REVIEW PAPER

Gut microbiota and artificial intelligence approaches: A scoping review

Ernesto Iadanza[1] · Rachele Fabbri[1] · Džana Bašić-ČiČak[2] · Amedeo Amedei[3] · Jasminka Hasic Telalovic[2]

Abstract

This article aims to provide a thorough overview of the use of Artificial Intelligence (AI) techniques in studying the gut microbiota and its role in the diagnosis and treatment of some important diseases. The association between microbiota and diseases, together with its clinical relevance, is still difficult to interpret. The advances in AI techniques, such as Machine Learning (ML) and Deep Learning (DL), can help clinicians in processing and interpreting these massive data sets. Two research groups have been involved in this Scoping Review, working in two different areas of Europe: Florence and Sarajevo. The papers included in the review describe the use of ML or DL methods applied to the study of human gut microbiota. In total, 1109 papers were considered in this study. After elimination, a final set of 16 articles was considered in the scoping review. Different AI techniques were applied in the reviewed papers. Some papers applied ML, while others applied DL techniques. 11 papers evaluated just different ML algorithms (ranging from one to eight algorithms applied to one dataset). The remaining five papers examined both ML and DL algorithms. The most applied ML algorithm was Random Forest and it also exhibited the best performances.

Keywords Microbiota · Microbiome · Artificial intelligence · Machine learning · Deep learning · Clinical decision support systems

1 Introduction

The idea for this work came from the recent authors' efforts in transnational scientific networks and research projects on the microbiome. In the last few years, the concept of applying computer-based algorithms for assessing medical problems has become a trending topic. The availability of large amounts of data, often referred to as *big data*, is a crucial enabling factor for this approach.

This article strives to provide a thorough overview of the use of Artificial Intelligence (AI) techniques in studying the gut microbiota and its role in the diagnosis and treatment of some important diseases.

The term *microbiota* refers to all microorganisms living in the same place, while microbiota habitat, the largest eukaryotic organism where the microbiota is located, is termed the *host* [1]. In animals, the site in which the largest amount of microorganisms resides is the gastro- digestive tract (mainly large intestine) [2].

A complex ecosystem consisting of bacteria, viruses, fungi and protozoans, is a human microbiota. It contains more than 100 times the human genome and gives us the functional properties we do not possess. It is composed by a number of genes (the microbiome). According to a recent estimation, the amount of bacteria contained in it could be higher than the amount of eukaryotic cells in the human body [3]: some 30 to 400 trillion microorganisms live in the gastrointestinal tract [4, 5]. Any surface exposed to the external environment, such as skin and mucosa (gastrointestinal, respiratory, and urogenital), is populated with the commensal microbiota, with the colon containing over 70% of all the bacteria in our body. An ecological connotation is thus assumed by the entire organism, which can now be redefined as a network of interactions and connections between various organisms (both eukaryotes and prokaryotes) [6].

✉ Ernesto Iadanza
   ernesto.iadanza@unifi.it

1 Department of Information Engineering, University
   of Florence, Via S. Marta 3, 50139 Florence, Italy

2 School of Science and Technology, University of Sarajevo,
   Sarajevo, Bosnia and Herzegovina

3 Department of Experimental and Clinical Medicine,
   University of Florence, Florence, Italy

Gut microbiota has essential dietary and metabolic functions (such as fermentation and digestion of carbohydrates, xenobiotic metabolism and vitamin synthesis) [7, 8]. It helps to safeguard against pathogenic issues. It is also important for the growth of the lymphoid tissue associated with the gut (GALTs) and the maturation of the innate and adaptive immune systems [9, 10]. The commensal bacteria are symbiotic, but they can cause a pathological state after translocation through the mucosa or in specific conditions such as immunodeficiency. The microbiota composition varies substantially between individuals, but it is also dynamic and susceptible to change. Moreover, the composition of the human microbiota is strictly personal, but the diversity in the structure of the bacterial population among the body sites is greater than it is between individuals. To date, although there have been over 50 bacterial phyla described, only two of them dominate the human gut normal flora: the *Bacteroidetes* and the *Firmicutes*, whereas *Actinobacteria*, *Proteobacteria*, *Fusobacteria*, *Verrucomicrobia* and *Cyanobacteria* appear in minor proportion [11].

Interestingly, a wide proportion of the human microbiota (about 70%) is composed of microbes that cannot be cultivated by common microbiological methods. Today, the advent of new molecular microbiota profiling tools, such as genomic Next-Generation Sequencing (NGS) and metagenomics shotgun sequencing, allows us to obtain more information about the microbiota impact in both healthy and pathological conditions [12]. As these techniques generate large amounts of data, they have sprung the development of bioinformatics techniques. However, the association between microbiota, diseases, and clinical relevance, is still a challenge. In this scenario, the advances in AI techniques, such as Machine Learning (ML) and Deep Learning (DL), can help clinicians in processing and interpreting information that can be extracted from these data sets.

ML and DL are two approaches to AI. However, defining the relation between them is not an easy and unambiguous task. Indeed, nowadays the field of AI is flourishing and the different techniques are continuously evolving. This results in a very dynamic scenario, within which the limits between one approach and another are evanescent. For example, in this paper DL is considered a subset of ML techniques [13] as Fig. 1 shows. However, this is not the only and most comprehensive representation of the field of AI and the reader can find other interpretations in the scientific literature.

1.1 Machine learning

Since the dawn of the technological era, computers' capabilities have been exploited for computer gaming and AI. The expression *machine learning* dates back to 1959 when Arthur Samuel used it for the first time in the IBM Journal of Research and Development [14]
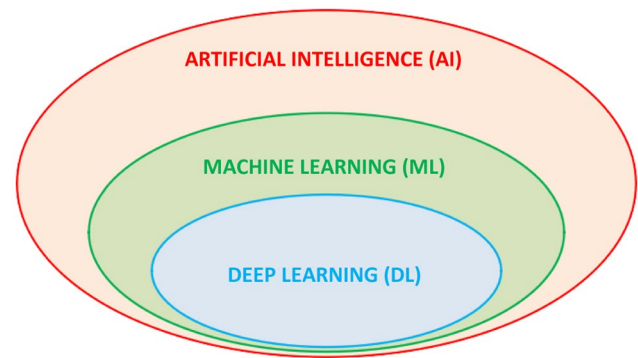
Fig. 1 The figure shows the relation between AI, ML and DL through a Venn diagram

ML approach is grounded on algorithms for solving problems of classification or prediction of patterns from data (regression models). It is common to talk about "learning from data" [15]. A typical distinction can be made between supervised or unsupervised ML algorithms. In the former approach, a number of input measures are used to predict the value of the output or for selecting one of the output classes. Unsupervised learning, instead, aims to "describe associations and patterns among a set of input measures" [15].

We can also say that in supervised algorithms we need some labels used during the training phase to instruct the machine on how to properly interpret the input. A typical label, in medicine, could be the diagnosis (healthy vs. ill, disease A vs. disease B, etc.). Conversely, the unsupervised approach is used to examine the data, searching for structures and patterns previously unknown (no labeled data needed).

