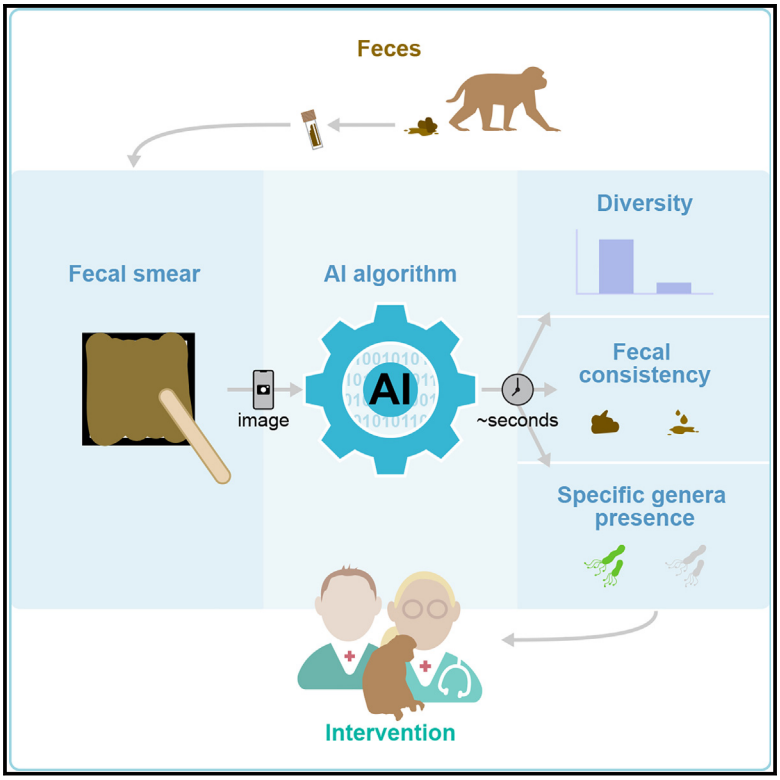
iScience

Article

A rapid, affordable, and reliable method for profiling microbiome biomarkers from fecal images

Graphical abstract



Highlights

- Successful proof-of-concept application for microbiome prediction
- High predictive accuracy for selection of bacterial genera
- High predictive accuracy for high-low microbial diversity and fecal consistency

Authors

Donghyeok Lee, Annemiek Maaskant, Huy Ngo, Roy C. Montijn, Jaco Bakker, Jan A.M. Langermans, Evgeni Levin

Correspondence

maaskant@bprc.nl (A.M.), evgeni.levin@horaizon.nl (E.L.)

In brief

Biological sciences; Microbiology; Microbiome

Lee et al., 2024, iScience *27*, 111310
December 20, 2024
https://doi.org/10.1016/j.isci.2024.111310

CellPress

iScience

CellPress
OPEN ACCESS

Article

A rapid, affordable, and reliable method for profiling microbiome biomarkers from fecal images

Donghyeok Lee,[1,4] Annemiek Maaskant,[2,3,4,*] Huy Ngo,[1] Roy C. Montijn,[1] Jaco Bakker,[2] Jan A.M. Langermans,[2,3] and Evgeni Levin[1,5,*]

[1]HORAIZON Technology BV, Marshallaan 2, 2625 GZ Delft, the Netherlands
[2]Biomedical Primate Research Centre, Lange Kleiweg 161, 2288 7 GJ Rijswijk, the Netherlands
[3]Department Population Health Sciences, Animals in Science and Society, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 2, 3584 CM Utrecht, the Netherlands
[4]These authors contributed equally
[5]Lead contact
*Correspondence: maaskant@bprc.nl (A.M.), evgeni.levin@horaizon.nl (E.L.)
https://doi.org/10.1016/j.isci.2024.111310

SUMMARY

Human and veterinary healthcare professionals are interested in utilizing the gut-microbiome as a target to diagnose, treat, and prevent (gastrointestinal) diseases. However, the current microbiome analysis techniques are expensive and time-consuming, and data interpretation requires the expertise of specialists. Therefore, we explored the development and application of artificial intelligence technology for rapid, affordable, and reliable microbiome profiling in rhesus macaques (*Macaca mulatta*). Tailor-made learning algorithms were created by integrating digital images of fecal samples with corresponding whole-genome sequenced microbial profiles. These algorithms were trained to identify alpha-diversity (Shannon index), key microbial markers, and fecal consistency from the digital images of fecal smears. A binary classification strategy was applied to distinguish between samples with high and low diversity and presence or absence of selected bacterial genera. Our results revealed a successful proof of concept for "high and low" prediction of diversity, fecal consistency, and "present or absent" for selected bacterial genera.

INTRODUCTION

Knowledge of the fundamental role of the gut microbiome in mammalian health and disease has significantly expanded over the last decades.[1] Gut microbiome bacteria are involved in several vital functions, including food fermentation, production of metabolites essential for growth and maintenance, and immune response.[2–4] Higher diversity within the microbiome enhances resistance to pathogenic invasions and improves host health.[2] While this diversity provides a broad overview of health in general, the presence of specific genera allows for a deeper understanding of microbiome structure, function, and impact on health.[2–7]

The presence or absence of specific bacteria in the gut microbiome can be used as biomarkers, serving as valuable tools for the diagnosis of disease.[7–9] Both human and veterinary healthcare professionals are interested in utilizing the microbiome as a target to diagnose, treat, and prevent diseases.[10–12] Existing DNA sequencing technologies, such as whole-genome sequencing or long-read sequencing provided by companies like Illumina or Oxford Nanopore, offer reliable analysis of the gut microbiome. However, these techniques are expensive and time consuming. Furthermore, interpretation of the data requires specialist expertise.

Rapidly evolving artificial intelligence (AI) technologies, including subsets, such as machine learning (ML) and deep neural network learning (DL), offer a potential alternative method for microbiome analysis. AI is a cost-effective method, and it generates data that is easy to interpret. ML and DL combined with metagenomic analysis can analyze high-dimensional data obtained from fecal images, allowing detailed characterization of microbiome dynamics and their association with diseases such as diarrhea.[13] This ML based technology facilitates the identification of complex relationships within the microbiome. It also employs non-linear multivariate models that account for interactions between microbial entities and the host.

