Benha Journal of Applied Sciences (BJAS) **print: ISSN 2356–9751**
Vol. (10) Issue (7) (2025), (87- 95) **online: ISSN 2356–976x**

https://bjas.journals.ekb.eg/
engineering sciences

Gut Microbiome Profile Prediction for Nonalcoholic Fatty Liver Disease Patients Based on Artificial Intelligence Techniques

Bahaa S.Rabi [1] , Hanim Khtar[2] , Ali H.El-Far[3]and Reda Albadawy[4]
[1] Electrical Engineering Dept., Benha Faculty of Engineering, Benha University, Benha, Egypt
[2] Pathology Dept., Faculty of Veterinary Medicine, Benha University, Benha, Egypt
[3] Biochemistry Dept., Faculty of Veterinary Medicine, Damanhour University, Damanhour, Egypt
[4] Gastrointestinal Diseases Dept., Faculty of Medicine, Benha University, Benha, Egypt
E-mail: bahaaalkhtaib@gmail.com

Abstract
The current conclusions and hypotheses on the roles of the gut microbiome in the pathophysiological understanding of different phenotypes of nonalcoholic Fatty Liver Disease (NAFLD) patients have greatly attracted research in multiple academic and research institutions. In this study, the gut microbiome profiles of different families and genetic categories that live in NAFLD patients were combined and saved in one dataset as a result of fecal and blood sample analysis. There were distinct gut microbiome patterns in the fecal and blood samples of obese individuals with NAFLD compared to lean individuals. Another important feature used in the dataset is body Mass Index (BMI), which is classified into lean (BMI <25 kg/m^2) and overweight (30>BMI >25 kg/m^2). One of the important branches of Artificial Intelligence is Machine Learning (ML).ML techniques have a great contribution to real-world applications. This approach implements Rules between Gut-Micro profiles for NAFLD (Lean obese). Besides the ML prediction model, Random Forest was deployed to predict the Gut-Micro profile for both lean and obese individuals in the case of NAFLD patients. The testing accuracy in the proposed model is more than 99%, which is considered an excellent performance parameter in gut microbiome investigation.

Keywords: Bioinformatics, Body Mass Index, Machine Learning, and Random Forest..

1. Introduction

Non-alcoholic Fatty Liver Disease (NAFLD) is the leading reason for chronic liver disease [1,2]. Patients with NAFLD are at an increased risk of developing cirrhosis, hepatocellular carcinoma, cardiovascular events, non-hepatocellular carcinoma malignancies, and increased mortality [3]. The connection between obesity and NAFLD seems more complex, regarding the absence of NAFLD in obese individuals without any metabolic anomalies and the presence of NAFLD in lean individuals. Although NAFLD is more common in certain ethnicities, such as Asians, it affects about 10% of the Western population [4].

According to a recent statistical analysis by Riazi et al. [5], 46.9 instances of NAFLD are present for every one thousand person-years.

According to [6], the accepted histological definition of grade 1 steatosis is hepatic steatosis of 5% or more; steatosis below 5% is considered normal. This is one of the diagnostic criteria for NAFLD. Ultrasound, currently the most popular diagnostic method for fatty liver, has a bottom limit of 30% at which steatosis can be reliably detected.

Apart from the imaging methods evaluated by these guidelines, computed tomography (CT) can also be used to identify fatty liver. By assessing the radiodensity, fat may be assessed (in Hounsfield units) since fat attenuates less than water when utilizing X-ray-based techniques, making the liver look darker on pictures.

The gut microbiota, comprising trillions of microorganisms, is a key regulator in the development of obesity, diabetes, and metabolic syndrome [7,8]. Current investigations suggest that gut microbiota plays a role in the development of NAFLD. It is believed that gut-derived endogenous alcohol may contribute to the pathogenesis of nonalcoholic steatohepatitis [9]. In NAFLD, changes in gut microbiome composition are relevant, such as a decrease in Firmicutes, an increase in Bacteroidetes in nonalcoholic steatohepatitis, and an increase in Ruminococcaceae in significant fibrosis [10]. Most previous studies on gut microbiome composition have focused on individuals with obesity. Therefore, Body Mass Index (BMI) is considered a significant determinant of changes in gut microbiota [11].