As highlighted by us in a recent review article on machine learning methods for microbiome host trait predictions, several ML methods have been applied in the latest years for microbiome prediction. The microbiome data are often arranged into Operational Taxonomic Units (OTUs) [16], each collecting similar sequences representing a specific bacteria [17]. A taxonomy is used to represent the relationship among the microbes and each OTU. This relationship can be used in ML for the taxonomy-informed feature selection. This approach facilitates the selection of features to be used as input for ML algorithms. A detailed description of these methods and techniques is out of the scope of this work, but the reader can find a comprehensive overview in [17].

An example of the use of taxonomy can be found in the paper by Vangay et al. (2019). They created a publicly available repository, being comprised of 33 curated regression and classification tasks involving human microbiome data from 15 public datasets [18]. The repository has been developed as a powerful tool for two different types of

users: ML algorithms developers owing a limited knowledge of microbiome and, on the other hand, microbiome researchers who want to find new datasets for performing a meta-analysis.

Many different algorithms can be used, alone or in combination, to perform automated data analyses. Some are statistical models, such as the Logistic Regression (LR), and are often used in ML to predict the risk of developing a certain disease. This method cannot be considered a classifier, since it models the probability of output, given an input, but it can be used as such by setting cutoff thresholds [15]. Another algorithm for performing a regression analysis is the Least Absolute Shrinkage and Selection Operator (LASSO) [15]. Some more algorithms can be used both for regression and for classification. This is the case of Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT) and Gradient Boosting Decision Tree (GBDT), which are some of the most popular supervised algorithms.

In DTs, each internal node is associated with a variable, also known as property. Each arch linking a node to other children nodes represents a decision (e.g. a possible value for that variable). DTs are often used in ensemble methods [19], techniques that combine multiple models or algorithms to achieve better predictive performances. The concept of "boosting" is also common, and used to convert the weak learners to the strong ones (e.g. Logitboost) [20]. Recent evolution is represented by the C5.0 algorithm, improving DT with feature selection and reduced pruning error [21, 22].

GBDT combines the predictions from a series of decision trees which are used as the base learners. A new decision tree is trained at each step. EXtreme Gradient Boosting (XGBoost) is an open-source implementation of the gradient boost algorithm that uses a second-order gradient to guide the boosting process. Adaptive Boosting (ADA) is also a boosting algorithm which is used both for classification and regression problems, where the "weak learners" are decision trees with a single split [20].

Logistic Model Tree (LMT) can also be considered an ensemble method since it combines LR and DT [15]. Some very simple classifiers are the Naïve Bayes (NB) probabilistic classifiers, grounded on the well-known Bayes' theorem, and used in ML since the very first approaches. Despite their simplicity, they are still often used in clinical decision support systems and can provide good performances. A variation of NB is the Multinomial Naïve Bayes (MNB) classifier, where the features represent the frequencies with which some events have been generated by a multinomial distribution. Both the above classifiers are widely used for text classification purposes as well [23].

Another classification method for high-dimensional data, such as microarray data, is the Nearest Shrunken Centroids (NSC). It calculates centroids for each class and somehow compresses them to zero, using a thresholding technique [24].

1.1.1 Support vector machines

SVMs are among the most adopted and best-performing algorithms, used as supervised binary classifiers. They were first introduced by Boser et al. in [25]. The space of features is separated into two regions that correspond to the binary classes of the training data. This is done by using a linear hyperplane of equation [26]:

$$w^t x + b = 0 \qquad (1)$$

The above equation (Eq. 1) is obtained as a result of a training process that optimizes the geometric margin (to achieve the best separation between classes). When an elementary hyperplane is used, it is common to talk about "linear-SVM".

1.1.2 Artificial Neural Networks

Artificial Neural Networks (ANNs), often simply referred to as Neural Networks (NNs), have been extensively used for many years as automated classifiers [27]. Today, they are living a new youth thanks to their wide use in DL.

Inspired by the biological neurons, which are connected by synapses and neurotransmitters, NNs are made of at least two layers of nodes: one dedicated to host the input values and the other for the classifier's output. Very often they are also provided with one or more intermediate layers, between the input and the output ones, named *hidden layers* (see Fig. 2).

The connections between nodes belonging to different layers are obtained through numerical weights. A nonlinear activation function, named *transfer function*, rep- resents the action potential firing in the cell [28]. During the supervised training process, these weights are fine-tuned to achieve satisfactory classification performances, *i.e.* a robust connection between an array of input and the corresponding output (label). For training NNs, an iterative method is commonly used to optimize parametric functions, named Stochastic Gradient Descent (SGD) [29].

1.1.3 Random Forest

Random Forest (RF) is an ensemble method introduced by Breiman in 2001 [30]. As mentioned in Section 1.1, ensemble methods are the combination of several algorithms for classification or regression. The overall result is the enhancement of the performances. Thus, RF improves the predictive power of a single DT, by training multiple trees and combining their outputs [31]. The training process of

Fig. 2 Schematic representation of an ANN with a single hidden layer. Source: www.learndatasci.com
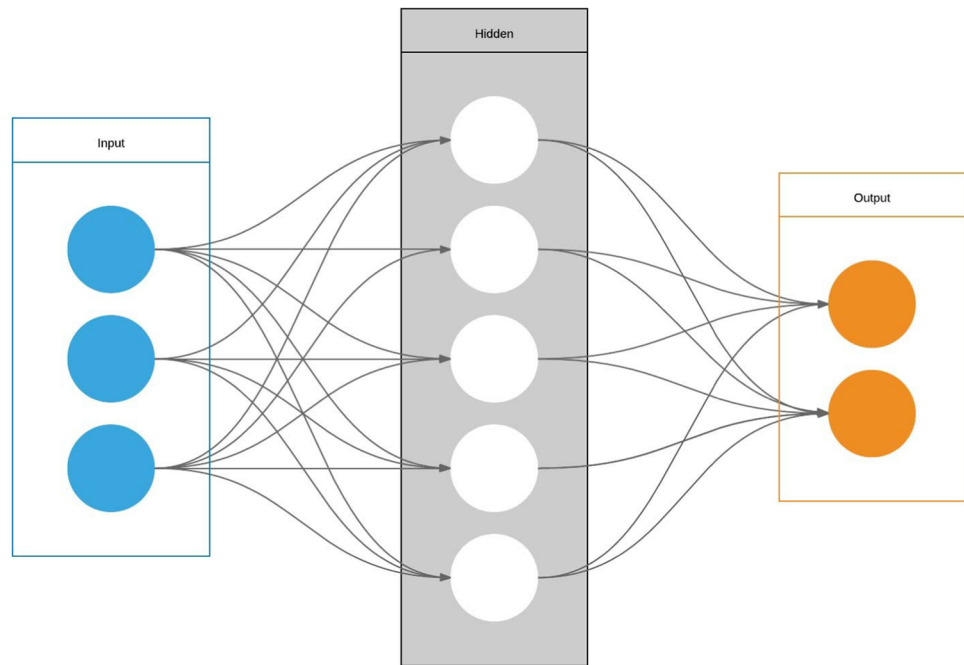
each tree selects a random subset from the training set. This procedure is called bagging, *i.e.* bootstrap aggregating. The final prediction is obtained as the average or the majority of the trees' estimations. RF splits the sample into groups, using features and associated thresholds. It shows very good performance, with only a few parameters to tune [32].

Thanks to their optimal characteristics, RFs are widely used and strongly applied also in the field of bioinformatics, metagenomics, and genomic data analysis [33–35].

