

Explainable AI for gut microbiome-based diagnostics

Acharjee, Animesh

DOI:

[10.1515/dx-2023-0062](https://doi.org/10.1515/dx-2023-0062)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Acharjee, A 2023, 'Explainable AI for gut microbiome-based diagnostics: colorectal cancer as a case study', *Diagnosis*. <https://doi.org/10.1515/dx-2023-0062>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Letter to the Editor

Animesh Acharjee*

Explainable AI for gut microbiome-based diagnostics: colorectal cancer as a case study

<https://doi.org/10.1515/dx-2023-0062>

Received May 30, 2023; accepted June 4, 2023;

published online June 19, 2023

Keywords: diagnostics; interpretable AI; microbiome

To the Editor,

An increasing number of studies have found a correlation between gut microbiome composition changes and colorectal cancer (CRC) [1–3]. Researchers have utilised machine learning (ML) algorithms to identify potential CRC biomarkers and to distinguish CRC patients from healthy individuals [4]. Identification of the diagnostics microbes will help tremendously to develop a non-invasive biomarker for the CRC patients. Recent microbiome studies frequently employ random forest [5, 6] due to its predictive ability and capacity to generate feature importance. Random forest is a decision tree-based algorithm and proven to be sophisticated enough to capture bacteria that are typically associated with CRC, however it is unable to identify species that are only significant for a subset of patients.

Recently, Rynazal et al., published an exciting paper on [7] on explainable AI and their application on microbiome data sets derived from CRC patients. This study investigated the feasibility of employing a method called as “Shapley Additive Explanations (SHAP)” technique for analysing gastrointestinal microbiome data [7]. The difference between normal ML methods compared to the

explainable AI method is interpretability. For example: Using Random Forest method, we can rank the important variable (here microbiome) called as “Gini importance” and we rely solely on global explanation techniques. Local explanations enable us to observe distinct bacterial contribution patterns among CRC patients. One of the examples provided Rynazal et al., would be, *Clostridium symbiosum* was found to be the most influential bacteria for the patient with subject ID SAMD00114911. In Figure 1, shows the application of the explainable AI and discovery of the interpretable microbes for therapeutic intervention.

In contrast, *Eubacterium eligens* displays the opposite pattern, indicating that a higher abundance of this species is associated with a lower CRC risk. This understanding of the direction of effects cannot be obtained using commonly employed global explanation methods, such as the random forest’s built-in feature importance.

This study further investigated the benefits of using SHAP for individual feature contribution analyses and CRC subtyping. In addition, a Python library was developed to assist microbiome researchers with comparable analyses. Based on the contribution of each bacterial species to the classifier, the study discovered that SHAP can generate feature contributions for every single ML prediction and classify the disease group into subgroups of CRC patients. This study investigated the feasibility of using explainable AI for CRC classification based on the gastrointestinal microbiome. SHAP was utilised to acquire more individualised feature importance that can be used to identify potential bacterial CRC biomarkers.

This method will benefit the microbiome community and encourage researchers to utilise local explanations for a more personalised identification of feature importance and subtyping and a huge potential for microbiome based diagnostics [8]. CRC is just an example of such applications however, it can be used other cancer types or other diseases like inflammatory bowel diseases [9] or metabolic diseases to make an interpretable model and hence a new avenue towards precision medicine [10].

*Corresponding author: Dr. Animesh Acharjee, Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK; Institute of Translational Medicine, University of Birmingham, Birmingham, UK; MRC Health Data Research UK (HDR UK), Midlands Site, Birmingham, UK; and University of Birmingham, B15 2TT, Birmingham, UK, Phone: +4407403642022, E-mail: a.acharjee@bham.ac.uk. <https://orcid.org/0000-0003-2735-7010>

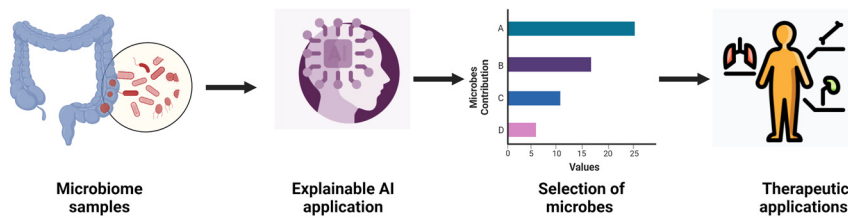


Figure 1: Shows the application of the explainable AI and discovery of the interpretable microbes for therapeutic intervention.

Research funding: None declared.

Author contributions: Animesh Acharjee is responsible for the writing, conceptualization, acquisition of data, and interpretation of the data.

Competing interests: The author declares no conflicts of interest regarding this article.

Informed consent: Not applicable.

Ethical approval: Not applicable.

References

1. Roelands J, Kuppen PJK, Ahmed EI, Mall R, Masoodi T, Singh P, et al. An integrated tumor, immune and microbiome atlas of colon cancer. *Nat Med* 2023;29:1273–86.
2. Rebersek M. Gut microbiome and its role in colorectal cancer. *BMC Cancer* 2021;21:1325.
3. Ternes D, Karta J, Tsenkova M, Wilmes P, Haan S, Letellier E. Microbiome in colorectal cancer: how to get from meta-omics to mechanism? *Trends Microbiol* 2020;28:401–23.
4. Bosch S, Acharjee A, Quraishi MN, Bijnsdorp IV, Rojas P, Bakkali A, et al. Integration of stool microbiota, proteome and amino acid profiles to discriminate patients with adenomas and colorectal cancer. *Gut Microb* 2022;14:2139979.
5. Acharjee A, Larkman J, Xu Y, Cardoso VR, Gkoutos GV. A random forest based biomarker discovery and power analysis framework for diagnostics research. *BMC Med Genom* 2020;13:178.
6. Breiman L Random forests. *Mach Learn* 2001;45:5–32.
7. Rynazal R, Fujisawa K, Shiroma H, Salim F, Mizutani S, Shiba S, et al. Leveraging explainable AI for gut microbiome-based colorectal cancer classification. *Genome Biol* 2023;24:21.
8. Acharjee A, Singh U, Choudhury SP, Gkoutos GV. The diagnostic potential and barriers of microbiome based therapeutics. *Diagnosis* 2022;9:411–20.
9. Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ Digit Med* 2019;2:69.
10. Petrosino JF. The microbiome in precision medicine: the way forward. *Genome Med* 2018;10:12.