In this proposed work, a new approach was suggested to compare the fecal and blood microbial profiles of obese and lean individuals for NAFLD patients based on rules implemented from the available dataset, which resulted in experimental approaches [1]. And applying Random Forest (RF) as an ML regression model [12] to predict the microbiome profile in fecal and blood analysis for both a given NAFLD (lean and obese). The next sections of this article are arranged into these main sections as below;

- Section 2 presents the Materials and methods applied in this research.
- Section 3 presents the results that were applied in this research.
- Section 4 presents the Discussion
- Section 5 presents the Conclusion and Future Work

2. Materials and methods

2.1 Machine Learning in Prediction

Prediction is a reflection of learning ability based on human judgment. With the advancement of society and technology, predictive science is now widely applied in various fields, including time series prediction, regression analysis, and pattern recognition. Classification and prediction are core branches of Machine Learning (ML), statistics, and data mining, as they involve prediction. The goal of a prediction model is to train a model with existing data to predict the output based on the input. With the development of ML technology, applying ML algorithms to prediction problems has become a research hotspot in the field of ML [13].Recently, there has been a growing interest in integrated learning from both industry and academia. ML as a state-of-the-art is widely used in various fields, including medical diagnosis, pattern recognition, scientific research, and data mining. The primary goal of integrated learning is to train multiple individual learners using a given dataset and then combine their respective prediction results to generate an outcome.

Gut-micro research has been utilizing ML to address various tasks [14], such as:

- Phenotyping (host phenotype).
- Microbial feature classification.

Intricate chemical and physical interactions among the constituents of the microbiome. Monitoring for changes in the microbiome composition [15].

Table 1 [14] presents a list of specific examples of ML tasks in Gut-Micro research. In this proposed work, the main objectives are focused on two tasks.

1. Analysis of the gut microbiome profile for NAFLD patients in cases of lean and obesity for a given dataset [1], and thus, deriving conclusions for a specific number of microbiome species found in blood or fecal analysis in lean and obese cases, as detailed in [1].
2. Applying the ML and RF models to predict the species of Gut-Micro in blood or fecal samples for NAFLD (lean and obese) cases.

Table (1) The ML methods used in microbiome research

Task	Predictive goal	Method	Reference
Phenotyping	Sponge bacterial density	Random forests	Moitinho-Silva et al. [16]
Phenotyping	Crop productivity	Random forests	Chang et al. [17]
Phenotyping	Food allergy	Recurrent neural network (LSTM)	Metwally et al. [18]
Phenotyping	Disease (inflammatory bowel disease)	Random forests, lasso, elastic nets.	Wirbel et al. [19]
Phenotyping	Disease (e.g., cirrhosis, type 2 diabetes, inflammatory bowel disease)	Convolutional neural networks	Sharma et al. [20], Reiman et al. [21, 22]
Microbial feature classification	Microbiome composition	Autoencoder	García-Jiménez et al. [23]
Microbial feature Classification	Metabolic profile	Autoencoder	Le et al. [24]
Interaction analysis	Microbe-metabolite interactions	Embedding	Morton et al. [25]
Interaction analysis	Microbe co-occurrence patterns	Embedding	Tataru and David [26]
Monitoring composition	Response to diet change	Autoencoder	Reiman and Dai [27]

2.2 Random Forest (RF)Technique.

The application of ML techniques to the interpretation of complicated and large-scale biological data has increased in modern biology. The approach is a widely used method in bioinformatics. It consists of an ensemble of decision trees (DTS) and naturally

integrates feature selection and interactions into the learning process [28,29]. Random Forest (RF) has multiple channels advantages, such as, efficient, interpretable, nonparametric, and having good prediction accuracy for a wide range of data types. Due to its unique advantages in handling high-dimensional feature spaces, complex data structures, and small sample sizes, RF has been increasingly utilized in computational biology research recently.