1.2 Deep Learning

The term *Deep Learning* identifies a subset of ML algorithms, characterized by multiple layers of representation between input and output. DL has been developed to overcome the limitations of the ML techniques. Indeed, one of the issues arising with the use of ML is the so-called "Curse of dimensionality" [16]. It refers to the extremely growing complexity of the ML algorithms when the number of the dimensions of the input data is high. The DL techniques can be supervised or unsupervised as well.

The majority of DL algorithms are built upon ANNs, a class of learning algorithms composed of multiple interconnected layers for reproducing the way the brain processes and spreads information, as explained in Section 1.1.2. Thus, Deep Neural Network (DNN) refers to the high number of layers that compose the net- work of the DL algorithm. Such a high number of levels and units enable higher complexity of function representation if compared to ML.

Convolutional Neural Networks (CNNs) are a wide class of DNNs, often applied to image analysis. These networks apply convolution to at least one of their layers. Fioravanti et al. [36] have developed a particular DL approach, based on CNNs, for the classification of metagenomics data. This algorithm is called Phylogenetic CNN (Ph-CNN).

Multilayer Perceptron Neural Networks (MLPNNs), also called deep forward neural networks or feedforward neural networks, are a particular class of ANNs that perform unsupervised learning, with no feedback connections: the output of the model is not fed back to the network [13].

Deep Belief Networks (DBNs) are a subset of DNN algorithms. DBNs are characterized by the connection between different layers, but not between the units within each layer.

One last type of DNN is the Autoencoder Neural Network (AutoNN). An autoencoder is the combination of an encoder function and a decoder function. This algorithm reproduces the input in a more compressed representation (*i.e.* with a lower number of features needed), allowing a dimensionality reduction.

Zhou and Feng [37] proposed multi-Grained Cascade Forest (gcForest), a novel decision tree ensemble approach, consisting of a combination of a traditional ML algorithm and DL. It exhibits excellent performance in a broad range of tasks, being comparable to a DNN. In particular, gcForest is less sensitive to the changes of the network parameters (hyper-parameters) and thus it is more robust to hyperparameter settings if compared to other DL algorithms.

1.3 Article structure

Following an overview in Section 1 on the main ML and DL algorithms, Section 2 of this article provides detailed

information about the sources of information, and the methods for analyzing the results. Section 3 provides an accurate synthesis of the results, thoroughly discussed in Section 4. Finally, Section 5 provides the reader with considerations on the use of ML and DL for analysis of microbiome, from a clinical standpoint.

2 Materials and methods

2.1 Information sources and literature search

Two research groups have been involved in this Scoping Review, working in two different areas of Europe: Florence (Italy) and Sarajevo (Bosnia and Herzegovina). The scoping review has been conducted according to the PRISMA[1] Guidelines for Scoping Reviews [38].

The research strategy has been defined jointly by the two teams and the literature search has been performed in parallel on the main biomedical databases: MEDLINE (Ovid[2]) and PubMed.[3]

The keywords used for the search are consistent with the scope of the review and are the following: artificial intelligence, machine learning, deep learning, transfer learning, neural network/s, expert system/s, automatic classifier, deep network/s, classification, clustering, regression, prediction, microbiota, microbiome, gut, colorectal, colon, Chron.

2.2 Eligibility criteria

The papers to be included in the review had to describe the use of ML or DL methods, applied to the study of human gut microbiota. Moreover, the following limitations have been adopted: publication year from 2004 to current and English language only. In particular, limiting the publishing time allows focusing on the techniques and algorithms developed in the latest years.

2.3 Selection of sources of evidence

After the literature search, all the records retrieved have been screened using the eligibility criteria. Two reviewers (D.B. and R.F.) independently screened the titles and abstracts of all the records in the output of the literature search. The ones considered as non-pertinent to the extent of the review have been removed. The results obtained separately by the two reviewers have been compared: the articles which both

considered eligible for the study have been directly included in the list for full-text download. A third reviewer (A.A.), the immunologist participating as a clinical partner, was asked for a decision about those papers selected by only one of the two reviewers.

The full texts of the above-mentioned list of papers potentially eligible for the review have been downloaded. Once more, the two reviewers read the full texts and excluded some more papers not consistent with the objectives of the review. In this way, the final list of papers to be included in the review was created.

The reviewers made use of a collaborative worksheet in a shared Google Drive folder. The process described above is presented in a dedicated flowchart in Section 3.

2.4 Data Charting Form

The Data Charting Form (DCF) was developed to select the variables to be extracted and analyzed from the papers. The two groups independently created a DCF based on the reading of a small subset of papers. Then, the two forms have been compared and merged in the final DCF, used for the analysis.

The analyzed variables concern article characteristics (*i.e.* first author name, year and country of publication, publisher), method's scope and limitation, sample (*i.e.* type, size, age), application (*e.g.* microbiota body site, diseases considered), analysis technique (*e.g.* algorithms), validation and metric used to assess the performance. A detailed list of all the variables can be found in the tables in Section 3. As for the other steps, Data Collection was also performed by the two groups independently. Two DCFs were compiled and then the results were discussed to agree on the final data form to be included in the review.

2.5 Synthesis of methods for results handling

The combined DCF was the source of the results that are reported in this paper. We included a number of useful variables that describe the different research results in great detail. For readability purposes, the variables were spread across three tables.

Different metrics can be applied to evaluate the performance of a binary classifier. These metrics are presented in Table 1 and outlined in the following work by Flach [84]. Four outcomes of binary classification are used to produce constituents for defining more complex performance metrics: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Precision specifies the proportion of positive identifications that are correct.

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

True Positive Rate (TPR), or *Sensitivity*, or *Recall*, specifies the proportion of actual positives that were correctly labeled.

$$TPR/Sensitivity/Recall = \frac{TP}{TP + FN} \qquad (3)$$

Accuracy is a percentage of correctly identified samples (either as positive or negative) and is not a good metric if sizes of two groups are unbalanced.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (4)$$

Specificity measures classifier's ability to correctly label negatives.

$$Specificity = \frac{TN}{TN + FP} \qquad (5)$$

False Positive Rate (FPR) is calculated as the ratio between false positive and the total number of actual negatives.

$$FPR = 1 - TNR = \frac{FP}{TN + FP} \qquad (6)$$

The *Receiver Operating Characteristic (ROC) curve* is a plot of FPR (x-axis) vs. TPR (y-axis). The *Area Under the ROC Curve* (AUC), is a threshold invariant aggregated measure of binary classifier performances that takes into account all possible threshold values. AUC values range from 0 to 1.

F1-score is a metric that combines recall and precision using harmonic mean:

$$F1 - score = \frac{2 * Precision * TPR}{Precision + TPR} \qquad (7)$$

F1-macro is a metric used by Lo and Marculescu [48] and described as follows: "We estimate F1macro by calculating the accuracy for each class and then finding their unweighted mean".