In the present study, we explore the development and application of AI/ML technology for rapid, affordable, and reliable microbiome profiling in rhesus macaques (*Macaca mulatta*). This state-of-the-art AI-driven approach integrates digital images of fecal samples with tailor-made learning algorithms that are trained by corresponding metagenomic sequences to identify key microbial markers in seconds. In addition, the algorithm was trained to grade the fecal consistency of the sample. This experimental technology successfully translates features detected on a fecal image that are representative of microbiome composition, microbial diversity, and fecal consistency. This

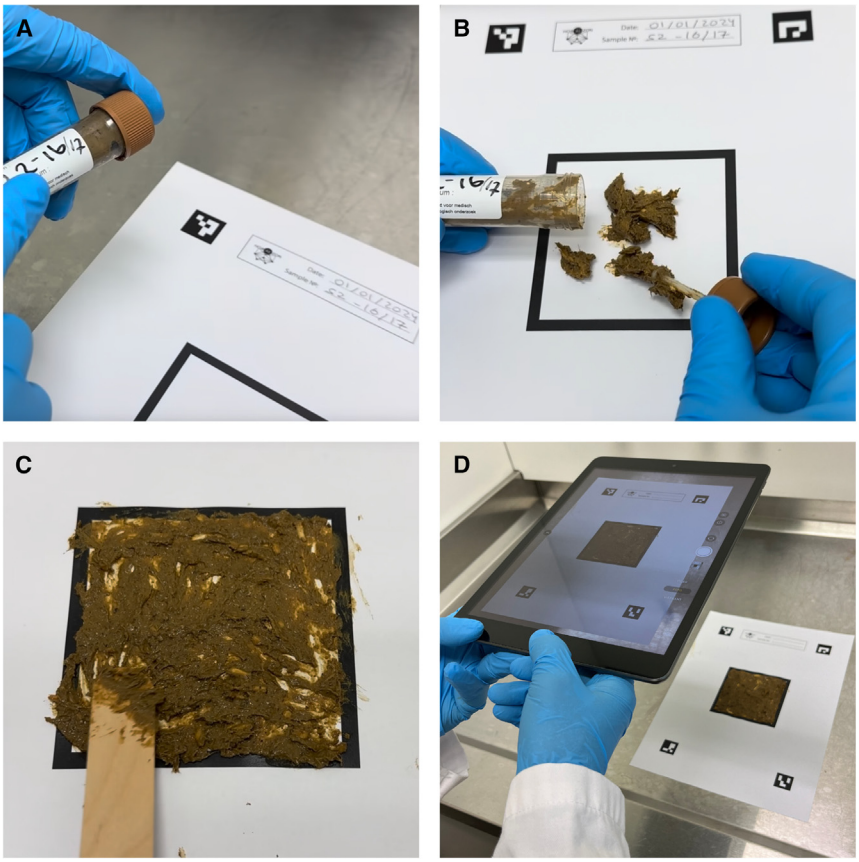


Figure 1. Fecal smear technique to acquire high quality digital images

(A) Paper template with printed square and fiducial marks on a flat surface, with the sample ID written on the top section.

(B) Feces applied within the borders of the square.

(C) Fecal layer smeared as evenly as possible using a spatula.

(D) Photographing the smear including the fiducial marks and sample identification.

proof of concept approach has the potential to improve microbiome diagnostics in both human and veterinary healthcare, paving the way for non-invasive, scalable, and cost-effective microbiome assessments.

RESULTS

Model input is a fecal smear image and is uploaded into experimental AI- algorithm

The input of the experimental AI algorithm is a digital image of a fecal smear. Figure 1 shows the smear technique that must be followed to acquire a high-quality image that can be uploaded.

Figure 2 shows the general overview of the process of our experimental model in comparison to the traditional microbiome screening method. Our method starts with an image of a fecal smear instead of DNA isolation and sequencing. Subsequently, the image was uploaded into the AI model. This AI model was trained to predict microbial composition, diversity, and fecal consistency.

Alpha-diversity and the presence of selected bacterial genera can be successfully predicted from a fecal smear image

The model training procedure involved seven distinct ViT (Vision Transformer) architectures pre-trained on ImageNet and fine-tuned on approximately 300 external fecal images. These external images were employed for model fine-tuning, to enhance generalizability and to avoid overfitting. The models were trained using a stratified random subsampling method and evaluated on study images. To reduce variability, the final prediction was derived by averaging the outputs of these models, forming an ensemble approach. This proved beneficial for the small dataset size, increasing robustness of the predictions.

For diversity prediction, our dataset comprised 66 images of fecal smears and their corresponding DNA sequences, which were analyzed to establish the microbial profiles. A binary classification strategy was used to distinguish between samples with "high" versus "low" diversity, based on the top and bottom 20% of the Shannon index. Thirteen samples were categorized as having high, and 12 samples as having low diversity, the remaining 41 samples were excluded from this analysis. Subsequently, the performance was evaluated on these 25 images.

The mean area under the receiver operator characteristic curve (AUROC) for the diversity prediction model is 0.73 ± 0.04 (Figure 3); data are presented as mean ± SD. The AUROC curve (Figure 3) illustrates the trade-off between sensitivity and specificity achieved by this model.

An initial prescreening of the genera present in the sequenced microbiome data was performed to identify the most promising subsets for detection. First, we selected the phylum *Firmicutes*, as this was the most abundant phylum in our dataset. In addition, for this phylum no differences were observed between fecal consistencies, as described elsewhere.[14] Subsequently, five potential genera were selected based on abundance to explore if the model could predict their presence. The detection task was translated into a binary classification task: determining whether the selected specific genera were "present" or "absent". The sample selection, sample size, and the evaluation procedure of the model were similar to the diversity prediction.

The results for the five selected genus presence detections, which showed different levels of accuracy, are shown in Figure 4. The models showed good predictive performance, with AUROC values of 0.83 ± 0.04 for *Coprococcus*, 0.80 ± 0.03 for *Phascolarctobacterium*, and 0.72 ± 0.02 for *Ruminococcus* (Figure 4).

