https://dergipark.org.tr/tr/pub/ijhmt Doi Number:10.31201/ijhmt.1252178

Int Journal Of Health Manag. And Tourism 2023, 8(1), 107-119

JHMT

Internatıonal Journal Of Health Management And Tourism

Editorial

Artificial Intelligence in Metabolomic Research

Hanifi KEBIROGLU* Hande HAYKIR**

*Dr., independent researcher Elazig, Turkey, hanifi007@hotmail.com

ORCID Number: 0000-0002-6764-3364

**PhD Student, Social Sciences Institute Sivas Cumhuriyet University, Sivas, Turkey

hande.haykirr@hotmail.com

ORCID Number: 0000-0001-9930-3420

Received: 17.02.2023

Accepted: 10.03.2023

Review Article

Abstract

The term "metabolomics" refers to high-throughput methods for detecting various metabolites and small molecules in biological samples. Undirected metabolomics, also known as unbiased global metabolome analysis, can be used to discover key metabolites as variables or measurements of human health and illness. From this vantage point, it is investigated how artificial intelligence and machine learning enable significant advances in non-targeted metabolic processes as well as significant findings in the early detection and diagnosis of diseases. Metabolomics is important for finding cures for many diseases. In the development of innovations in the field of biotechnology, it is of great importance to collect, filter, analyse, and use biological information in smart data. For this reason, many biotechnology companies and various healthcare organizations around the world have created large biological databases. This biological data accelerates the development of products in many areas. Algorithms are being developed for biological data analysis. It is thought that many disease treatments will be found when the human genome is edited. Machine learning techniques are effective tools for metabolomic investigation; however, they can only be used in straightforward computing scenarios. When used functionally, data formatting frequently calls for the use of sub-computational resources that are not covered in this area.

Corresponding author:Hanifi Kebiroglu, Dr., independent researcher Elazig, Turkey, e-mail: hanifi007@hotmail.com

Cite This Paper:

Kebiroglu, H., Haykır, H. (2023) Artificial Intelligence in Metabolomic Research. International Journal of Health Management And Tourism, 8(1): 107-119

Keywords: Metabolomics, Artificial intelligence, Medicine, Machine learning

INTRODUCTION

Modern science has a form in which it is thought possible to formulate "how" the events observed in nature occur, and therefore to "know" how things will happen before they happen. A paradigm shifts in science began when computers made possible operations that could not be calculated by hand. Computer-aided experimental equipment has developed very rapidly and has begun to scan and produce data at a speed that was unimaginable before. This situation has revealed many previously unforeseen problems. Making sense of the produced data requires a completely different scientific study methodology. Now, numerical approaches such as mathematical modeling, complex networks, and machine learning and the application of these approaches as software tools have become inevitable for almost every branch of science. With technological developments, reading of all DNA sequences in the cell (genomics), determination of all RNA amounts (transcriptomics), or determination of all protein amounts (proteomics)-omic data has become available. All these have led to the establishment of an approach to evaluate the system instead of understanding the whole by interpreting and combining the intracellular mechanisms into small parts. This new approach is called systems biology. Apart from the genomic, transcriptomic, and proteomic data that correspond to the main activities in the cell, there are many different layers of big data. With the emergence of metabolic models at the genome-scale, metabolomic data has emerged and it is aimed to simultaneously analyze complex structures consisting of, for example, 7500 reactions and 5000 metabolomics for a human cell.

Interactions of proteins, whose number is approaching 20,000 in the cell, are examined in the form of network structures with the help of interatomic data, and unknown mechanisms are tried to be clarified (Jung-Ming G. Lin, et al. 2022). Analyzing the three-dimensional (3D) structures of proteins is of vital importance in the determination of their functional parts and, therefore, in the development of specific drugs for diseases caused by proteins. Computational structural biology, which tries to predict the structure of proteins whose amino acid sequence is known but whom three-dimensional (3D) structure is unknown, by machine learning methods using proteins of known three-dimensional (3D) structure, has become an almost independent discipline. (Lander ES et al., 2001).