The *Matthews Correlation Coefficient* (MCC) is, in essence, a correlation coefficient between the actual and predicted binary classifications and it assumes a value in $[-1, 1]$. A value of 1 represents a complete agreement of prediction and observation, 0 a random prediction, and -1 means opposite values of prediction and observation.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \qquad (8)$$

3 Results

3.1 Selection of sources of evidence

The output of the literature search sums up to 1109 articles in total. After the screening, based on title and abstract, 22 papers were assessed as eligible: their full-texts have been downloaded and read. Further 10 articles have been excluded after reading. Among these, the work by Zhou and Gallins [17] is a review. Although it has been excluded from the study, it provided four new articles [48, 52, 53, 85] to be added in full-text to the list of the eligible papers. Also, the paper [18] has been excluded as it is also a review. It presents a repository of classification and regression tasks from human microbiome datasets publicly available and, as for the previously mentioned review, this paper has been read and some of the studies it reviews have been considered for the assessment.

The remaining excluded articles were eliminated either because they did not apply ML or DL algorithms but just performed statistical analysis on microbiota data [86–88], or because the analysis was not focused on microbiota (e.g. enzymes profiles or metagenomics were analyzed). A detailed list of the reasons for exclusion for each of the ten articles is discussed in Section 4.
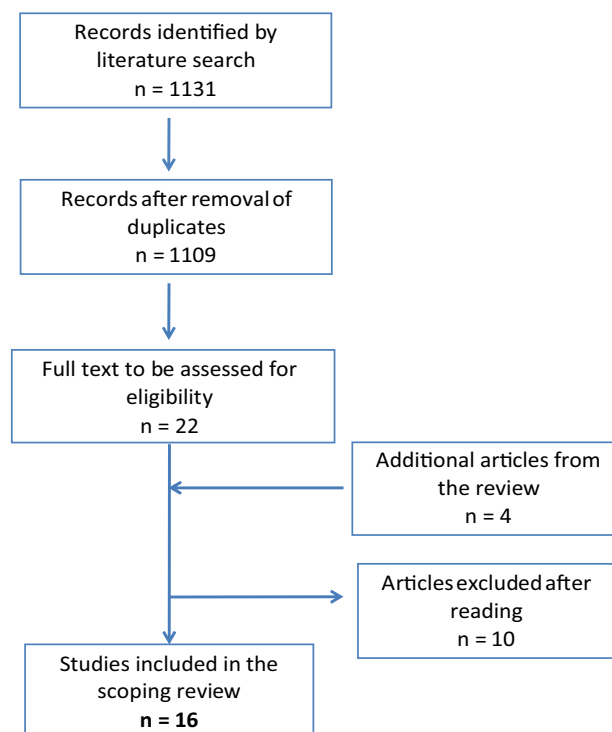
Fig. 3 The figure summarizes the process of selection of sources of evidence for the study

The final set of articles included and examined in this scoping review is made up of 16 articles. The process of selection of sources of evidence is reported in the flowchart in Fig. 3.

3.2 Synthesis of results

Table 1 gives a high-level overview for all the studies. For each study it states its type, cross-validation method, examined taxonomy level, and studied trait (if applicable). Table 2 summarizes all the data sets that were used in the examined studies while Table 3 summarizes all the different algorithms that have been used together with indicators of their performances.

4 Discussion

A total of 26 papers have been reviewed for this study. Out of those on the final list, 10 papers were identified as cases whose findings are not fit for this review. Table 4 lists those papers that were excluded together with the reasons for their exclusion. The remaining 16 papers were fully examined. The high-level data from Table 1 shows that 4 papers reported new methods that can be used to analyze the gut microbiota data while the remaining 12 papers applied the existing methods to analyze gut microbiota of humans with different traits. All papers performed some cross-validation methods: 10-folds cross-validation was the most common.

Table 1 High level information on studies

Paper	Study type	Cross- validation method	Taxonomy level	Trait
[39]	Dataset study	Tenfold, 10,000 times	Species	Fibromyalgia (FM)
[40]	Dataset study	Tenfold, 3 times	Genus	Stroke
[41]	Dataset study	Tenfold, 20 times	Species	Primary Sclerosing Cholangitis (PSC)
				Ulcerative Colitis (UC)
[36]	New methods	Fivefold, 10 times	Phylum	Crohn's Disease flare (iCDf)
				colon Crohn's Disease flare (cCDf)
				ileal Crohn's Disease remission (iCDr)
				colon Crohn's Disease remission (cCDr)
				Ulcerative Colitis flare (UCf)
				Ulcerative Colitis remission (UCr)
[42]	Dataset study	Tenfold	Species	Inflammatory Bowel Disease (IBD)
				Two body sites (skin vs gut)
[43]	New methods	80–20% split	Various	Three body sites
[44]	Dataset study	NA	Species	Liver Cirrhosis
		Fivefold		Type 2 Diabetes (T2D)
		Fivefold		Obesity
		NA		IBD
[45]	Dataset study	Tenfold	Genus	Obesity
[46]	New methods	Tenfold, 10 times	Species	5 body sites
				24 environments
				Colorectal cancer
[47]	New methods	Tenfold	Genus	Classification of body site
				Classification of subjects
				Classification of disease states
[48]	Dataset study	Tenfold	Species	Colorectal Cancer (CRC) and Colorectal Adenoma (CRA)
[49]	Dataset study	NA	Genus	Crohn's Disease (CD)
[50]	Dataset study	Tenfold, 10 times	Species	Colorectal Polyps
[51]	Dataset study	Leave-one-out	Species	Colorectal Adenoma
				Colorectal Cancer
[52]	Dataset study	Fivefold, 5 times	Genus	Type 2 diabetes (T2D)
				Liver Cirrhosis
				Rheumatoid Arthritis (RA)
[53]	Dataset study	Leave-one-out	Genus	Oral malodour

Table 2 Description of data sets that each study used

Paper	Dataset	Samples	Site	Study design	Cases	Controls
[39]	[39]	156	Gut	Case–control	77	79
[40]	[54]	233	Gut	Case–control	141	92
	[55]	58	Gut	Case–control	29	29
	[56]	88	Gut	Case–control	49	39
	[57]	224	Gut	Case–control	191	33
	[58]	404	Gut	Case–control	263	141
	[59]	73	Gut	Case–control	29	44
[41]	[41]	48	Saliva	Case–control	24	24
	[41]	40	Saliva	Case–control	16	
[36]	[60]	82	Gut	Case–control	44	38
	[60]	54	Gut	Case–control	16	
	[60]	97	Gut	Case–control	59	
	[60]	56	Gut	Case–control	18	
	[60]	79	Gut	Case–control	41	
	[60]	82	Gut	Case–control	44	
[42]	[42]	112	Gut	Case–control	56	56
	[42]	94	Skin and Gut	Classification	NA	NA
[43]	[61]	1,967	Gut, oral, skin	Classification	NA	NA
[44]	[62]	232	Gut	Case–control	118	114
	[63]	344	Gut	Case–control	170	174
	[64]	253	Gut	Case–control	164	89
	[65]	110	Gut	Case–control	25	85
[45]	[45]	66	Gut	Classification	NA	NA
[46]	[66]	4,485	Different body sites	Case–control	NA	NA
	[67]	10,101	Different environments	Classification	NA	NA
	[68]	182	Gut	Case–control	90	92
	[69]	60	Gut	Case–control	30	30
[47]	[70]	552	Different body sites	Classification	NA	NA
	[70]	357	Different body sites	Classification	NA	NA
	[66]	1,025	Different body sites	Classification	NA	NA
	[70]	140	Different body sites	Classification	NA	NA
	[71]	104	Different body sites	Classification	NA	NA
	[71]	98	Skin	Classification	NA	NA
	[72]	1,025	Gut	Case–control	500	500
	[73]	200	Gut	Case–control	172	28
[48]	[48]	141	Gut	Case–control	97	52
	NA	141	Gut	Case–control	80	61
[49]	[49]	217 (45subjects)	Gut	Cohort study	NA	NA
[50]	[50]	552	Gut (231 rectal swab, 183 stool, 138 biopsy)	Case–control	316	236
[51]	CA,CC	509	Gut	Case–control	79	235
	[63, 68, 69, 73–82]		Gut	Case–control	195	
[52]	[65]	170	Gut	Case–control	170	383
	[62]	130	Gut	Case–control	130	
	[83]	123	Gut	Case–control	123	
[53]	[53]	90	Saliva	Case–control	45	45