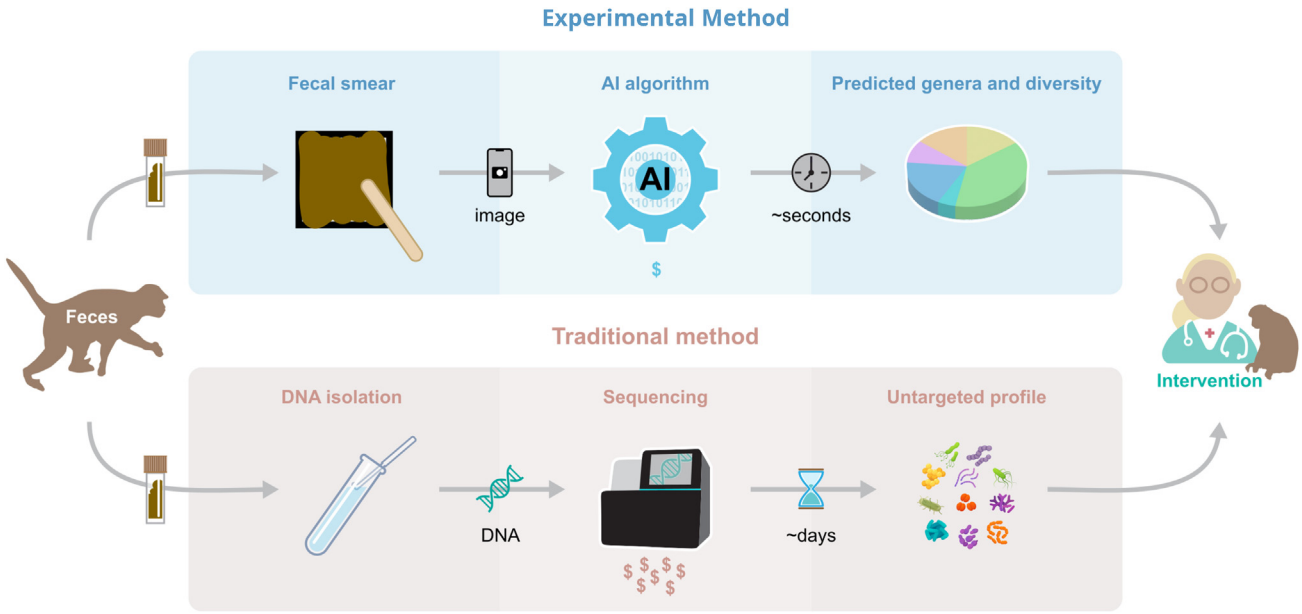


Figure 2. General overview of the process of the experimental AI/ML method in comparison to the conventional screening of the microbiome using sequencing techniques

For the experimental method, feces is smeared on a paper template, subsequently a photograph is taken and uploaded into the AI algorithm. Within seconds, targeted results such as diversity and genera are obtained. This targeted result is easier to use for clinicians to diagnose, intervene, or treat patients with gastrointestinal problems.

The model did not achieve sufficient AUROC values to predict the presence of *Intestinimonas* and *Oscillibacter*, 0.68 ± 0.02 and 0.67 ± 0.01, respectively.

Fecal consistency can be successfully predicted from a fecal smear image

The dataset consisted of 111 fecal images with corresponding Waltham scores. The images were ordered by this score in the top and bottom 25% and were categorized in "high" and "low" class, consisting of 27 fecal images each class. The remaining 57 samples were excluded from analysis.

This dataset was divided into 10 partitions via a 10-fold cross-validation (CV), with one partition used for evaluation and the others used for training. Data augmentation techniques were employed on the training partitions to synthetically expand the sample size, effectively increasing a small number of images to several hundred.[15]

The model achieved an AUROC of 0.72 ± 0.12 (Figure 5). Despite the use of data augmentation, we observed a higher standard deviation for the fecal consistency prediction compared to the diversity (SD 0.04) or genus presence prediction (SD range 0.02–0.04) models.

DISCUSSION

The study shows that our experimental AI/ML technology is a successful proof of concept in translating features, detected from a digital fecal image, into high-low microbial diversity (AUROC 0.73 ± 0.04), high-low fecal consistency (AUROC 0.72 ± 0.12), and predicting the presence of three different bacterial genera i.e., *Coprococcus* (AUROC 0.83 ± 0.04), *Phascolarctobacterium* (AUROC 0.80 ± 0.03), and *Ruminococcus* (AUROC 0.72 ± 0.02), in rhesus macaques. With only a limited number of available samples, these results are very promising for further development of the technology and hold major potential for rapid microbiome analysis of companion animals and livestock.

Diversity and genus presence detection

Our ensemble model successfully predicted microbial diversity, categorizing samples as having high or low diversity with an AUROC of 0.73 ± 0.04, as validated through a repeated stratified split process. The ensemble's aggregated performance surpassed that of its individual components for both diversity and genus presence prediction, demonstrating the robustness of this approach. Overall, the results suggest that while the model is effective, expanding the dataset will further improve its precision, allowing for more accurate distinctions in microbial diversity predictions.

Healthy gut microbial communities are in general characterized by high taxa diversity, high microbial gene richness, and a stable functional core microbiome.[3,16] A variety of acute and chronic disorders including inflammatory bowel diseases, obesity, and diabetes are associated with intestinal dysbiosis characterized by reduction of alpha-diversity.[3,4,8] Manor et al. showed in humans that gastrointestinal problems such as diarrhea were negatively associated with microbiome diversity.[6]

The model achieved accurate prediction of the presence of *Coprococcus*, *Phascolarctobacterium*, and *Ruminococcus*. This illustrates the model's ability to distinguish between the presence and absence of these bacterial taxa with high specificity and

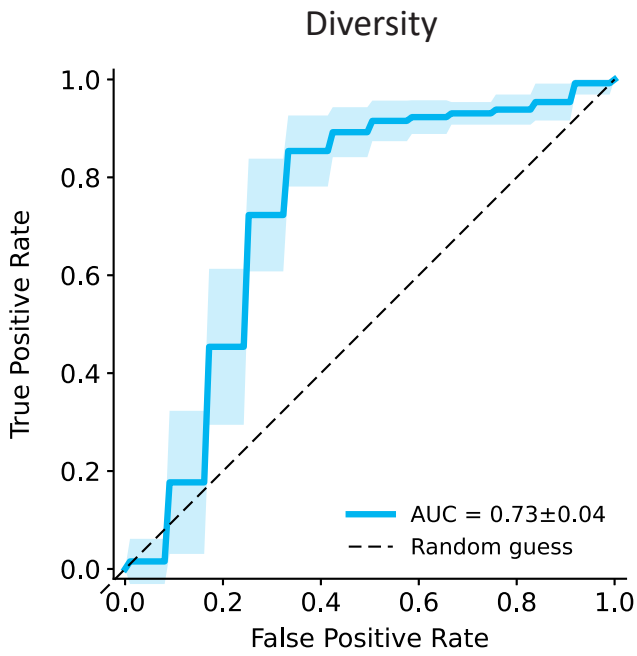


Figure 3. Alpha-diversity prediction model (Shannon-index) with binary classification strategy high versus low diversity showing good predictive accuracy

The area under the receiver operating characteristic curve (AUROC), data are presented as mean ± SD, for the diversity prediction model.

sensitivity. Specific genera are linked to various health outcomes; *Lactobacillus* spp and *Bifidobacterium* spp are known to have beneficial effects.[2,6] Therefore, genera prediction was of importance to be included in our AI/ML models. The three genera that were successfully predicted, i.e., *Coprococcus*, *Phascolarctobacterium*, and *Ruminococcus*, contribute significantly to the fermentation of dietary fibers, producing short chain fatty acids (SCFAs) that are vital for gut barrier integrity and metabolic and immune health.[17,18] Our technique could serve as a valuable tool for evaluating intestinal health and assessing related microbial biomarkers. However, prescreening for genus presence was necessary to manage the substantial computational costs associated with designing and running these predictive models. In future work, with an expanded dataset, the model is likely to predict many additional genera and potentially identify disease-specific biomarkers.