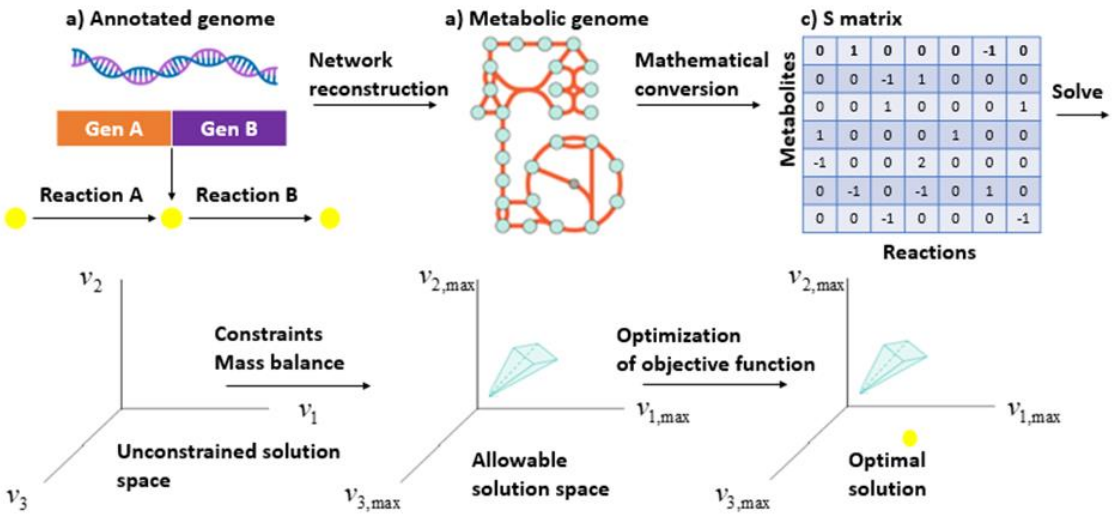

Figure 1. Structure of proteins with known amino acid sequence but unknown three-dimensional structure using proteins with known three-dimensional structure (Xin Fang et al., 2020).

Metabolomic applications include monitoring clinical disease, evaluating therapeutic applications, and understanding the effects of genetic modifications. All metabolomics found in the body are defined by metabolomics techniques. Metabolomics is very diverse and exists in different concentrations. In this technique, detection, separation, identification, and analysis of metabolomics are performed. Performing all these processes in a short time without loss of sensitivity and high efficiency is one of the important parameters of metabolism (Bren L., 2005).

Purpose

The process through which the body's chemical reactions create the numerous metabolites required for human life is known as metabolism. The two fundamental processes of metabolism are anabolism, or the creation of compounds required by the cell, and catabolism, or the breakdown of molecules for energy. Inactivation, detoxification, and the elimination of foreign or undesirable substances are additional functions of metabolism. Understanding these systems and how they relate to human physiology in terms of health and disease is crucial. The measurement of small molecules in biological samples is the basis of metabolomics, a thorough, high-throughput study that may be used to explore these processes individually or collectively (usually blood, urine, or

saliva) (J.D. (2018)). Phenotypic changes in metabolite profiles may result from disruption of the ohmic layer. Exposomics (shaded in blue above) expands this to measure phenotype-related metabolite profiles, including exogenous small molecules and the effect of exogenous and non-genetic variants on "omic" cascades. Sample collection times based on phenotypic identification can help determine whether metabolite biomarkers or pathways are relevant to disease etiology, diagnosis, or progression (Topfer, N. 2021).