2.2.1 Bioinformatics with Random Forest Solutions

Among well-liked ML techniques, RF offers a special blend of model interpretability and prediction accuracy. The use of ensemble and random sampling techniques in RF allows for improved generalizations and precise prediction-making. This aspect of generalization originates from the bagging scheme, which enhances the generalization by reducing variance; comparable techniques, such as boosting, accomplish this via reducing bias [32].

The main merits of RF as next;

1.It can evaluate the significance of each feature through model training.

2.It can make reliable predictions for a wide range of applications

3. The trained model can measure the pairwise proximity between samples.

According to these facts about RF, it is recommended to apply RF in solving bioinformatics studies (Gut-Micro) in this study, because of the huge amount of data received in experimental procedures [33]

2.2.2 Random Forest Criteria

.

The fundamental units of an RF algorithm are DTs. A decision support method that resembles a tree is called a DT. It is easy to comprehend the operation of RF algorithms by first reviewing DTs.Three nodes make up a DT: a root node, leaf nodes, and decision nodes. A training dataset is divided into branches by a decision tree algorithm, and these branches subsequently subdivide into other branches. This process keeps going until a leaf node is reached. There is no way to segregate the leaf node further.

The decision tree's nodes stand for characteristics that are utilized to forecast the result. The decision nodes connect to the leaves. The three different kinds of decision tree nodes are depicted in the following Figure (1).

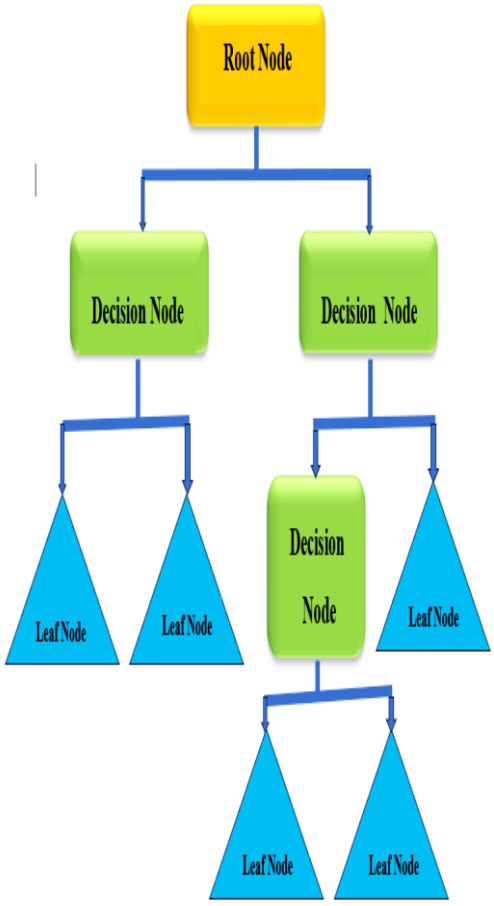

Fig. (1) Decision Tree components.

The foundation of a DT is entropy and information gain. Having a basic understanding of these ideas can help researchers better comprehend how DTs are constructed.

Uncertainty can be measured using entropy. The degree to which a set of independent variables reduces uncertainty in the target variable is measured as information gain.

Using independent variables (features) to learn more about a target variable (class), is known as the information gain idea. The information gain is estimated using the entropy of the target variable (Y) and the conditional entropy of Y (given X). In this instance, the entropy of Y is subtracted from the conditional entropy.

DTs are trained using the information gained. It aids in lowering these trees' level of uncertainty.

A substantial information gain indicates a significant reduction in uncertainty (information entropy). Splitting branches, a crucial step in the creation of decision trees, requires consideration of entropy and information gain [32].