Fecal consistency prediction by Waltham score system

As traditional fecal consistency scoring is performed subjectively by observers, we aimed to reduce reliance on subjective human assessment by automating these observations.

The results (AUROC 0.72 ± 0.12) show the potential of DL techniques to predict high and low Waltham score from fecal image data, even with a relatively limited dataset comprising 27 samples in both classes. Although data augmentation effectively increases the training data, it may not fully address issues related to small sample sizes. This could explain the model's higher standard deviation compared to the other two models.

The total sample size (n = 111) accompanied with the unbalanced distribution across the Waltham scores was not sufficient for inclusion of additional Waltham classes. Yet, there is potential for enhancing the model's accuracy with the incorporation of a larger dataset, suggesting that increased data volume could significantly refine the predictive precision of this approach. In addition, the original 5-point scale system can be reinstated when sufficient data are available across all scales.

Diarrhea is a common clinical problem in the veterinary field. Different studies reported a diarrhea incidence of companion animals admitted to veterinary practices between 2.2% and 19.1%.[19–21] In non-human primate colonies, idiopathic diarrhea incidence is also high and can be up to 5–15%.[22] Multiple fecal scoring systems have previously been used in research studies and clinical applications for the objective characterization of clinical signs and responses to veterinary interventions.[23–28] In the current study, we used the WALTHAM feces scoring system because it is publicly available and included both written and visual descriptions of each score.[29] All scoring metrics have in common that they rely on pet owners' observation and interpretation. Cavett et al.[30] observed that disagreement in fecal scores occurred more frequently in pet owners compared to veterinarians. Yet, veterinarians have to rely on owner observations to follow up on progress after implementation of a therapy. Our model provides a proof of concept for an easy approach in which owners or caretakers can use AI to establish the fecal consistency of their pets.

Conclusion

This study demonstrates the successful application of AI/ML technology for microbiome analysis using digital fecal images, achieving proof of concept for rapid, affordable, and reliable microbial profiling in rhesus macaques. Our results show high predictive accuracy for key genera such as *Coprococcus*, *Phascolarctobacterium*, and *Ruminococcus*, as well as microbial diversity and fecal consistency.

Our approach represents an advancement in the field of microbiome analysis, particularly for veterinary applications. By bypassing the need for costly and time-intensive DNA sequencing, this method could enable more routine, accessible microbiome diagnostics in clinical settings. Furthermore, it opens new avenues for non-invasive health monitoring in both animals and humans, offering a scalable solution for microbiome-based interventions.

Future work should focus on expanding the dataset to further refine the model's accuracy and broaden its applicability to other species, including companion animals and livestock. Integrating multi-omics data could provide an even more comprehensive understanding of microbiome dynamics, improving health outcomes through targeted interventions.

Limitations of the study

While the results of this study demonstrate the potential of our method to predict microbial diversity, fecal consistency, and the presence of specific bacterial genera using AI/ML techniques, there are also limitations. Firstly, the relatively small

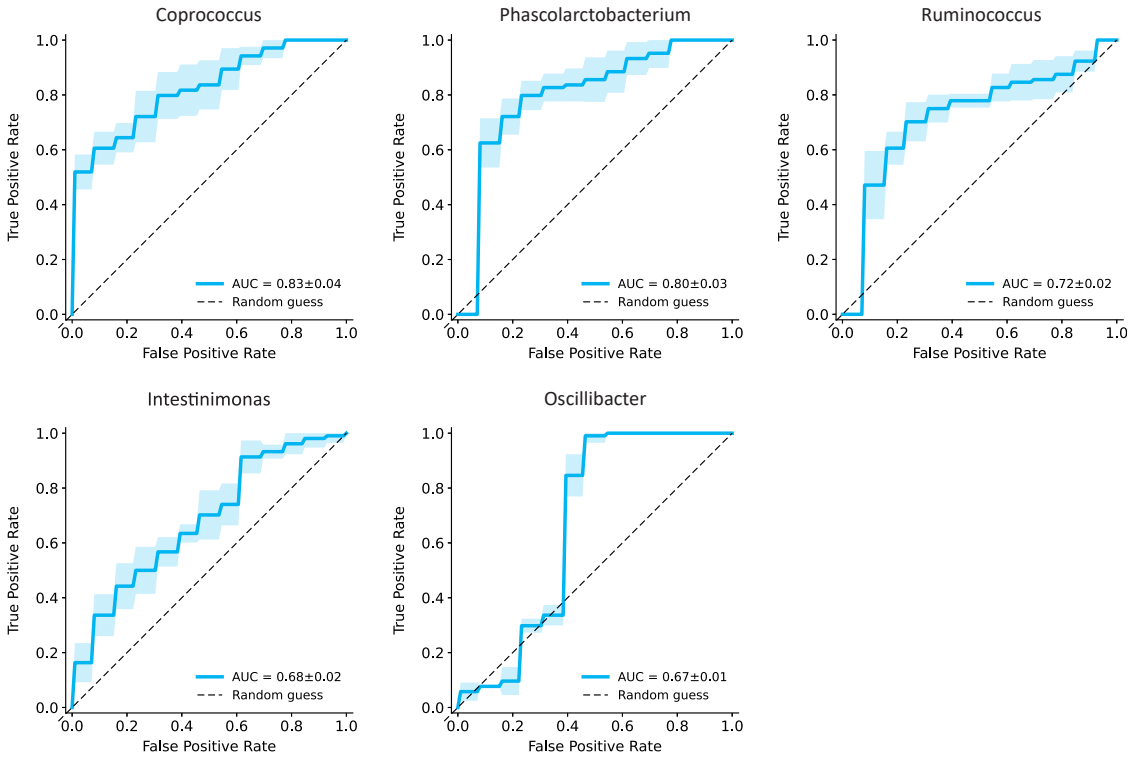


Figure 4. Good predictive model performance for bacterial genera: Coprococcus, Phascolarctobacterium, and Ruminococcus

The area under the receiver operator characteristic curve (AUROC), data are presented as mean ± SD, for the genus presence predictions. The model shows good predictive performances (AUC >0.7) for the genera Coprococcus, Phascolarctobacterium, and Ruminococcus. Less successful predictive performances were observed for Intestinimonas and Oscillibacter (AUC <0.7).

sample size of fecal images and corresponding DNA sequences constrains the generalizability of our results. With a limited dataset, the model's performance might not fully capture the variability and complexity of the gut microbiome across different individuals and conditions. Larger, more diverse datasets are necessary to train, validate, and refine our models further, ensuring robustness and applicability across broader populations. Yet, collecting large and gender balanced sample sizes is a common challenge in animal research using non-human primates and due to logistical and ethical considerations, expanding the dataset was not possible. Secondly, the fecal consistency scoring, although automated, was initially based on subjective human assessments. This could introduce variability. Lastly, while our AI models achieved good predictive accuracy, the black-box nature of these models poses challenges in interpretability and clinical acceptance. Future work will focus on enhancing the interpretability of AI models and integrating multi-omics data to provide a more comprehensive understanding of the gut microbiome dynamics and their clinical implications.