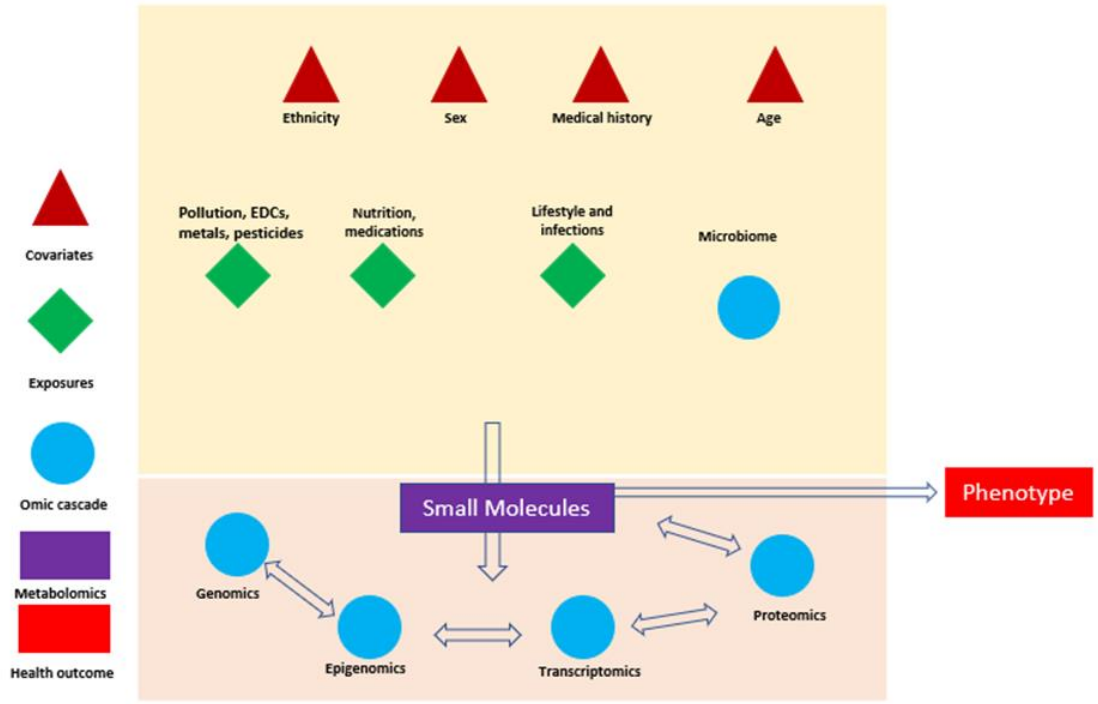

Figure 2. Metabolomics supports the discovery process. Metabolomics (pink shading, bottom) focuses on endogenous small molecules as outputs of metabolic systems.

Many researchers wonder what the potential for artificial intelligence could be in the future of medicine and healthcare. Looking at AI studies, the general implication is that the way modern machines draw inferences is very different, and that complementarity is ultimately a source of strength, as well as an opportunity for researchers to make better decisions, but that there are no more important decisions than the decisions physicians have to make every day. Therefore, it is believed that physicians who partner with artificial intelligence in decision-making will see their healing powers increase even more. It can be said that the artificial intelligence revolution of this century is really a renaissance, a rebirth today. Artificial intelligence and machine learning have

significant advantages for medicine, health, science, and engineering today. Artificial intelligence and machine learning-derived technologies enable faster and larger processing of data as well as advanced analytics. It allows the identification of inconsistencies, minimizing the need for human intervention. Based on data from many patients, the proposed approach is trained, and feature distributions are examined to identify the most important features (Jacobs DM. 2017). It is then trained using selected interesting features to construct a diagnostic classifier and evaluate data from different patients (Jialal, I. 2019).

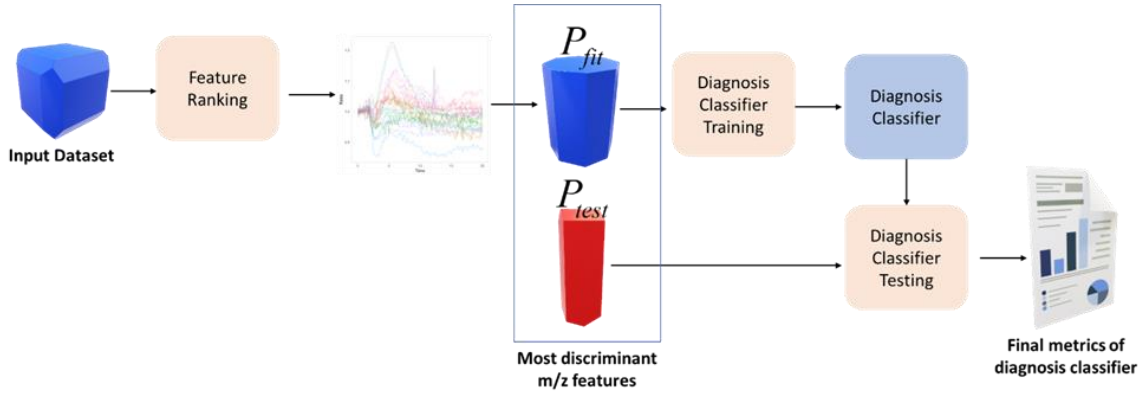

Figure 3. Steps to train and test the diagnostic classifier.

