Am J Cancer Res 2023;13(4):1617-1619
www.ajcr.us /ISSN:2156-6976/ajcr0150228

Commentary

Precision medicine in oncology - machine learning recommendations

Michal Ordak

Department of Pharmacotherapy and Pharmaceutical Care, Faculty of Pharmacy, Medical University of Warsaw, Warsaw, Poland

Received March 18, 2023; Accepted April 7, 2023; Epub April 15, 2023; Published April 30, 2023

Abstract: The article describes recommendations related to machine learning methods in oncology.

Keywords: Oncology, precision medicine, machine learning methods

In October 2022, Zhang, et al. posed the question, published in Medicine (Baltimore), what are the global research trends on precision oncology [1]. Precision medicine, also known as personalized medicine when taking into account factors such as environmental exposure, lifestyle and biological susceptibility, could revolutionize the prevention and treatment of many illness. Schuman et al. in Lancet Global Health pointed out that it is important to bear in mind that precision medicine mainly concerns high-income countries. Unfortunately, this may then run the risk of ignoring the needs of low-and middle-income countries [2]. How can the development of personalized medicine, including oncology, be accelerated? One option is the machine learning methods that have been widely developed in recent years. However, it is important to bear in mind some of the recommendations listed below.

According to an article published in ecancermedicalscience, the authors reviewed applications of machine learning techniques in oncology and explore how they work. Despite the progress that is occurring, it will still take many years to be able to accurately assess the potential of such methods [3]. Analysis of a range of different data sources could enable us to answer questions such as why do patients with cancer respond differently to identical treatments? Why does the therapy have positive effects in some patients, while others lack the expected effects or have negative side effects? It is important to remember that the data of oncological patients are very sensitive. For this reason, in order to build the trust of the individual patient, data should be collected in such a way that the patient is certain how and for what purpose the data would be used. With this approach, there is an opportunity to apply machine learning methods more effectively, i.e. with the aim of developing personalized medicine even further.

Another recommendation is to implement the teaching of machine-learning methods in medical universities, including on the basis of oncological patient datasets. According to one recent study published in the BMJ Health & Care Informatics, medical students reported limited awareness and education in this area. Four out of ten respondents had not heard the term machine learning, while two out of three said they had not spent time learning machine learning methods throughout their studies [4]. For this reason, medical universities should consider implementing additional classes for students and doctoral students in although basic machine learning methods. Demonstrating the practical application of machine learning methods to this group of people could be presented through two means. One way to do this is to look at an article published, for example, in ecancermedicalscience [3], in which the authors synthesized the machine learning techniques in oncology.

Another way is to carry out the most important machine learning methods on a simple instructional basis, so that the results can then be related to specific patients, such as those suffering from cancer. In other words, there would be an interesting demonstration of the importance of machine learning methods in the development of precision medicine. Based on a dataset of different groups of oncological patients, the students would apply the appropriate machine learning method of their choice, so that they could then try to assign an appropriate course of action for a particular person based on the results.

The fourth recommendation is to make the scientific community aware of the assumptions of the various machine learning methods. An example here is cluster analysis, for which one of the critical issues is sample representativeness and collinearity [5]. The effect of lack of representativeness may be to falsify the structure of the medical patient clusters created. Co-linearity, a situation where independent variables are highly correlated with each other, can make it difficult to assess their true impact. The resulting cluster system, e.g. of people suffering from cancer, may not find a place in reality. Another example is the sensitivity to outlier points, the presence of which can create an erroneous cluster of individuals [6]. If a dataset contains, for example, many outlier cases of oncological patients, the resulting number of clusters without analysis of these cases will not necessarily reflect the validity of the results obtained, and thus it will not be possible to implement the appropriate treatment for a particular individual. Such examples of unfulfilled assumptions may give rise to inconclusive findings related to personalized medicine. One further effect of this may be to reduce the confidence of a particular patient in all aspects of their treatment. When reviewing submitted manuscripts in which the authors have applied machine learning methods, expert statistical review should also be carried out. Otherwise, as was the case with the COVID-19 articles, a smaller percentage of them will meet statistical validity [7] and thus, instead of implementing precision medicine, we will be standing still all the time. This may result in the unnecessary media spread of erroneous premises, i.e. resulting from the publication of inconclusive results related to the misapplication of machine learning methods.

A final recommendation is to adequately educate health practitioners about key research findings related to the application of machine learning methods. For example, as the authors mentioned in Scientific Reports, nowadays psychiatrists rely mainly on symptoms in the differential diagnosis of anxiety and depression [8]. In addition, the gradual dissemination to this group of doctors of the latest research findings related to machine learning methods would allow personalized medicine to be developed even further in future and thus facilitate the selection of an appropriate course of action for a specific psychiatric patient, including those with depression and anxiety disorders mentioned above. For this reason, one of the recommendations is to represent the most important scientific developments related to the application of machine learning methods at the regularly held World Cancer Congress, so that the knowledge gained can then be translated into even greater development of precision medicine, also known as personalized medicine.

Disclosure of conflict of interest

None.

Address correspondence to: Michal Ordak, Department of Pharmacotherapy and Pharmaceutical Care, Faculty of Pharmacy, Medical University of Warsaw, Warsaw, Poland. E-mail: michal.ordak@wum.edu.pl

References

[1] Zhang B, Ao B, Lu X, Yang S, Bao P, Wang H, Li R and Huang Y. Global research trends on precision oncology: a systematic review, bibliometrics, and visualized study. Medicine (Baltimore) 2022; 101: e31380.

[2] Schumann G, Benegal V, Yu C, Tao S, Jernigan T, Heinz A, Araya R, Yu L and Calhoun V; Global Imaging Genetics in Adolescents Consortium. Precision medicine and global mental health. Lancet Glob Health 2019; 7: e32.

[3] Nardini C. Machine learning in oncology: a review. Ecancermedicalscience 2020; 14: 1065.

[4] Blease C, Kharko A, Bernstein M, Bradley C, Houston M, Walsh I, Hägglund M, DesRoches C and Mandl KD. Machine learning in medical education: a survey of the experiences and opinions of medical students in Ireland. BMJ Health Care Inform 2022; 29: e100480.

[5] Hummel M, Edelmann D and Kopp-Schneider A. Clustering of samples and variables with

mixed-type data. PLoS One 2017; 12: e0188274.
[6] Nowak-Brzezińska A and Gaibei I. How the outliers influence the quality of clustering? Entropy (Basel) 2022; 24: 917.
[7] Ordak M. COVID-19 research: quality of biostatistics. Arch Med Sci 2021; 18: 257-259.
[8] Richter T, Fishbain B, Markus A, Richter-Levin G and Okon-Singer H. Using machine learning-based analysis for behavioral differentiation between anxiety and depression. Sci Rep 2020; 10: 16381.