RESOURCE AVAILABILITY

Lead contact

Further information and request for resources should be directed to lead contact Evgeni Levin (evgeni.levin@horaizon.nl).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Metagenomic sequence data are publicly accessible through the European Nucleotide Archive (ENA). Accession number is listed in the key resources table.
- The analysis was performed using proprietary model founded on the same statistical principles as those pre-trained architecture available at GitHub as presented in the key resources table.
- Fecal images of the rhesus macaques are available upon reasonable request from the lead contact.

ACKNOWLEDGMENTS

The authors would like to thank the Animal Science department of the BPRC, especially the animal caretakers for excellent care of the animals and for assisting with sample collection and Nicole Heijmans for assisting with the fecal smears. Francisca van Hassel for figure design and Thea de Koning for her assistance with editing the manuscript. No external funding was received.

AUTHOR CONTRIBUTIONS

Conceptualization: D.L., A.M., and E.L.; methodology: D.L. and E.L.; software: D.L., H.N., and E.L.; formal analysis and validation: D.L., H.N., and E.L.; investigation: D.L., A.M., and J.B.; resources: D.L., A.M., J.A.M.L., J.B., R.C.M., and E.L.; data curation: D.L., A.M., and E.L.;writing-original draft preparation: D.L., A.M., and E.L.; writing-review and editing: J.A.M.L., J.B., H.N., R.M., and E.L.;

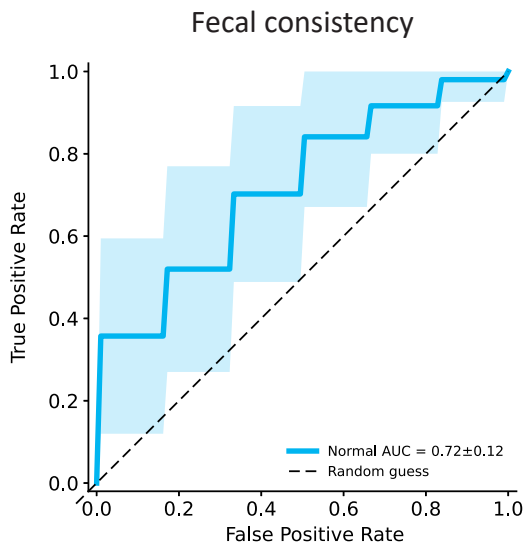


Figure 5. Fecal consistency prediction model based on Waltham score with binary classification strategy high versus low Waltham score shows good predictive accuracy

The area under the receiver operator characteristic curve (AUROC), data are presented as mean ± SD for high versus low Waltham score.

supervision: J.A.M.L., J.B., R.M., and E.L.; all authors have read and approved the final manuscript.

DECLARATION OF INTERESTS

E.L. is the founder of HORAIZON BV, Netherlands, which owns a patent related to the technology used (patent no. WO2023055238A1).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL SUBJECT DETAILS
  - Animals, husbandry and housing
  - Ethics approval
- METHOD DETAILS
  - Fecal sample collection and consistency scoring
  - Fecal imaging
  - DNA isolation and metagenomic sequencing and profiling
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data processing, visualization, and statistical analysis
  - Model training
  - Diversity prediction

Received: June 27, 2024
Revised: September 24, 2024
Accepted: October 30, 2024
Published: November 8, 2024

REFERENCES

1. Cenit, M.C., Matzaraki, V., Tigchelaar, E.F., and Zhernakova, A. (2014). Rapidly expanding knowledge on the role of the gut microbiome in health and disease. Biochim. Biophys. Acta *1842*, 1981–1992. https://doi.org/10.1016/j.bbadis.2014.05.023.
2. Hou, K., Wu, Z.X., Chen, X.Y., Wang, J.Q., Zhang, D., Xiao, C., Zhu, D., Koya, J.B., Wei, L., Li, J., and Chen, Z.S. (2022). Microbiota in health and diseases. Signal Transduct. Target. Ther. *7*, 135. https://doi.org/10.1038/s41392-022-00974-4.
3. Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. Nature *489*, 220–230. https://doi.org/10.1038/nature11550.
4. Pickard, J.M., Zeng, M.Y., Caruso, R., and Núñez, G. (2017). Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. Immunol. Rev. *279*, 70–89. https://doi.org/10.1111/imr.12567.
5. Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., Tobe, T., Clarke, J.M., Topping, D.L., Suzuki, T., et al. (2011). Bifidobacteria can protect from enteropathogenic infection through production of acetate. Nature *469*, 543–547. https://doi.org/10.1038/nature09646.
6. Manor, O., Dai, C.L., Kornilov, S.A., Smith, B., Price, N.D., Lovejoy, J.C., Gibbons, S.M., and Magis, A.T. (2020). Health and disease markers correlate with gut microbiome composition across thousands of people. Nat. Commun. *11*, 5206. https://doi.org/10.1038/s41467-020-18871-1.
7. Newman, T.M., Shively, C.A., Register, T.C., Appt, S.E., Yadav, H., Colwell, R.R., Fanelli, B., Dadlani, M., Graubics, K., Nguyen, U.T., et al. (2021). Diet, obesity, and the gut microbiome as determinants modulating metabolic outcomes in a non-human primate model. Microbiome *9*, 100. https://doi.org/10.1186/s40168-021-01069-y.
8. Gilbert, J.A., Quinn, R.A., Debelius, J., Xu, Z.Z., Morton, J., Garg, N., Jansson, J.K., Dorrestein, P.C., and Knight, R. (2016). Microbiome-wide association studies link dynamic microbial consortia to disease. Nature *535*, 94–103. https://doi.org/10.1038/nature18850.
9. Nagpal, R., Shively, C.A., Appt, S.A., Register, T.C., Michalson, K.T., Vitolins, M.Z., and Yadav, H. (2018). Gut Microbiome Composition in Non-human Primates Consuming a Western or Mediterranean Diet. Front. Nutr. *5*, 28. https://doi.org/10.3389/fnut.2018.00028.
10. Janse, M.E.M., Zinkweg, D.B., Larsen, O.F.A., and van de Burgwal, L. (2022). Innovations in the veterinary intestinal health field: A patent landscape analysis. One Health *15*, 100419. https://doi.org/10.1016/j.onehlt.2022.100419.
11. Pilla, R., and Suchodolski, J.S. (2019). The Role of the Canine Gut Microbiome and Metabolome in Health and Gastrointestinal Disease. Front. Vet. Sci. *6*, 498. https://doi.org/10.3389/fvets.2019.00498.
12. Turjeman, S., and Koren, O. (2022). Using the microbiome in clinical practice. Microb. Biotechnol. *15*, 129–134. https://doi.org/10.1111/1751-7915.13971.
13. Maasland, T., and Levin, E. (2023). Method and System for Analyzing Intestinal Microflora of a Subject (Google Patents).
14. Maaskant, A., Voermans, B., Levin, E., de Goffau, M.C., Plomp, N., Schuren, F., Remarque, E.J., Smits, A., Langermans, J.A.M., Bakker, J., and Montijn, R. (2024). Microbiome signature suggestive of lactose-intolerance in rhesus macaques (Macaca mulatta) with intermittent chronic diarrhea. Anim. Microbiome *6*, 53. https://doi.org/10.1186/s42523-024-00338-z.
15. Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. Commun. ACM *60*, 84–90. https://doi.org/10.1145/3065386.
16. Di Vincenzo, F., Del Gaudio, A., Petito, V., Lopetuso, L.R., and Scaldaferri, F. (2024). Gut microbiota, intestinal permeability, and systemic inflammation: a narrative review. Intern. Emerg. Med. *19*, 275–293. https://doi.org/10.1007/s11739-023-03374-w.
17. Morrison, D.J., and Preston, T. (2016). Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. Gut Microb. *7*, 189–200. https://doi.org/10.1080/19490976.2015.1134082.
18. Tan, J., McKenzie, C., Potamitis, M., Thorburn, A.N., Mackay, C.R., and Macia, L. (2014). The role of short-chain fatty acids in health and disease.