Collecting, filtering, analyzing, transforming, and transforming biological information into smart data is of great importance in the development of innovations in the field of biotechnology. For this reason, many biotechnology companies and various healthcare organizations around the world have created large biological databases. This biological data accelerates the development of products in many areas. The scope of health biotechnology, which is one of the most important areas of biotechnology, can be listed as the development and production of biopharmaceuticals; chemical analysis of different compounds; diagnosis and treatment models based on RNA and DNA data; and planning; personalized medicine and others; the development of health methods; enzyme research and other similar biological processes. The main elements that feed the bioinformatics methods are the omic data presented within the scope of biological sciences. All methods of identifying hereditary material, genomes, and studying structural and functional features in organisms are defined as genomics. Studies to elucidate the biological function journey of information in the genome have advanced with other omics technologies such as transcriptomics, proteomics, and metabolomics. There are metabolomic approaches that enable the

identification, detection, and analysis of relevant small molecules, as well as transcriptomic technologies that allow the study of all RNA products created in vivo and proteomic technologies that examine the structures, modifications, positions, and functions of translated proteins. The structures of functional components in the organism, such as carbohydrates, lipids, and vitamins, are used to understand biological processes in the organism. makes a major contribution (Genetic Engineering and Biotechnology News publishing).

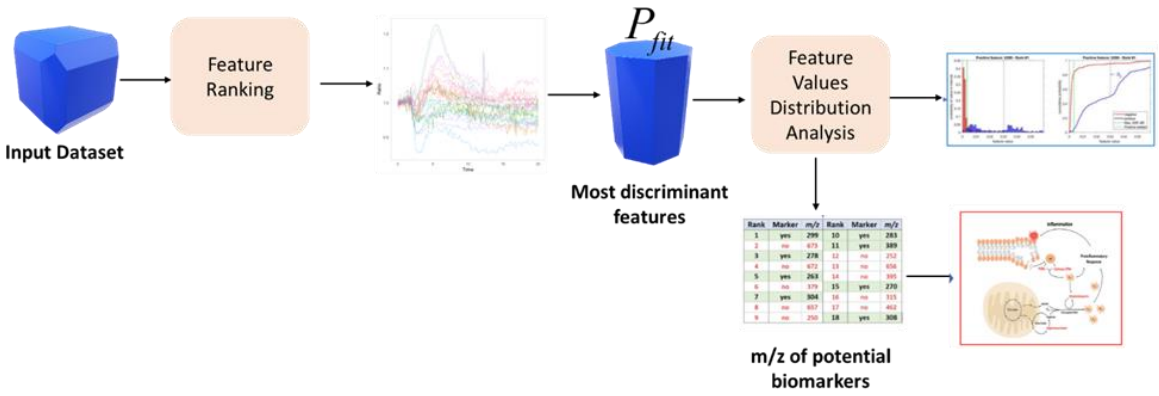

Figure 4. Identification of possible biomarkers using the proposed machine learning approach and the most salient features (Dabaja, M. Z., et al., 2019).

1. RESEARCH METHODOLOGY

Vaccines, drugs, and other therapeutic applications developed within the framework of health biotechnology use living organisms, cells, molecules (e.g., DNA, RNA, peptides, proteins) and biological information obtained from these structures. Information technology tools such as artificial intelligence and machine learning are used at very important stages in all stages, from molecular discovery, which is the first stage of drug development, to in vitro tests, from animal experiments to clinical research. The discovery of drug molecules, whether a conventional chemical drug or a biotech drug, is a very long and arduous step. For classical molecules, only 1 out of 8,000 molecules on average is released as a drug. It takes a lot of time, effort, and money to develop even this single molecule. Thanks to artificial intelligence, these molecular scanning processes have been reduced from 5 to 10 years to months or even days. The best example of this is the work of Alex Zhavoronkov and colleagues, who reduced years of molecular scanning to 21