Table 3 Studies' algorithms and their performance

Paper	Data set—trait	Algorithm	Metric	Performance
[39]	[39]—FM	LASSO	AUROC	0.8720
		SVM	AUROC	0.8780
[40]	[54]—Stroke	LogitBoost	Accuracy	0.9832
		SVM	Accuracy	0.9875
		KNN	Accuracy	0.9655
		LMT	Accuracy	0.9832
	[55]—JIA	LogitBoost	Accuracy	0.9852
		SVM	Accuracy	0.9828
		KNN	Accuracy	0.9727
		LMT	Accuracy	0.9895
	[55]—ME/CFS	LogitBoost	Accuracy	0.9693
		SVM	Accuracy	0.9646
		KNN	Accuracy	0.9473
		LMT	Accuracy	0.9626
	[55]—HIV1	LogitBoost	Accuracy	0.9971
		SVM	Accuracy	0.9885
		KNN	Accuracy	0.9727
		LMT	Accuracy	0.9866
	[57]—CRC	LogitBoost	Accuracy	0.9684
		SVM	Accuracy	0.9559
		KNN	Accuracy	0.9028
		LMT	Accuracy	0.9593
	[57]—MS	LogitBoost	Accuracy	0.9828
		SVM	Accuracy	0.9808
		KNN	Accuracy	0.9641
		LMT	Accuracy	0.9818
[41]	[41]—PSC	RF	AUROC	0.7423
	[41]—UC	RF	AUROC	0.875
[36]	[60]—iCDF	Ph-CNN	MCC	0.944
		RF	MCC	0.89
		linear-SVM	MCC	0.941
		MLPNN	MCC	0.901
	[60]—eCDF, cCDF	Ph-CNN	MCC	0.572
		RF	MCC	0.72
		linear-SVM	MCC	0.595
		MLPNN	MCC	0.648
	[60]—iCDr	Ph-CNN	MCC	0.91
		RF	MCC	0.82
		linear-SVM	MCC	0.907
		MLPNN	MCC	0.837
	[60]—cCDr	Ph-CNN	MCC	0.89
		RF	MCC	0.765
		linear-SVM	MCC	0.88
		MLPNN	MCC	0.788

Paper	Data set—trait	Algorithm	Metric	Performance
	[60]—UCf	Ph-CNN	MCC	0.628
		RF	MCC	0.76
		linear-SVM	MCC	0.565
		MLPNN	MCC	0.606
	[60]—UCr	Ph-CNN	MCC	0.346
		RF	MCC	0.4
		linear-SVM	MCC	0.408
		MLPNN	MCC	0.489
[42]	[42]—IBD	SVM	Accuracy	0.81
		RF	Accuracy	0.81
		NSC	Accuracy	0.79
		LR	Accuracy	0.78
	[42]—skin/gut	SVM	Accuracy	0.98
		RF	Accuracy	0.99
		NSC	Accuracy	0.99
		LR	Accuracy	1.00
[43]	[61]—3 body sites	RNN	Accuracy	0.8313
		RNN	Accuracy	0.8418
		DBN	Accuracy	0.9685
		DBN	Accuracy	0.9721
		MLPNN	Accuracy	0.8844
		RF	Accuracy	0.9914
		CNN-1D	Accuracy	0.9467
		CNN-2D	Accuracy	0.9947
[43]	[62]—Cirrhosis	NA	NA	NA
	[63]—T2D	SVM	Accuracy	0.643
		RF	Accuracy	0.657
		XGBoost	Accuracy	0.64
		gcForest	Accuracy	0.655
		autoNN	Accuracy	0.663
	[64]—Obesity	SVM	Accuracy	0.637
		RF	Accuracy	0.648
		XGBoost	Accuracy	0.635
		gcForest	Accuracy	0.64
		autoNN	Accuracy	0.624
	[65]—IBD	NA	NA	NA
[45]	[45]—Obesity	Ensemble method	Sensitivity	0.64
			Specificity	0.86
		SVM	Sensitivity	0.57
			Specificity	0.83
		C5.0	Sensitivity	0.71
			Specificity	0.86

Table 3 (continued)

Table 3 (continued)

Paper	Data set—trait	Algorithm	Metric	Performance
[46]	[66]—5 body sites	DT	AUROC	0.985
		NB	AUROC	0.995
		RF	AUROC	0.999
	[67]—24 environments	DT	AUROC	0.96
		NB	AUROC	0.949
		RF	AUROC	0.999
	[68]—CRC	DT	AUROC	0.678
		NB	AUROC	0.721
		RF	AUROC	0.795
	[69]—CRC	DT	AUROC	0.657
		NB	AUROC	0.583
		RF	AUROC	0.975
[47]	[70]—different body sites	SVM	F1-macro	0.78
		RF	F1-macro	0.73
		GB	F1-macro	0.74
		MNB	F1-macro	0.66
		LR1	F1-macro	0.41
		LR2	F1-macro	0.17
	[70]—different body sites	SVM	F1-macro	0.63
		RF	F1-macro	0.58
		GB	F1-macro	0.48
		MNB	F1-macro	0.49
		LR1	F1-macro	0.26
		LR2	F1-macro	0.24
	[66]—different body sites	SVM	F1-macro	0.97
		RF	F1-macro	0.97
		GB	F1-macro	0.95
		MNB	F1-macro	0.95
		LR1	F1-macro	0.94
		LR2	F1-macro	0.93
	[70]—different subjects	SVM	F1-macro	0.88
		RF	F1-macro	0.87
		GB	F1-macro	0.74
		MNB	F1-macro	0.76
		LR1	F1-macro	0.16
		LR2	F1-macro	0.19
	[71]—different subjects	SVM	F1-macro	0.94
		RF	F1-macro	1.00
		GB	F1-macro	0.91
		MNB	F1-macro	0.98
		LR1	F1-macro	.60
		LR2	F1-macro	.58

Table 3 (continued)