Adv. Immunol. *121*, 91–119. https://doi.org/10.1016/B978-0-12-800100-4.00003-9.

19. Jones, P.H., Dawson, S., Gaskell, R.M., Coyne, K.P., Tierney, A., Setzkorn, C., Radford, A.D., and Noble, P.J.M. (2014). Surveillance of diarrhoea in small animal practice through the Small Animal Veterinary Surveillance Network (SAVSNET). Vet. J. *201*, 412–418. https://doi.org/10.1016/j.tvjl.2014.05.044.
20. Lund, E.M., Armstrong, P.J., Kirk, C.A., Kolar, L.M., and Klausner, J.S. (1999). Health status and population characteristics of dogs and cats examined at private veterinary practices in the United States. J. Am. Vet. Med. Assoc. *214*, 1336–1341.
21. Rakha, G.M.H., Abdl-Haleem, M.M., Farghali, H.A.M., and Abdel-Saeed, H. (2015). Prevalence of common canine digestive problems compared with other health problems in teaching veterinary hospital. Vet. World *8*, 403–411. https://doi.org/10.14202/vetworld.2015.403-411.
22. Ardeshir, A., Oslund, K.L., Ventimiglia, F., Yee, J., Lerche, N.W., and Hyde, D.M. (2013). Idiopathic microscopic colitis of rhesus macaques: quantitative assessment of colonic mucosa. Anat. Rec. *296*, 1169–1179. https://doi.org/10.1002/ar.22727.
23. Allenspach, K., Wieland, B., Gröne, A., and Gaschen, F. (2007). Chronic enteropathies in dogs: evaluation of risk factors for negative outcome. J. Vet. Intern. Med. *21*, 700–708. https://doi.org/10.1892/0891-6640(2007)21[700:ceideo]2.0.co;2.
24. Grellet, A., Feugier, A., Chastant-Maillard, S., Carrez, B., Boucraut-Baralon, C., Casseleux, G., and Grandjean, D. (2012). Validation of a fecal scoring scale in puppies during the weaning period. Prev. Vet. Med. *106*, 315–323. https://doi.org/10.1016/j.prevetmed.2012.03.012.
25. Hernot, D.C., Biourge, V.C., Martin, L.J., Dumon, H.J., and Nguyen, P.G. (2005). Relationship between total transit time and faecal quality in adult dogs differing in body size. J. Anim. Physiol. Anim. Nutr. *89*, 189–193. https://doi.org/10.1111/j.1439-0396.2005.00544.x.
26. Meyer, H., Zentek, J., Habernoll, H., and Maskell, I. (1999). Digestibility and compatibility of mixed diets and faecal consistency in different breeds of dog. Zentralbl. Veterinarmed. A *46*, 155–165. https://doi.org/10.1046/j.1439-0442.1999.00201.x.
27. Propst, E.L., Flickinger, E.A., Bauer, L.L., Merchen, N.R., and Fahey, G.C., Jr. (2003). A dose-response experiment evaluating the effects of oligofructose and inulin on nutrient digestibility, stool quality, and fecal protein catabolites in healthy adult dogs. J. Anim. Sci. *81*, 3057–3066. https://doi.org/10.2527/2003.81123057x.
28. Rolfe, V.E., Adams, C.A., Butterwick, R.E., and Batt, R.M. (2002). Relationships between fecal consistency and colonic microstructure and absorptive function in dogs with and without nonspecific dietary sensitivity. Am. J. Vet. Res. *63*, 617–622. https://doi.org/10.2460/ajvr.2002.63.617.
29. Moxham, G. (2001). Waltham feces scoring system-A tool for veterinarians and pet[owners: How does your pet rate. Waltham Focus *11*, 24–25.
30. Cavett, C.L., Tonero, M., Marks, S.L., Winston, J.A., Gilor, C., and Rudinsky, A.J. (2021). Consistency of faecal scoring using two canine faecal scoring systems. J. Small Anim. Pract. *62*, 167–173. https://doi.org/10.1111/jsap.13283.
31. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., and Antiga, L. (2019). Pytorch: An imperative style, high-performance deep learning library. Adv. Neural. Inf. Process. Syst. *32*, 8026–8037.
32. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
33. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Preprint at arXiv. https://doi.org/10.48550/arXiv.2010.11929.
34. Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. Adv. Neural. Inf. Process. Syst. *25*, 1097–1105.
35. Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., and Dong, L. (2022). Swin Transformer V2: Scaling up Capacity and Resolution. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12009–12019.
36. Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). A Convnet for the 2020s. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11976–11986.
37. Wu, K., Zhang, J., Peng, H., Liu, M., Xiao, B., Fu, J., and Yuan, L. (2022). Tinyvit: Fast Pretraining Distillation for Small Vision Transformers (Springer), pp. 68–85.
38. Steiner, A., Kolesnikov, A., Zhai, X., Wightman, R., Uszkoreit, J., and Beyer, L. (2021). How to train your vit? data, augmentation, and regularization in vision transformers. Preprint at arXiv. https://doi.org/10.48550/arXiv.2106.10270.
39. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., and Clark, J. (2021). Learning Transferable Visual Models from Natural Language Supervision (PMLR)), pp. 8748–8763.
40. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.
41. Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. *20*, 257. https://doi.org/10.1186/s13059-019-1891-0.
42. Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Connor, R., Funk, K., Kelly, C., Kim, S., et al. (2022). Database resources of the national center for biotechnology information. Nucleic Acids Res. *50*, D20–D26. https://doi.org/10.1093/nar/gkab1112.