days. Likewise, artificial intelligence solutions are used in the design and execution of in vitro studies. Thanks to artificial intelligence and deep learning, laboratory studies are better optimized, and the repetition of experiments is reduced, resulting in significant time and cost savings. In the animal testing phase, AI-powered programs ensure that research is time-consuming and done with fewer animals. In addition, it is thought that the use of organ-on-chip and tissue-on-chip will reduce the need for animal experiments thanks to biochip technology. When it comes to the clinical trial phase, AI plays an undeniable accelerator role in the entire human phase study design, the right recruitment of volunteers (patients), and the collection of efficacy and safety data. Artificial intelligence-assisted predictive measures can be planned by collecting, reviewing, and analyzing information about all adverse events that occur during the clinical trial process and after drug release. For example, the unexpected situations of a drug or vaccine developed according to a certain age, gender, concomitant disease or some other conditions can be determined by big data analysis, and these special groups can be detected quickly with artificial intelligence (Kurnaz I. Ed. (2019); Zhavoronkov A (2019); Zacharoula & Yannakakis, Georgios (2020); Tekade, R. K. (2021)). They are used in the processes of collecting, analyzing, interpreting, and transforming biological information into a solution with artificial intelligence (Ünver and Kurnaz, 2019).

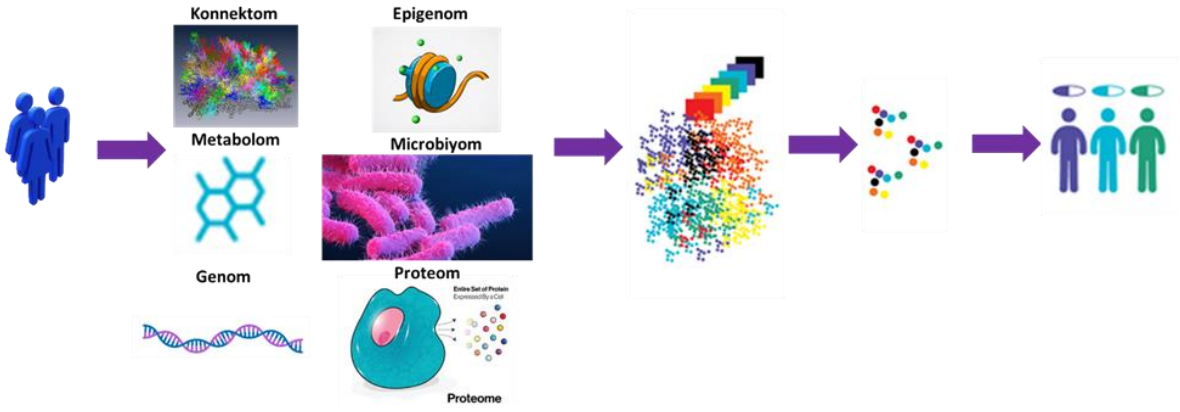

Figure 5. Many new technologies, from implantable biosensors to wearable sensors

To survive while balancing a wide range of varied environmental variables, cellular life requires a vast diversity of metabolic phenotypes (Gray A et al. (2009); Barnhill S et al. (2002); Li JV et al. (2011); Porfirio B et al. (2009)). Due to the technology's capacity to quantitatively test numerous metabolomic targets simultaneously, liquid chromatography-mass spectrometry (LC-

MS) is the most effective method for differentiating these phenotypes (Brandon TR et al. (2005); Neumann S et al. (2008); Siuzdak G et al. (2011). As a result, non-targeted metabolomic profiles are excellent at detecting environmental, stress, or disease-related indicators. However, this may be complicated by the typical biological variance in metabolic response (Bothner B et al. 2014). The best way to properly apply machine learning to metabolomic datasets is likely to combine it with data mining techniques like principal component analysis and hierarchical clustering. This is primarily since data mining methods may be used unsupervised, which reduces the danger of human bias and enables the machine to discover significant associations between samples. Machine learning techniques are effective tools for metabolomic investigation; however, they can only be used in straightforward computing scenarios. It is frequently necessary to use sub-computational resources not addressed in this section when preparing data for functional purposes (Istanbul, February 16, 2021— Harvard T.H.). Harvard Chan School of Public Health researchers have utilized machine learning, a branch of artificial intelligence, to pinpoint the elements more precisely in walnuts that may be lowering the risk of type 2 diabetes and cardiovascular disease (two of the leading causes of death in the US). This study, funded by the California Walnut Commission and featured in The Journal of Nutrition, identified 19 metabolomics linked to walnut consumption using a novel method called agnostic machine learning. Various metabolites are produced by the body based on the type of food eaten. A 37% reduced incidence of type 2 diabetes and a 63% lower risk of cardiovascular disease were linked to the walnut metabolomics profile. This work adds to the three decades of prior research on walnuts and heart health by being the first to investigate the link between walnut metabolomics and the risk of cardiometabolic illness.