Paper	Data set—trait	Algorithm	Metric	Performance
	[71]—different subjects	SVM	F1-macro	0.68
		RF	F1-macro	0.63
		GB	F1-macro	0.55
		MNB	F1-macro	0.50
		LR1	F1-macro	0.17
		LR2	F1-macro	0.17
	[72]—IBD	SVM	F1-macro	0.68
		RF	F1-macro	0.57
		GB	F1-macro	0.65
		MNB	F1-macro	0.43
		LR1	F1-macro	0.47
		LR2	F1-macro	0.43
	[73]—PDX	SVM	F1-macro	0.29
		RF	F1-macro	0.28
		GB	F1-macro	0.35
		MNB	F1-macro	0.18
		LR1	F1-macro	0.15
		LR2	F1-macro	0.15
[48]	[48]—CRC/CRA	BN	AUROC	0.93
		RF	AUROC	0.94
		SL	AUROC	0.975
		LMT	AUROC	0.975
	NA	BN	AUROC	0.858
		RF	AUROC	0.86
		SL	AUROC	0.714
		LMT	AUROC	0.762
[49]	[49]—CD	RF	AUROC	0.87
[50]	[50]—Colorectal polyps	NB	AUROC	0.86
		NN	AUROC	0.87
[51]	[63, 68, 69, 73–82]—CRC/CRA	RF	AUROC	0.80
[52]	[62, 65]—T2D, Cirrhosis, RA	KNN	F1-score	0.8602
		LR	F1-score	0.9142
		RF	F1-score	0.8341
		SVM	F1-score	0.9138
		GBDT	F1-score	0.8741
		SGD	F1-score	0.836
		ADA	F1-score	0.8959
[53]	[53]—Oral malodour	Deep learning	Accuracy	0.9670
		SVM	Accuracy	0.7890

Table 4 Excluded papers with the reason for exclusion

Paper	Reason for exclusion
[86]	No AI algorithms, only statistical analysis performed
[17]	A review for non-experts
[18]	Technical note
[89]	ML was used solely for the unknown OTU discovery (taxonomic assignment)
[87]	No AI algorithms, only statistical analysis performed
[88]	No AI algorithms, only statistical analysis performed
[90]	ML performed on a database of substrate molecules, not microbiome data
[85]	Did not use human microbiome
[48]	Findings reported from [48] version of the paper
[91]	ML solely used for feature subset selection

The majority of the examined papers used species level as the final taxa resolution. This is consistent with findings in the literature: species is the last taxonomic level at which 16 s rRNA sequencing technique is accurate. As for the traits examined, there is a great variance, but datasets of individuals with either Chron's disease or Colorectal cancer have been examined multiple times.

As reported in Table 2, most of the studies examined multiple datasets, counting from one to eight. All but two papers (that examined saliva microbiota) analyzed gut microbiota data. Classification of different body sites is a problem that has been examined in two studies. While three other papers included different classification problems, and one used cohort study as the design method, the overwhelming majority was either fully or partially based on the case–control design method. This is due to AI techniques being suitable for the design of computational classifiers that can distinguish the samples from case and control groups with high probability. The total number of samples in a single dataset varied from 40 to 10, 101. In the articles that used case–control study design, the case groups contained 16 to 500 samples, while the control groups were made up of 24 to 500 samples.

A different range of AI algorithms was applied in the reviewed papers. Some of them were from the ML group while the others were from DL. 11 papers evaluated just different ML algorithms (ranging from one to eight algorithms applied to one dataset). The remaining five papers examined both ML and DL algorithms.

The most often applied ML algorithm was Random Forest. This algorithm had also most often the best performance reported. For the papers that applied both ML and DL algorithms, it is inconclusive which ones performed better.

When it comes to reporting the performance of AI algorithms on microbiome data, as summarized in Table 3, the metrics that were reported most frequently were *AUC* and *Accuracy*. Additional metrics include *Sensitivity*, *Specificity*, *MCC*, *F1-score* and *F1-macro*.

5 Conclusion

It is commonly accepted that the "ecosystem" microbiome plays a central role in health and disease development. The human microbiome consists of promising biomarkers for various pathological states and there is an overflow of metagenomics results. Translating these data into clinical practice is now a big challenge for the future. Microorganisms and host cells communicate by producing and sharing metabolites and generating metabolic networks that we can use to develop meta- metabolic network models. Studying network biology using ML represents a great opportunity for exploring the "human health condition".

Some models could be used for understanding the microbial-host interplay, as well as for predicting and gaining insights into the synergistic and dysbiotic connections. Some models could be used to inspect how the abnormal growth of a specific microbial species might perturb the metabolic balance of the ecosystem by secreting beneficial metabolites that promote health or, conversely, toxic ones that could damage the host tissues. Some models could also be used to foster the development of innovative diagnostic applications. The huge amount of data produced by these models is often referred to as *big data*.

However, such a big amount of data needs to be reported in an intelligible way. Each prediction allows for more extensive analysis, which in turn may let clinicians make informed and accurate decisions. Using a method for explaining individual classifier decisions for complex microbiota analysis may assist in performing treatment management for every single patient. This approach can also help the physician in improving his/her clinical expertise (with new and fine stratification of patients' sub-types), thus opening new perspectives on personalized therapy.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval N/A.

References

1. Rosenberg E, Zilber-Rosenberg I. Microbes drive evolution of animals and plants: the hologenome concept. mBio. 2016;7(2):e01395. https://doi.org/10.1128/mBio.01395-15.
2. Shapira M. Gut microbiotas and host evolution: scaling up symbiosis. Trends Ecol Evol. 2016;31(7):539–49. https://doi.org/10.1016/j.tree.2016.03.006.
3. Sender R, Fuchs S, Milo R. Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. Cell. 2016;164(3):337–40. hhttps://doi.org/10.1016/j.cell.2016.01.013.
4. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI. The long-term stability of the human gut microbiota. Science. 2013;341(6141):1237439. https://doi.org/10.1126/science.1237439.
5. Seedorf H, Griffin NW, Ridaura VK, et al. Bacteria from diverse habitats colonize and compete in the mouse gut. Cell. 2014;159(2):253–66. https://doi.org/10.1016/j.cell.2014.09.008.
6. van Baalen M, Huneman P. Organisms as ecosystems/ecosystems as organisms. Biol Theory. 2014;9(4):357–60. https://doi.org/10.1007/s13752-014-0194-7.
7. Rowland I, Gibson G, Heinken A, et al. Gut microbiota functions: metabolism of nutrients and other food components. Eur J Nutr. 2018;57(1):1–24. https://doi.org/10.1007/s00394-017-1445-8.
8. Martin FP, Sprenger N, Yap IK, et al. Panorganismal gut microbiome-host metabolic crosstalk. J Proteome Res. 2009;8(4):2090–105. https://doi.org/10.1021/pr801068x.
9. Cebra JJ. Influences of microbiota on intestinal immune system development. Am J ClinNutr. 1999;69(5):1046S-1051S. https://doi.org/10.1093/ajcn/69.5.1046s.
10. Niccolai E, Boem F, Emmi G, Amedei A. The link "cancer and autoimmune diseases" in the light of microbiota: evidence of a potential culprit. Immunol Lett. 2020;222:12–28. https://doi.org/10.1016/j.imlet.2020.03.001.
11. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. Microbiology: diversity of the human intestinal microbial flora. Science. 2005;308(5728):1635–8. https://doi.org/10.1126/science.1110591.
12. Espinoza JL, Kotecha R, Nakao S. Microbe-induced inflammatory signals triggering acquired bone marrow failure syndromes. Front Immunol. 2017;8:186. https://doi.org/10.3389/fimmu.2017.00186.
13. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning, vol. 1. Cambridge: MIT press; 2016.
14. Samuel AL. Some studies in machine learning using the game of checkers. IBM J Res Dev. 1959;3(3):210–29.
15. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning, vol. 1, no. 10. New York: Springer series in statistics; 2001.
16. Schmitt S, Tsai P, Bell J, et al. Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. ISME J. 2012;6(3):564–76. https://doi.org/10.1038/ismej.2011.116.
17. Zhou YH, Gallins P. A review and tutorial of machine learning methods for microbiome host trait prediction. Front Genet. 2019;10:579. https://doi.org/10.3389/fgene.2019.00579.
18. Vangay P, Hillmann BM, Knights D. Microbiome learning Repo (ML Repo): a public repository of microbiome regression and classification tasks. Giga-Science. 2019;8(5):1–12. https://doi.org/10.1093/gigascience/giz042.
19. Zhou ZH. Ensemble methods: foundations and algorithms. New York: CRC Press; 2012.
20. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann Stat. 2000;28(2):337–407.
21. Pandya R, Pandya J. C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. Int J Comput Appl. 2015;117(16):18–21.
22. Iadanza E, Mudura V, Melillo P, et al. An automatic system supporting clinical decision for chronic obstructive pulmonary disease. Health Technol. 2020;10:487–98. https://doi.org/10.1007/s12553-019-00312-9.
23. Aghila G. A survey of naïve bayes machine learning approach in text document classification.arXiv preprint arXiv:1003.1795; 2010.
24. Tibshirani R, Hastie T, Narasimhan B, Chu G. Class prediction by nearest shrunken centroids, with applications to DNA microarrays. Stat Sci. 2003;18:104–17.
25. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory; 1992. p. 144–52.
26. Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press; 2000.
27. Michie D, & Spiegelhalter DJ. Machine learning. Neural and Statistical Classification. New York: Ellis Horwood; 1994.
28. Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol. 1952;117(4):500–44. https://doi.org/10.1113/jphysiol.1952.sp004764.
29. Goldt S, Advani M, Saxe AM, Krzakala F, Zdeborová L. Dynamics of stochastic gradient descent for two-layer neural networks in the teacher-student setup. In: Advances in neural information processing systems; 2019. p. 6981–91.
30. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
31. Rokach L. Decision forest: twenty years of research. Inf Fusion. 2016;27:111–25.
32. Genuer R, Poggi JM, Tuleau-Malot C, Villa-Vialaneix N. Random forests for big data. Big Data Research. 2017;9:28–46. https://doi.org/10.1016/j.bdr.2017.07.003.
33. Qi Y. Random forest for bioinformatics. In: Ensemble machine learning; 2012. p. 307–23. Boston: Springer.