2.3 Dataset

In this proposed work, the dataset from [1] is imported and prepared for analysis and predictions for the gut microbiome. In this study [1], Eligibility screening for this study was conducted between June and September 2014 at Kangbuk Samsung Hospital Total Healthcare Screening Centers in Seoul, South Korea.

Also, this study included three main datasets for 268 patients. Medical tests showed that 68 patients were suffering from NAFLD, and 192 patients did not have NAFLD. Besides, the gut microbiome profile of 268 patients was associated with this, as explained in the next section.

2.3.1 Preparing the Dataset

- From Table 2 [1]. This represents the demographic and clinical characteristics of subjects with NAFLD vs. control according to their body habits [1]. The BMI feature is extracted and imported for (lean and obese) in the case of NAFLD in 76 cases (27 lean and 49 obese).
- From the supporting information section [1], which contains the S1 table. Significant taxa in fecal microbial related to NAFLD groups are linked to the original link for the S1 Table as mentioned in [1] through this link [33].
- From the supporting information section [1], which contains the S2 table. Significant taxa in blood microbial related to NAFLD groups are linked to the original link for the S2 Table as mentioned in [1] through this link [33].
- From a, b, and c, we import all the required data about BMI, fecal microbiota profiles in taxa, and blood microbiota profiles in taxa. Then the new dataset, which contains the microbiome profile in fecal and blood samples gathered with BMI. The next section explains this dataset.

2.3.2 Dataset Applied

From the last section, the dataset is ready for processing and analysis procedures and predictions. We offer this dataset through this link [34].

The dataset contains many species of the microbiome as next;

(F__f_Desulfovibrionaceae&F_g__Fastidiosipila&F_g__Roseburia&F__f_Enterobacteriaceae&F_g__Erysipelothrix&F_g__Citrobacter&F_g__Acidaminococcus&F_g__Faecalibacterium&F_g__Turicibacter),(B_f__Deferribacterales_incertae_sedis&B_f__Succinivibrionaceae&B_f__Nocardioidaceae&B_f__Deinococcaceae&B_f__Leuconostocaceae&B_f__Beijerinckiaceae&B_f__Desulfobacteraceae&B_g__Nocardioides&B_g__Deinococcus&B_g__Leuconostoc&B_g__Caldithrix&B_g__Clostridium_sensu_stricto&B_g__Anaerobiospirillum)

For Fecal analysis, it begins with a capital letter (F), and for Blood analysis, it begins with a capital letter (B).

For the family level, it is represented by the letter (f) in the microbiome name; for the genius level, it is represented by the letter (g) in the microbiome name.

3. Experiment and Results

3.1 Approach 1: The rules for gut microbiome abundance for NAFLD patients.

The dataset is now ready for visualization, We use the available resources for processing data with utilities of a processor Intel Core i5, 16 gigabytes of RAM and a NVIDIA 310 graphics processing unit. The programming tool is Python 3.9 through Spider, an Anaconda platform.

As shown in Figure 2, it is clear that there is a diversity of microbiome profiles related to BMI and Blood & Fecal analysis taxa.

As shown in Figure 2, it is obvious that BMI is a linear factor that represents lean and obese cases.

From the figure, there is a negative correlation between BMI < 25 kg/m^2 and some blood microbiome, but for Fecal, the correlation is positive, such as;

- B_f__Deferribacterales_incertae_sedis
- B_f__Nocardioidaceae
- B_f__Beijerinckiaceae
- B_f__Desulfobacteraceae
- B_g__Nocardioides
- B_g__Callithrix

From the figure, there is a negative correlation between BMI > 25 kg/m^2 and some blood microbiomes, but for Fecal, the correlation is negative, such as;

- B_f__Deferribacterales_incertae_sedis
- B_f__Nocardioidaceae
- B_f__Beijerinckiaceae
- B_f__Desulfobacteraceae
- B_g__Nocardioides
- B_g__Callithrix

This indicates that previous microbe activities decrease in the case of overweight (30>BMI >25 kg/m^2) case, but increase in the case of lean (BMI <25 kg/m^2) case.