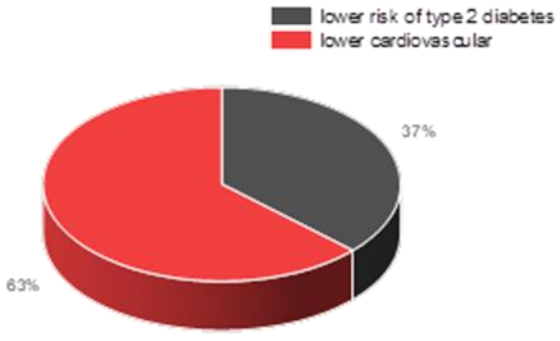

Figure 6. The relationship between walnut metabolomics and the risk of cardiometabolic disease.

The PREvención con Dieta Mediterránea (PREDIMED) project, a large-scale, multi-year investigation of the benefits of the Mediterranean diet on avoiding cardiovascular disease in persons at high risk of heart disease, included 1,833 participants, whose data were reviewed by the researchers.

Table 1. Participants were between the ages of 55 and 80 and followed one of the three diets.

1	Adding mixed nuts to a Mediterranean diet (50 percent walnuts, 25 percent almonds, and 25 percent hazelnuts)
2	Olive oil extra virgin enriches the Mediterranean diet.
3	Low-fat diet: walnut metabolomics reveals a walnut metabolomic signature linked to a decreased risk of cardiovascular disease and type 2 diabetes.

These results underline the connection between eating walnuts as part of a balanced diet and cardiometabolic health. This epidemiology study's new methods will aid in establishing associations between food and illness. The results, however, do not imply causation. Because this study only looked at older Hispanic people, more research in other demographics is needed. Future research will be required to uncover other walnut intake indicators that were not monitored in this study as well as to comprehend individual metabolic reactions following walnut eating, given the quickly developing area of metabolomics.

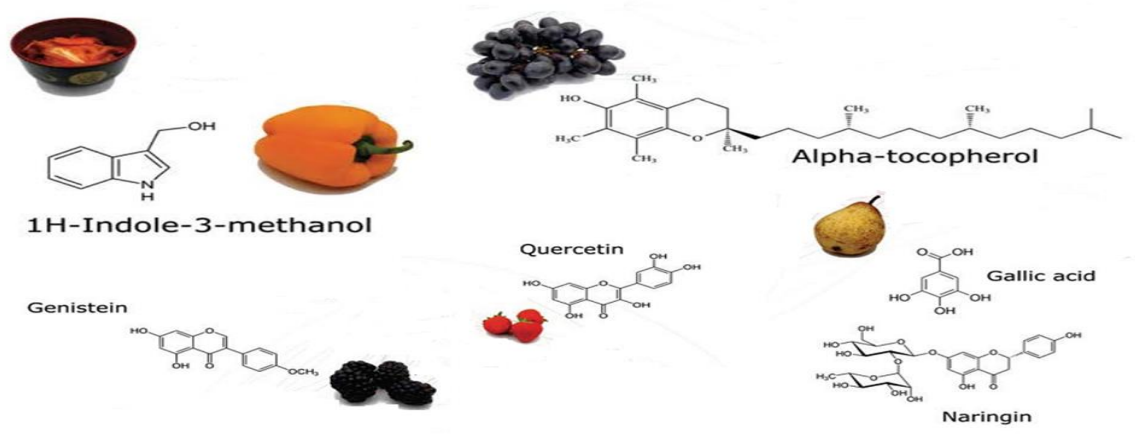

Figure 7. Profiles of compounds in certain foods with predicted efficacy in targeting the SARS-CoV-2-host interactome (Xian Y, et al., 2020; Gordon DE, et al. 2020).