34. Zhu Q, Zhu Q, Pan M, Jiang X, Hu X, He T. The phylogenetic tree based deep forest for metagenomic data classification. In: 2018 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE, p. 279–82.

35. Chen X, Ishwaran H. Random forests for genomic data analysis. Genomics. 2012;99(6):323–9. https://doi.org/10.1016/j.ygeno.2012.04.003.

36. Fioravanti D, Giarratano Y, Maggio V, et al. Phylogenetic convolutional neural networks in metagenomics. BMC Bioinforma. 2018;19(Suppl 2):49. https://doi.org/10.1186/s12859-018-2033-5.

37. Zhou ZH, Feng J. Deep forest. arXiv preprint arXiv:1702.08835; 2017.

38. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467–73. https://doi.org/10.7326/M18-0850.

39. Minerbi A, Gonzalez E, Brereton NJB, et al. Altered microbiome composition in individuals with fibromyalgia. Pain. 2019;160(11):2589–602. https://doi.org/10.1097/j.pain.0000000000001640.

40. Bang S, Yoo D, Kim SJ, Jhang S, Cho S, Kim H. Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data. Sci Rep. 2019;9(1):10189. https://doi.org/10.1038/s41598-019-46249-x.

41. Iwasawa K, Suda W, Tsunoda T, et al. Dysbiosis of the salivary microbiota in pediatric-onset primary sclerosing cholangitis and its potential as a biomarker. Sci Rep. 2018;8(1):5480. https://doi.org/10.1038/s41598-018-23870-w.

42. Eck A, Zintgraf LM, de Groot EFJ, et al. Interpretation of microbiota-based diagnostics by explaining individual classifier decisions. BMC Bioinforma. 2017;18(1):441. https://doi.org/10.1186/s12859-017-1843-1.

43. Reiman D, Metwally A, Dai Y. Using convolutional neural networks to explore the microbiome. Conf Proc IEEE Eng Med BiolSoc. 2017;2017:4269–72. https://doi.org/10.1109/EMBC.2017.8037799.

44. LaPierre N, Ju CJ, Zhou G, Wang W. MetaPheno: a critical evaluation of deep learning and machine learning in metagenome-based disease prediction. Methods. 2019;166:74–82. https://doi.org/10.1016/j.ymeth.2019.03.003.

45. Fernández-Navarro T, Díaz I, Gutiérrez-Díaz I, et al. Exploring the interactions between serum free fatty acids and fecal microbiota in obesity through a machine learning algorithm. Food Res Int. 2019;121:533–41. https://doi.org/10.1016/j.foodres.2018.12.009.

46. Oudah M, Henschel A. Taxonomy-aware feature engineering for microbiome classification. BMC Bioinforma. 2018;19(1):227. https://doi.org/10.1186/s12859-018-2205-3.

47. Lo C, Marculescu R. MetaNN: accurate classification of host phenotypes from metagenomic data using neural networks. BMC Bioinforma. 2019;20(Suppl 12):314. https://doi.org/10.1186/s12859-019-2833-2.

48. Ai L, Tian H, Chen Z, Chen H, Xu J, Fang JY. Systematic evaluation of supervised classifiers for fecal microbiota-based prediction of colorectal cancer. Oncotarget. 2017;8(6):9546–56. https://doi.org/10.18632/oncotarget.14488.

49. Braun T, Di Segni A, BenShoshan M, et al. Individualized dynamics in the gut microbiota precede Crohn's disease flares. Am J Gastroenterol. 2019;114(7):1142–51. https://doi.org/10.14309/ajg.0000000000000136.

50. Dadkhah E, Sikaroodi M, Korman L, et al. Gut microbiome identifies risk for colorectal polyps. BMJ Open Gastroenterol. 2019;6(1):e000297. https://doi.org/10.1136/bmjgast-2019-000297.

51. Shah MS, DeSantis TZ, Weinmaier T, et al. Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer. Gut. 2018;67(5):882–91. https://doi.org/10.1136/gutjnl-2016-313189.

52. Wu H, Cai L, Li D, et al. Metagenomics biomarkers selected for prediction of three different diseases in Chinese population. Biomed Res Int. 2018;2018:2936257. https://doi.org/10.1155/2018/2936257.

53. Nakano Y, Suzuki N, Kuwata F. Predicting oral malodour based on the microbiota in saliva samples using a deep learning approach. BMC Oral Health. 2018;18(1):128. https://doi.org/10.1186/s12903-018-0591-6.

54. Yin J, Liao SX, He Y, et al. Dysbiosis of gut microbiota with reduced Trimethylamine-N-Oxide level in patients with large-artery atherosclerotic stroke or transient ischemic attack. J Am Heart Assoc. 2015;4(11):e002699. https://doi.org/10.1161/JAHA.115.002699.