43. Lu, J., Breitwieser, F.P., Thielen, P., and Salzberg, S.L. (2017). Bracken: estimating species abundance in metagenomics data. PeerJ Computer Science *3*, e104.
44. Gruning, B., Dale, R., Sjodin, A., Chapman, B.A., Rowe, J., Tomkins-Tinch, C.H., Valieris, R., Koster, J., and Bioconda, T. (2018). Bioconda: sustainable and comprehensive software distribution for the life sciences. Nat. Methods *15*, 475–476. https://doi.org/10.1038/s41592-018-0046-7.
45. Hanley, J.A. (1989). Receiver operating characteristic (ROC) methodology: the state of the art. Crit. Rev. Diagn. Imaging *29*, 307–335.
46. Hanley, J.A., and McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology *143*, 29–36. https://doi.org/10.1148/radiology.143.1.7063747.
47. Spellerberg, I.F., and Fedor, P.J. (2003). A Tribute to Claude Shannon (1916-2001) and a Plea for More Rigorous Use of Species Richness, Species Diversity and the 'Shan-Non-Wiener' Index. Global Ecol. Biogeogr. *12*, 177–179. https://doi.org/10.1046/j.1466-822X.2003.00015.x.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Fecal samples (Macaca mulatta)	Biomedical Primate Research Center, Rijswijk, Netherlands	N/A
Fecal images	This paper	N/A
Deposited data		
Microbiome sequence data	Maaskant et al.[14]	https://www.ebi.ac.uk/ena/browser/view/PRJEB70928
Software and algorithms		
PyTorch	Paszke et al.[31]	2.3.1+cu121
Scikit-learn	Pedregosa et al.[32]	1.4.2
Python		3.11.9
Vision Transformers (ViT)	Dosovitskiy et al.[33]	https://doi.org/10.48550/arXiv.2010.11929.
ImageNet	Krizhevsky et al.[34]	https://www.image-net.org/
Technology patent WO2023055238A1	Maasland and Levin[13]	https://worldwide.espacenet.com/patent/search/family/083659093/publication/WO2023055238A1?q=pn%3DWO2023055238A1
Diversity: maxvit_tiny_tf_224	Liu et al.[35,36]	https://huggingface.co/timm/maxvit_tiny_tf_224.in1k
Diversity: maxvit_small_tf_224	Liu et al.[35,36]	https://huggingface.co/timm/maxvit_small_tf_224.in1k
Diversity: tiny_vit_5m_224	Wu et al.[37]	https://huggingface.co/timm/tiny_vit_5m_224.dist_in22k_ft_in1k
Diversity: tiny_vit_21m_224	Wu et al.[37]	https://huggingface.co/timm/tiny_vit_21m_224.dist_in22k_ft_in1k
Diversity: vit_small_patch8_224	Steiner et al.[38]	https://huggingface.co/timm/vit_small_patch8_224.dino
Diversity: vit_small_patch16_224	Steiner et al.[38]	https://huggingface.co/timm/vit_small_patch16_224.augreg_in21k
Diversity: vit_small_patch32_224	Steiner et al.[38]	https://huggingface.co/timm/vit_small_patch32_224.augreg_in21k
Coprococcus: maxxvitv2_rmlp_base_rw_224	Liu et al.[35]; Liu et al.[36]; Tu et al.[31]	https://huggingface.co/timm/maxxvitv2_rmlp_base_rw_224.sw_in12k
Phascolarctobacterium: maxvit_tiny_tf_224	Liu et al.[35]	https://huggingface.co/timm/maxvit_tiny_tf_224.in1k
Ruminococcus: vit_relpos_base_patch16_clsgap_224	Wu et al.[37]	https://huggingface.co/PrunaAI/vit_relpos_base_patch16_clsgap_224.sw_in1k-turbo-tiny-green-smashed/tree/main
Intestinimonas: vit_small_patch8_224	Steiner et al.[29]	https://huggingface.co/timm/vit_small_patch8_224.dino
Oscillibacter: vit_base_patch32_clip_quickgelu_224	Radford et al.[39]	https://huggingface.co/openai/clip-vit-base-patch32
Fecal consistency: tiny_vit_11m_224	Wu et al.[37]	https://huggingface.co/timm/tiny_vit_11m_224.dist_in22k
Other		
The Waltham Feces Scoring System	Moxham[29]	https://www.waltham.com/s3media/2020-05/waltham-scoring.pdf

EXPERIMENTAL MODEL SUBJECT DETAILS

Animals, husbandry and housing

The feces used for the images and the corresponding metagenomic sequence, utilized for the AI algorithms, originated from a broader research project focused on diarrhea and nutrition in rhesus macaques at the Biomedical Primate Research Center (BPRC). As previously detailed by Maaskant et al.[14] our research included 12 male and 2 female captive-bred rhesus macaques (*Macaca mulatta*). At the start of the study, all animals were between 3 and 7 years old and had a weight range between 6 and

12 kg. All animals were of Indian origin, bred and raised at the BPRC in social, multi-generational family structures. Feces samples were collected systematically between 2020 and 2022, following a standardized protocol.

All animals were socially housed. The enclosures of the animals featured freely accessible indoor and outdoor areas. The indoor areas had wood fiber bedding (Lignocel3–4, JRS, Rosenberg, Germany) and the outdoor areas had sand bedding. The enclosures were equipped with various enrichment structures such as climbing frames, beams, fire hoses, and platforms. The climate control was set to maintain an indoor temperature of approximately 18°C with a 12-h light-dark cycle. The diet consisted of standard commercial monkey pellets, supplemented with fruits, vegetables, and grains. Water was supplied *ad libitum* via automatic dispensers. Caretakers monitored the macaques twice daily for any signs of injury or illness, reporting any abnormalities to the veterinary team. Detailed electronic health records were maintained for each individual.