Potential bioactive anti-COVID-19 compounds in foods were found using the network machine learning approach in the study "Network machine learning maps phytochemically rich "Hyperfoods" to combat COVID-19," which was published in the journal Human Genomics. Among the 5658 experimental and clinically licensed medications that are effective against COVID-19, it has been calibrated using a machine learning model to forecast anti-COVID-19 candidates with a balanced classification accuracy of 80–85%. These have been found to be the most promising therapeutic candidates that may be "reused" to treat COVID-19, including popular medications like simvastatin, atorvastatin, and metformin that are cycled to treat metabolic and cardiovascular diseases. 52 biologically active compounds from different chemical classes, including flavonoids, terpenoids, coumarins, and indoles, were found after a database of 7694 bioactive food-based molecules was put through a calibrated machine learning algorithm. These compounds are predicted to target the SARS-CoV-2-host interactome networks. Based on the variety and relative concentrations of probable compounds with antiviral capabilities, this was utilized to build a "food map" with the theoretical anti-COVID-19 potential of each projected component. According to the conventional approach to developing antiviral drugs, a medication must target a viral protein. In this regard, comprehensive computational molecular docking simulations have been conducted to identify plant-based bioactive compounds for certain SARS-CoV-2 protein targets. The durability of intricate virus-host interaction networks against specific protein degradation is one of this method's many shortcomings. In addition, escaping viral variants may confuse the potential effects of vaccinations and medications against genes or protein targets unique to SARS-CoV-2. In order to find food-based bioactive compounds targeting the SARS-CoV-2 and human interactome networks, we integrated network-based machine learning techniques with mobile supercomputing and interatomic data. First, experimentally confirmed medicines' anti-COVID-19 characteristics were predicted using the suggested machine learning technique. The models were used to find drug-like compounds in food after calibration. Medication repurposing and reuse in cancer research, population stratification based on mutations, drug repurposing, and food-based anti-cancer molecular therapies have all benefited from similar network propagation techniques. Machine learning algorithm parameters were tuned to forecast experimentally confirmed medications against COVID-19 in the cross-validation context (Vázquez-Calvo et al. 2017; Gysi DM et al., 2020; Xian Y, et al. 2020; Boozari M, et al. 2020; Veselkov, K. et al. 2021).

2. CONCLUSION

The algorithm will thus improve as we gather more metabolomics data and come to comprehend the shift in metabolomics, helping us find the best medicines for the right patients. It will get closer to achieving its goal of saving lives with every line of code written. It will be easier to find connections between nutrition and disease with the use of new epidemiological study technologies. This study suggested new avenues for metabolomics, which need high-end tools and skilled workers, but also suggested simplified sample preparation, a shorter turnaround time for metabolite analysis, and the use of cutting-edge data processing methods for decision-making. Important metabolomics discoveries have already been made and are anticipated to be made thanks to developments in artificial intelligence and machine learning in computers. A new era of discovery was ushered in by the success of deep learning algorithms on unstructured data as well as the fusion of new artificial intelligence and machine learning techniques that aren't commonly used in metabolomics with existing datasets or samples that contain low and high concentrations of well-known chemicals, trained training and peak aggregation algorithms, environments, and precision health. They serve as crucial launching pads for refueling.

Conflict of Interest: The authors have no conflicts of interest to declare.

Funding: The authors declared that this study had received no financial support.

References

Bertini I, Calabro A, De Carli V, Luchinat C, Nepi S, Porfirio B et al (2009). The metabonomic signature of celiac disease. *J Proteome Res* 8:170–177.

Boozari M, Hosseinzadeh H. (2020). Natural products for COVID¬19 prevention and treatment regarding to previous coronavirus infections and novel studies. *Phytotherapy Res*. p. 1-13

Bren L. (2005). Metabolomics: Working toward personalized medicine. *FDA Consum*, 39: 28-33.

Dias-Audibert, F. L., Navarro, L. C., Oliveira, D. N., Melo, C. F. O. R., Guerreiro, T. M., Dabaja, M. Z., et al., (2019). Combining Machine Learning and Metabolomics to IdentifyWeight Gain Biomarkers. Campinas: *IEEE Dataport*.

Giannakos, Michail & Voulgari, Iro & Papavlasopoulou, Sofia & Papamitsiou, Zacharoula & Yannakakis, Georgios. (2020). *Games for Artificial Intelligence and Machine Learning Education: Review and Perspectives*.

Gordon DE, et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 583(7816):459–68.

Guan W, Zhou M, Hampton CY, Benigno BB, Walker LD, Gray A et al (2009). Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines. *BMC Bioinformatics* 10:259.

Guyon I, Weston J, Barnhill S et al (2002). Gene selection for cancer classification using support vector machines. *Mach Learn* 46:389–422.

Gysi DM, Í Do V, Zitnik M, Ameli A, Gan X, Varol O, Ghiassian SD, Patten JJ, Davey R, Loscalzo J, Barabási AL. (2020). *Network medicine framework for identifying drug repurposing opportunities for COVID-19*.