55. Di Paola M, Cavalieri D, Albanese D, et al. Alteration of fecal microbiota profiles in Juvenile idiopathic arthritis. Associations with HLA-B27 allele and disease status. Front Microbiol. 2016;7:1703. https://doi.org/10.3389/fmicb.2016.01703.

56. Giloteaux L, Goodrich JK, Walters WA, Levine SM, Ley RE, Hanson MR. Reduced diversity and altered composition of the gut microbiome in individuals with myalgic encephalomyelitis/chronic fatigue syndrome. Microbiome. 2016;4(1):30. https://doi.org/10.1186/s40168-016-0171-4.

57. Noguera-Julian M, Rocafort M, Guillén Y, et al. Gut microbiota linked to sexual preference and HIV Infection. EBioMedicine. 2016;5:135–46. https://doi.org/10.1016/j.ebiom.2016.01.032.

58. Baxter NT, Ruffin MT 4th, Rogers MA, Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med. 2016;8(1):37. https://doi.org/10.1186/s13073-016-0290-3.

59. Jangi S, Gandhi R, Cox LM, et al. Alterations of the human gut microbiome in multiple sclerosis. Nat Commun. 2016;7:12015. https://doi.org/10.1038/ncomms12015.

60. Sokol H, Leducq V, Aschard H, et al. Fungal microbiota dysbiosis in IBD. Gut. 2017;66(6):1039–48. https://doi.org/10.1136/gutjnl-2015-310746.

61. Caporaso JG, Lauber CL, Costello EK, et al. Moving pictures of the human microbiome. Genome Biol. 2011;12(5):R50. https://doi.org/10.1186/gb-2011-12-5-r50.

62. Qin N, et al. Alterations of the human gut microbiome in liver cirrhosis. Nature. 2014;513(7516):59–64.

63. Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490(7418):55–60. https://doi.org/10.1038/nature11450.

64. Le Chatelier E, Nielsen T, Qin J, et al. Richness of human gut microbiome correlates with metabolic markers. Nature. 2013;500(7464):541–6. https://doi.org/10.1038/nature12506.

65. Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59–65. https://doi.org/10.1038/nature08821.

66. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207–14. https://doi.org/10.1038/nature11234.

67. Henschel A, Anwar MZ, Manohar V. Comprehensive meta-analysis of ontology annotated 16S rRNA profiles identifies beta diversity clusters of environmental bacterial communities. PLoS Comput Biol. 2015;11(10):e1004468. https://doi.org/10.1371/journal.pcbi.1004468.

68. Zeller G, Tap J, Voigt AY, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol. 2014;10(11):766. https://doi.org/10.15252/msb.20145645.

69. Zackular JP, Rogers MA, Ruffin MT 4th, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. Cancer Prev Res (Phila). 2014;7(11):1112–21. https://doi.

org/10.1158/1940-6207.CAPR-14-0129. Epub 2014 Aug 7. PMID: 25104642; PMCID: PMC4221363.

70. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. Science. 2009;326(5960):1694–7. https://doi.org/10.1126/science.1177486.

71. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Forensic identification using skin bacterial communities. Proc Natl Acad Sci U S A. 2010;107(14):6477–81. https://doi.org/10.1073/pnas.1000162107.

72. Gevers D, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014;15(3):382–92.

73. Ahn J, Sinha R, Pei Z, et al. Human gut microbiome and risk for colorectal cancer. J Natl Cancer Inst. 2013;105(24):1907–11. https://doi.org/10.1093/jnci/djt300.

74. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PLoS ONE. 2012;7(6):e39743. https://doi.org/10.1371/journal.pone.0039743.

75. Wu N, Yang X, Zhang R, et al. Dysbiosis signature of fecal microbiota in colorectal cancer patients. Microb Ecol. 2013;66(2):462–70. https://doi.org/10.1007/s00248-013-0245-9.

76. Weir TL, Manter DK, Sheflin AM, Barnett BA, Heuberger AL, Ryan EP. Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. PLoS ONE. 2013;8(8):e70803. https://doi.org/10.1371/journal.pone.0070803.

77. Brim H, Yooseph S, Zoetendal EG, et al. Microbiome analysis of stool samples from African Americans with colon polyps. PLoS ONE. 2013;8(12):e81352. https://doi.org/10.1371/journal.pone.0081352.

78. Mira-Pascual L, Cabrera-Rubio R, Ocon S, et al. Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. J Gastroenterol. 2015;50(2):167–79. https://doi.org/10.1007/s00535-014-0963-x.

79. Flemer B, Lynch DB, Brown JM, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut. 2017;66(4):633–43. https://doi.org/10.1136/gutjnl-2015-309595.

80. Sobhani I, Tap J, Roudot-Thoraval F, et al. Microbial dysbiosis in colorectal cancer (CRC) patients. PLoS ONE. 2011;6(1):e16393. https://doi.org/10.1371/journal.pone.0016393.

81. Chen HM, Yu YN, Wang JL, et al. Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma. Am J Clin Nutr. 2013;97(5):1044–52. https://doi.org/10.3945/ajcn.112.046607.

82. Goedert JJ, Gong Y, Hua X, et al. Fecal microbiota characteristics of patients with colorectal adenoma detected by screening: a population-based study. EBioMedicine. 2015;2(6):597–603. https://doi.org/10.1016/j.ebiom.2015.04.010.

83. Zhang X, Zhang D, Jia H, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. Nat Med. 2015;21(8):895–905. https://doi.org/10.1038/nm.3914.

84. Flach P. Machine learning: the art and science of algorithms that make sense of data. Cambridge: Cambridge University Press; 2012.

85. Moitinho-Silva L, Steinert G, Nielsen S, et al. Predicting the HMA-LMA status in marine sponges by machine learning. Front Microbiol. 2017;8:752. https://doi.org/10.3389/fmicb.2017.00752.

86. Ferrocino I, Ponzo V, Gambino R, et al. Changes in the gut microbiota composition during pregnancy in patients with gestational diabetes mellitus (GDM). Sci Rep. 2018;8(1):12216. https://doi.org/10.1038/s41598-018-30735-9.

87. Hu Y, Peng J, Li F, Wong FS, Wen L. Evaluation of different mucosal microbiota leads to gut microbiota-based prediction of type 1 diabetes in NOD mice. Sci Rep. 2018;8(1):15451. https://doi.org/10.1038/s41598-018-33571-z.

88. Le Roy CI, Bowyer RCE, Castillo-Fernandez JE, et al. Dissecting the role of the gut microbiota and diet on visceral fat mass accumulation. Sci Rep. 2019;9(1):9758. https://doi.org/10.1038/s41598-019-46193-w.

89. Vervier K, Mahé P, Tournoud M, Veyrieras JB, Vert JP. Large-scale machine learning for metagenomics sequence classification. Bioinformatics. 2016;32(7):1023–32. https://doi.org/10.1093/bioinformatics/btv683.

90. Sharma AK, Jaiswal SK, Chaudhary N, Sharma VK. A novel approach for the prediction of species-specific biotransformation of xenobiotic/drug molecules by the human gut microbiota. Sci Rep. 2017;7(1):9751. https://doi.org/10.1038/s41598-017-10203-6.

91. Ditzler G, Morrison JC, Lan Y, Rosen GL. Fizzy: feature subset selection for metagenomics. BMC Bioinforma. 2015;16:358. https://doi.org/10.1186/s12859-015-0793-8.