doi 10.18699/bgrs2024-8.1-12

Application of machine learning algorithms to study the gut microbiota of patients with depressive disorders

Galanova O.O.[1, 2]*, Kovtun A.S.[1]

[1] *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology, State University, Dolgoprudny, Russia*

* *galanova.oo@phystech.edu*

Key words: depression; gut-microbiota; gut-brain axis; machine learning

Motivation and Aim: Depression remains one of the most important health problems of our time due to its widespread and devastating impact on the human psyche. According to the World Health Organization, 5 % of adults worldwide suffer from depression. Scientific research usually looks at the biochemical and neural aspects of a given disease. At the same time, the potential impact of the gut microbiota (GM) is often not taken into account, despite its enormous metabolic potential. Gut bacteria are able to influence the development and functioning of various human organs and systems, including the nervous system through two-way communication implemented through the gut-brain axis [1]. It means that disorders in this complex system may have a significant impact on human mental health, and vice versa [2, 3]. Studies of GM metabolites with neuromodulating activity affecting mental health are becoming increasingly relevant [4, 5]. In this context, understanding of the role of GM opens up the possibility of developing more effective methods for the diagnosis and treatment of depression, as well as other mental disorders [1, 4, 6]. Identifying and analyzing the relationships between the development of mental illness in humans and changes in GM, both taxonomic and functional, is a complex task that requires processing and interpretation huge amounts of data. Machine learning provides a promising approach to analyzing large-scale metagenomic data and identifying taxonomic and functional biomarkers associated with various diseases, including depressive disorders. The general idea of many ML algorithms is to train a model based on a list of parameters and a special training dataset and then use it for the analysis of a general dataset. One of the most important parts of application of the ML is the assembly of an appropriate training dataset, which avoids such problems as biases, batch effect, noisiness and dataset size. Random forest classification is one of the most widely used algorithms in metagenomics thanks to its ability to work with relatively small datasets while delivering stable results. With the help of various ML approaches, researchers can solve the tasks of taxonomic characterization, sample clusterization and correlation of microbiota composition with various diseases. A number of studies have shown a correlation between changes in GM and major depressive disorder [7–9]. The results obtained using machine learning may contribute to a better understanding of the role of GM in major depressive disorder and pave the way for new diagnostic and therapeutic strategies.

Methods and Algorithms: A cohort of 74 samples of gut microbiota, obtained from 36 patients with depression and 38 healthy controls from Moscow Region, was used for the research. The metagenomic analysis was carried out using the signature approach that allows for identification of both genes, their bacterial source and abundance of the pairs (gene; taxon). For the application of the machine learning approach, the cohort was first divided into two groups, training and validation, and then broadened with the

simulated data generated using MOSTLY AI tool based on GAN. The models were generated using several machine learning algorithms, including random forest, elastic net, and YOLO.

Results: Compositional changes in the GM of patients with MDD and subsequent changes in its neurometabolic potential were indicated. The results demonstrated a decrease in beneficial bacteria, such as *Faecalibacterium*, *Roseburia*, and *Lachnospira*, along with an increase in abundances of opportunistic pathogens, such as *Ruthenibacterium* and *Escherichia* in the PwD group compared with healthy controls. These changes characterized the dysbiosis of GM and could be contributing factors to the development of depression. The signature approach at the "Species" level revealed a set of genes with significantly decreased abundance in the patients' group, most of which were identified in species *Faecalibacterium prausntizii*. These data represent the neurometabolic signature of the GM of depressive patients and can be used as biomarkers. Further application of the machine learning algorithms allowed for creation of the models that can determine the depressed condition of a patient using the information about their metagenomics signature. The best results were observed for the models generated by the random forest algorithm with the prediction accuracy of 80 % and YOLO with the prediction accuracy of more than 90 %.

Conclusion: Application of machine learning algorithms in metagenomics studies is gaining popularity. They can be used for classification tasks introducing novel approaches for taxonomic annotation of microbiota, and in association studies for identification of correlations between changes in microbiota and the host's condition, including such disorders, as depression. In our research we have made a complex description of changes of gut microbiota, which can be connected with the depressed condition. The obtained models indicate that the specific genes and taxa of the gut microbiota can be viewed as the novel group of biomarkers for diagnostics of depression in addition to the currently used methods. The currently developed models can already be used for the *in silico* diagnostics using the metagenomics sequencing and the signature approaches. The development of new methods of treatment for the depression by the correction of gut microbiota with the use of psychobiotics is another possible application for these algorithms.

Funding: The study is supported by the Russian Science Foundation (№ 20-14-00132).

References

1. The neuroscience of depression genetics, cell biology, neurology, behavior, and diet. Eds. Martin C.R., Hunter L.-A. London, UK, 2021
2. Sasso J.M. et al. Gut microbiome-brain alliance: a landscape view into mental and gastrointestinal health and disorders. *ACS Chem Neurosci.* 2023;14:1717-1763. doi 10.1021/acschemneuro.3c00127
3. Kovtun O., Leneva N., Bykov Y.S. et al. Structure of the membrane-assembled retromer coat determined by cryo-electron tomography. *Nature*. 2018;561:561-564. doi 10.1038/s41586-018-0526-z
4. Poluektova E.U., Yunes R.A., Danilenko V.N. the putative antidepressant mechanisms of probiotic bacteria: relevant genes and proteins. *Nutrients*. 2021;13:1591. doi 10.3390/nu13051591
5. Averina O.V., Zorkina Y.A., Yunes R.A. et al. Bacterial metabolites of human gut microbiota correlating with depression. *Int J Mol Sci*. 2020;21:9234. doi 10.3390/ijms21239234
6. Liu L., Wang H., Chen X., Zhang Y., Zhang H., Xie P. Gut microbiota and its metabolites in depression: From pathogenesis to treatment. *EBioMedicine*. 2023;90:104527. doi 10.1016/j.ebiom.2023.104527
7. Angelova I.Y. et al. Unveiling the connection between microbiota and depressive disorder through machine learning. *Int J Mol Sci*. 2023;24(22):16459
8. Radjabzadehc D. et al. Gut microbiome-wide association study of depressive symptoms. *Nat Comm.* 2022;13:7128. doi s41467-022-34502-3
9. Liang S. et al. Multi-cohort analysis of depression-associated gut bacteria sheds insight on bacterial biomarkers across populations. *Cell Mol Life Sci*. 2022;80(1):9. doi 10.1007/s00018-022-04650-2