Heinemann J, Mazurie A, Tokmina-Lukaszewska M, Beilman GJ, Bothner B et al., (2014). Application of support vector machinesto metabolomics experiments with limited replicates. *Metabolomics* 10:1121–1128.

Jacobs DM, Smolders L, Lin Y, de Roo N, Trautwein EA, van Duynhoven J, Mensink RP, Plat J and Mihaleva VV (2017) Effect of Theobromine Consumption on Serum Lipoprotein Profiles in Apparently Healthy Humans with Low HDL-Cholesterol Concentrations. Front. Mol. Biosci. 4:59. doi: 10.3389/fmolb.2017.00059

Jang, C., Chen, L., and Rabinowitz, J.D. (2018). Metabolomics and isotope tracing. *Cell* 173, 822–837.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Funke R. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921.

Laponogov, I., Gonzalez, G., Shepherd, M., Qureshi, A., Veselkov, D., Charkoftaki, G., ... & Veselkov, K. (2021). Network machine learning maps phytochemically rich "Hyperfoods" to fight COVID-19. *Human genomics*, 15(1), 1-11.

Lent-Schochet, D., McLaughlin, M., Ramakrishnan, N., and Jialal, I. (2019). Exploratory metabolomics of metabolic syndrome: a status report. *World J. Diabetes* 10:23.

Lin J. M. G., Kourtis S., Ghose, R., Lorente N.P., Kubicek, S., Sdelci, S. (2022). Metabolic Modulation of Transcription: The Role of One-Carbon Metabolism, Cell Chemical Biology, Volume 29, Issue 12, P1664-1679

Lin X,Wang Q, Yin P, Tang L, Tan Y, Li., (2011). A method for handling metabonomics data from liquid chromatography/mass spectrometry: combinational use of support vector machine recursive feature elimination, genetic algorithm and random forest for feature selection. *Metabolomics* 7(4):549–558.

Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discovery Today*, 26(1), 80–93.

Sahu, A., Bla¨ tke, M.A., Szyma nski, J.J., and Topfer, N. (2021). Advances in flux balance analysis by integrating machine learning and mechanism-based models. *Comput. Struct. Biotechnol. J*. 19, 4626–4640.

Salman, Ünver S., and Aksan, Kurnaz I. Ed. (2019). Adım Adım Biyogirişimcilik: Biyoteknoloji Girişimci ve Yatırımcılarına Yol Haritası. *İstanbul: ABA Yayınları*.

Smith C, O'Maille G, Want EJ, Qin C, Trauger S, Brandon TR et al (2005). METLIN: a metabolomike mass spectral database. *Ther Drug Monit* 27(6):747–751

Tautenhahn R, Bo¨ttcher C, Neumann S et al (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* 9:504.

Vázquez-Calvo, Á.; De Oya, N.J.; Martín-Acebes, M.A.; Garcia-Moruno, E.; Saiz, J.-C. (2017). Antiviral Properties of the Natural Polyphenols Delphinidin and Epigallocatechin Gallate against the Flaviviruses West Nile Virus, Zika Virus, and Dengue Virus. Front. *Microbiol*. 8, 1314.

VeselKov KA, Vingara LK, Masson P, Robinette SL, Want E, Li JV et al., (2011). Optimizing preprocessing of ultraperformance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Anal Chem* 83:5864–5872.

Xian Y, et al. (2020). Bioactive natural compounds against human coronaviruses: a review and perspective. *Acta Pharm Sin B.* 10(7):1163–74.

Xin Fang, Colton J. Lloyd & Bernhard O. Palsson, (2020). Reconstructing organisms in silico: genome-scale models and their emerging applications, *Nature Reviews Microbiology*.

Yanes O, Tautenhahn R, Patti GJ, Siuzdak G et al (2011). Expanding coverage of the metabolome for global metabolomike profiling. *Anal Chem* 83(6):2152–2161.

Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, Terentiev VA, Polykovskiy DA, Kuznetsov MD, Asadulaev A, Volkov Y, Zholus A, Shayakhmetov RR, Zhebrak A, Minaeva LI, Zagribelnyy BA, Lee LH, Soll R, Madge D, Xing L, Guo T, Aspuru-Guzik A., (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat Biotechnol*. Sep;37(9):1038-1040.