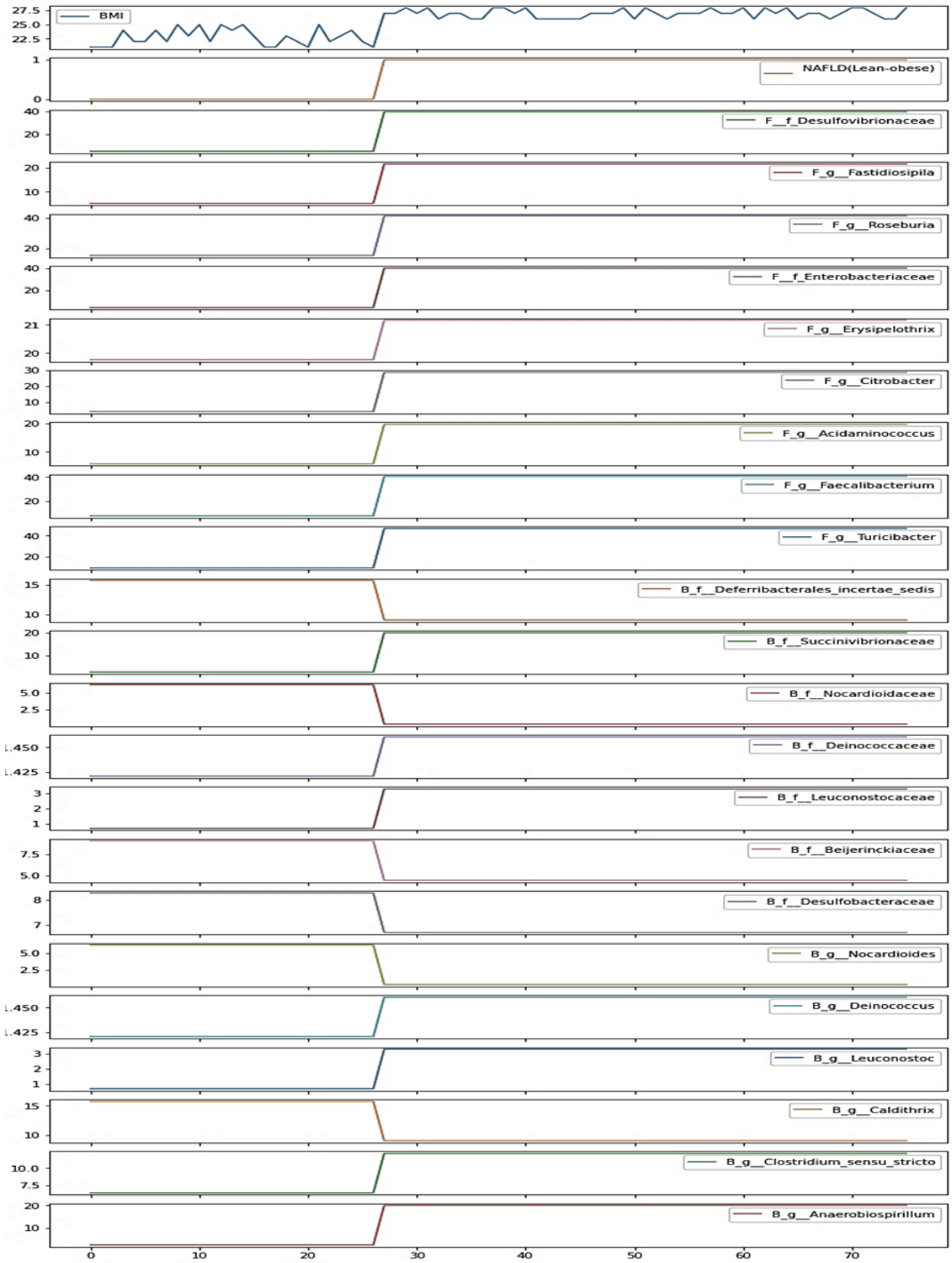

Table 2 is a dedicated table that shows the ratio between (lean: obese) and gut microbe Abundance. As shown in Table 2, in general, the gut microbiome is more active and more populated in the case of obese NAFLD patients, specifically in fecal taxa. E.g., F__f_Enterobacteriaceae increases by 8.3 times if the patient is obese than a lean one.That led to 1st result, which defines that. The environment of the gastrointestinal tract is very conducive to the growth of the gut microbiome population.