Ethics approval

All animals were housed in accordance with Dutch law and international ethical and scientific standards and guidelines (EU Directive 63/2010). All procedures and husbandry were compliant with the above standards and legislation. No interventions other than required for veterinary care were performed on these animals. Therefore, no approval from the competent authorities was required. Nevertheless, additional approval was obtained by the institutional animal welfare body (IvD 018A). The Biomedical Primate Research Center (BPRC) is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) International.

METHOD DETAILS

Fecal sample collection and consistency scoring

The methodology used in this study regarding the sample collection, scoring and sequencing, has been previously described in detail in Maaskant et al.[14] In summary, serial fecal samples were collected between 9 and 11 a.m. For microbiome analysis the feces were divided into three cryovials containing 1g each, and then frozen at −80°C for subsequent microbiome analysis. Samples used for fecal imaging were stored at −20°C.

The consistency scoring of each fecal sample was determined by means of the Waltham score, a Bristol Stool Scale equivalent, with a scale ranging from 1 (very hard and dry) to 5 (completely liquid).[29,30] The scoring was performed by two independent observers. The average Waltham score of the fecal samples determined the classification of each animal's fecal consistency.[29]

Fecal imaging

For fecal imaging, a specially designed paper template was used on which the feces were smeared. The template featured a distinct square printed in the center and fiducial marks at the corner edges. The paper was placed on a clean, stable and flat surface in a well-lit flow cabinet. Each fecal sample was gradually thawed and subsequently applied into the central square by using a disposable spatula. The fecal layer was applied as evenly as possible, completely covering but without extending the borders of the square. With an iPad camera (8-megapixel camera, 9th generation, model MK2K3NF/A Apple Inc. Cupertino, CA, USA) the entire template was captured in the image, including the square and fiducial marks. The images were subsequently uploaded into the AI-model.

DNA isolation and metagenomic sequencing and profiling

Total DNA was extracted from the samples using an Agowa/PurePrep protocol. To each 150 μL sample, 500 μL zirconium beads (0.1 mm) and 800 μL CD1 solution (DNeasy 96 Powersoil Pro QIAcube HT kit) were added. Cells were disrupted by bead beating twice for 2 min, with cooling on ice in between and afterward. After centrifugation for 6 min at 3,000 RPM, 350 μL supernatant was mixed with 300 μL Agowa binding buffer and 10 μL Agowa magnetic beads. Samples were further purified using the PurePrep 96 system (Molgen, The Netherlands) with two wash steps and a final elution step in 65 μL.

Libraries for whole-genome sequencing were prepared using the Illumina DNA prep protocol according to the instructions of Illumina (Illumina DNA Prep Reference Guide).

DNA concentrations were standardized across samples. After the tagmentation and clean-up steps, PCR-mediated standard indexed i5 and i7 adapters were added and the library was amplified. Next, the libraries were cleaned up and pooled. Whole-genome sequencing was performed using the Illumina MiSeq sequencer applying MiSeq V3 chemistry.

For metagenomic profiling we used fastp to preprocess the reads.[40] This filters out reads that are low-quality and too short. In addition, fastp trims out all reads from the front and the tail and cuts adapter sequences if present. Profiling of trimmed reads was done using kraken2 with a custom database built through kraken2's proprietary method.[41] We used a database consisting of Archaea, Bacteria, Plasmid, Virus, human, fungi and UniVec Core nucleotide/protein sequences. The latter is a subset of the NCBI-supplied database of vector, adapter, linker and primer sequences that may be contamination factors. The GRCh38 human genome assembly, as available on the NCBI database, was used to filter out potential contaminations of human reads, that may have ended up in the samples due to contamination during sample collection.[42] Parameter choices for kmer length were unchanged from the default settings of kraken2. After profiling, Bayesian re-estimation of abundance was performed using the Bracken tool.[43] All profiling tools were installed from bioconda channels.[44]

QUANTIFICATION AND STATISTICAL ANALYSIS

Data processing, visualization, and statistical analysis

To evaluate the performance of diversity, fecal consistency, and genera predictions, binary non-parametric receiver operating characteristic (ROC) curves were constructed. The area under the ROC (AUROC) curve was calculated as a measure of discriminative ability.[45,46] AUROC values of >0.9 were considered excellent, 0.8–0.9 as good, 0.7–0.8 as fair, and <0.7 as failed for model predictions. Analysis and visualization were performed using Python, utilizing PyTorch[31] and Scikit-learn.[32]

Model training

DL architectures for images are typically designed to extract features and define sharp boundaries for distinguishing discrete categories, rather than to predict continuous target variables. Models pre-trained on large external datasets, like ImageNet, are especially designed and trained for classification tasks.[15] Therefore, a DL-model with a classification approach was employed and trained on fecal images, to estimate microbial diversity, genus presence, and fecal consistency. Considering the limited sample size, we opted for binary classification rather than multi-class classification.

Vision Transformers (ViT)[33] were selected for this task because they are widely adopted for image analysis in various domains and have proven to be beneficial for capturing global patterns through the self-attention mechanism.[33] Since the fecal sample is evenly smeared on the template, it is crucial to capture these overall characteristics i.e., the global pattern, such as color, shape, and consistency, rather than specific details i.e., the local spatial pattern.

For diversity prediction, the model training procedure included seven distinct ViT architectures, each pre-trained on ImageNet and fine-tuned on approximately 300 external fecal images. The models were trained on 80% of these data, which were selected using a stratified random subsampling method. Subsequently, the performance was evaluated on the study images. This procedure was repeated 10 times to estimate the mean and standard deviation (SD) of the model's final performance. The final prediction was made by averaging the outputs from these models. This ensemble approach is particularly beneficial for small sample sizes, as it reduces variability in performance by averaging the outputs.

For the genus presence prediction task, the training and evaluation procedures, as well as the sample size, are identical to those used in the diversity prediction task. The key difference lay in the construction of the ensemble model. Specifically, for each genus prediction task, a particular ViT architecture was selected that demonstrated a significantly better performance compared to others. The selected ViT model was subsequently trained on different subsets of the training data to form the ensemble model. This approach leveraged the superior performance of the selected ViT architecture and enhanced the robustness of the predictions with an ensemble approach.

For the fecal consistency prediction, the external fecal images and the ensemble approach were not employed. Instead, the study images were divided into 10 partitions via a 10-fold cross-validation (CV), with one partition used for evaluation and the others used for training.

The detailed ViT architectures for each prediction task e.g., diversity, genus presence, fecal consistency are summarized in the key resources table.

Diversity prediction

The primary dataset of 66 fecal smears and their corresponding metagenomic profiles were used for diversity prediction. These profiles were quantitatively assessed using the Shannon Index, a metric of ecological diversity i.e., alpha-diversity.[47] Rarefaction has been applied to ensure comparable number of reads per sample.