of stress in a running test in athletes from the modality.

Tumour necrosis factor alpha (TNF-α) was initially recognized as a factor that causes the necrosis of tumours (Jang et al., 2021). However, but recently other studies have shown that this cytokine and adipokine have been positive for psychiatric diseases and depression (Uzzan & Azab, 2021), with intervertebral disc hernias (Wang et al., 2020), it may have therapeutic potential for the prevention of Alzheimer's disease (Torres-Acosta et al., 2020), can participate in the complaint of bones and cartilage of the temporomandibular joint (Wang et al., 2021), among other findings. They indicate that knowledge about TNF is still embryonic and suggest future studies in different fields of knowledge.

Regarding physical exercise, it is known that stimuli of moderate intensity and duration can have an anti-inflammatory effect with control of TNF levels, but that stimuli with high intensity, long duration, or high-intensity intervals can increase its levels, indicating dose-dependence (Paloucci et al., 2018).

The oxidation of HDL and LDL plasma lipoproteins plays an essential role in the modulation of the inflammatory response by macrophages, with a reduction in TNF secretion (Girona et al., 1997).

The present study showed a negative correlation between TNF and HDL and a similarity between the same TNF and LDL. These findings do not indicate a cause-and-effect relationship but suggest a possible modulation of the immune system, lipid metabolism, and exercise that requires further investigation.

CONCLUSION

The difficulty of treating large databases with different individuals, multiple biomarkers and collection times, in addition to physical parameters and sample characteristics, added to the decrease in new findings induced by the application of statistical tools of univariate analysis, indicate the need to apply strategies exploratory machine learning, generating holistic and integrated analysis of the results.

The present study showed a negative correlation between TNF and HDL and a similarity between the same TNF and LDL. These findings do not indicate a cause-and-effect relationship but suggest a possible modulation of the immune system, lipid metabolism, and exercise that requires further investigation.

Limitations

Because a convenience sample was chosen, and after applying the inclusion and exclusion criteria, there was a reduction in the sample size. However, the results are highly relevant with the advent of selected statistical methods.

AUTHOR CONTRIBUTIONS

ASG, RLN, LCOG and AMMN: essential contributions to the conception and design of the study protocol; acquisition, analysis and interpretation of data; and involvement in drafting of the manuscript. ASG, RLN, LCOG, MKMG, MCSF, LSM, DMO, MVAV and AMMN: critical revisions for important intellectual content. All authors read and approved the final manuscript.

SUPPORTING AGENCIES

No funding agencies were reported by the authors.

DISCLOSURE STATEMENT

No potential conflict of interest were reported by the authors.

REFERENCES

Antink CH, Braczynski AK, Ganse B. (2021). Learning from machine learning: prediction of age-related athletic performance decline trajectories. Geroscience. 43(5):2547-2559. https://doi.org/10.1007/s11357-021-00411-4

Cust EE, Sweeting AJ, Ball K, Robertson S. (2019). Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. Journal of Sports Sciences. 37(5):568-600. https://doi.org/10.1080/02640414.2018.1521769

Deo RC. (2015). Machine Learning in Medicine. Circulation. 132(20):1920-1930. https://doi.org/10.1161/circulationaha.115.001593

Girona J, Ville AE, Heras M, Olivé S, Masana L. (1997). Oxidized lipoproteins including HDL and their lipid peroxidation products inhibit TNF-alpha secretion by THP-1 human macrophages. Free Radical Biology & Medicine. 23(4):658-667. https://doi.org/10.1016/S0891-5849(97)00061-0

Gonçalves LCO, Bessa A, Freitas-Dias R, Luzes R, Werneck-de-Castro JPS, Bassini A, Cameron LC. (2012) A sportomics strategy to analyze the ability of arginine to modulate both ammonia and lymphocyte levels in blood after high-intensity exercise. J Int Soc Sports Nutr. 9(1):30. https://doi.org/10.1186/1550-2783-9-30

Gonçalves LCO, Santiago DDC, Neto AMM, Ferreira HSP, Verli MVA, Muniz-Santos R, Lopes JSS, Andrade CMB. (2020). Sportomics analysis of a high-intensity functional training method, the CrossFit. Europub Journal of Health Research. 1(1):2-20. https://doi.org/10.54747/ejhrv1n1-001

Gonçalves LCO, Neto AMM, Bassini A, Prado ES, Muniz-Santos R, Verli MVA, Jurisica L, Lopes JSS, Jurisica I, Andrade CMB, Cameron LC. (2022) Sportomics suggests that albuminuria is a sensitive

biomarker of hydration in cross combat. Scientific Reports. 12(8150):1-12. https://doi.org/10.1038/s41598-022-12079-7

Jang DI, Lee AH, Shin HY, Song HR, Park JH, Kang TB, Lee SR, Yang SH. (2021). The Role of Tumor Necrosis Factor Alpha (TNF-α) in Autoimmune Disease and Current TNF-α Inhibitors in Therapeutics. International Journal of Molecular Sciences. 22(5):2719. https://doi.org/10.3390/ijms22052719

Paloucci EM, Loukov D, Bowdish DME, Heisz JJ. (2018). Exercise reduces depression and inflammation but intensity matters. Biological Psychology. 133:79-84. https://doi.org/10.1016/j.biopsycho.2018.01.015

Tanisawa K, Wang G, Seto J, et al. (2020). Sport and exercise genomics: the FIMS 2019 consensus statement update. British Journal of Sport Medicine. 54(16):969-975. https://doi.org/10.1136/bjsports-2019-101532

Torres-Acosta N, O'Keefe JH, O'Keefe EL, Isaacson R, Small G. (2020). Therapeutic Potential of TNF-α Inhibition for Alzheimer's Disease Prevention. Journal of Alzheimer's Disease. 78(2):619-626. https://doi.org/10.3233/jad-200711

Uzzan S, Azab AN. (2021). Anti-TNF-α Compounds as a Treatment for Depression. Molecules. 26(8):2368. https://doi.org/10.3390/molecules26082368

Wang Y, Che M, Xin J, Zheng Z, Li J, Zhang S. (2020). The role of IL-1β and TNF-α in intervertebral disc degeneration. Biomedicine & Pharmacoterapy. 131: 110660. https://doi.org/10.1016/j.biopha.2020.110660

Wang Y, Bao M, Hou C, Wang Y, Zheng L, Peng Y. (2021). The Role of TNF-α in the Pathogenesis of Temporomandibular Disorders. Biological & Pharmaceutical Bulletin. 44(12):1801-1809. https://doi.org/10.1248/bpb.b21-00154