On the other side gut microbiome is less active and has a larger population in the case of obese NAFLD

patients, exactly in the Blood taxa.For example, B_g__Nocardioides decreases by 0.05 times if the patient is obese compared to a lean one.

That led to 2nd result, which defines that. The environment of blood vessels for the obese isn't very good for the gut microbiome population is the environment.

Table 2. Microbiome species in the taxa ratio for (lean, obese) NAFLD Cases.

Microbiome type	Taxa	Appropriated Ratio between (lean: obese)
F__f_Desulfovibrionaceae	Fecal	≈ (1: 7.8)
F_g__Fastidiosipila		≈ (1: 4.25)
F_g__Roseburia		≈ (1: 2.7)
F__f_Enterobacteriaceae		≈ **(1: 8.3)**
F_g__Erysipelothrix		≈ (1: 1.07)
F_g__Citrobacter		≈ (1: 7.6)
F_g__Acidaminococcus		≈ (1: 3.4)
F_g__Faecalibacterium		≈ (1: 5.3)
F_g__Turicibacter		≈ (1: 5)
B_f__Deferribacterales_incertae_sedis	Blood	≈ (1: 0.6)
B_f__Succinivibrionaceae		≈ (1: 7.5)
B_f__Nocardioidaceae		≈ (1: 0.05)
B_f__Deinococcaceae		≈ (1: 1)
B_f__Leuconostocaceae		≈ (1: 4.75)
B_f__Beijerinckiaceae		≈ (1: 0.5)
B_f__Desulfobacteraceae		≈ (1: 0.8)
B_g__Nocardioides		≈ (1: 0.05)
B_g__Deinococcus		≈ (1: 1)
B_g__Leuconostoc		≈ (1: 4.75)
B_g__Caldithrix		≈ (1: 0.6)
B_g__Clostridium_sensu_stricto		≈ (1: 1.9)
B_g__Anaerobiospirillum		≈ (1: 7.5)

3.2 Evaluation Metrics

To evaluate the performance of the proposed approach, several performance metrics are employed [35].

Recall (Rec): This metric is also frequently called sensitivity.

$$\text{Recall} = \frac{TP}{FN+TP} \quad (1)$$

Precision (Perc): This metric is also frequently called the positive predictive value.

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

Specificity (Spec): This metric is frequently called the true negative rate.

$$\text{Specificity} = \frac{TP}{TN+TP} \quad (3)$$

Accuracy (Acc): The percentage of correctly identified class labels:

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

F1 score (F1): A measure of a test's accuracy by calculating the harmonic mean of the precision and recall:

$$\text{F1 score} \frac{2TP}{FP+FN+2TP} \quad (5)$$

Where True positive (TP) is the number of correctly identified positive samples.

True negative (TN) is the number of correctly identified negative samples.

False Positive (FP) is the number of incorrectly identified positive samples. False Negative (FN) is the number of incorrectly identified negative samples.

3.3 Approach 2: Predicting Gut Microbiome Abundance in NAFLD Patients.

- **Scenario 1: In the case of an NAFLD patient, his BMI is known.**

The first dataset is split into training and testing by 0.75–0.25 to prepare the dataset for prediction. We use cross-validation in training processing; the number of estimators in DT 100 is a random state of 42. The x, the input value for the classifier RF, represents the BMI for NAFLD patients [33].

Still, the y, which represents the output value, represents the gut microbiome in blood or fecal, as mentioned in Table 2.

In the next Figure 3, the confusion matrix between predicted and actual values of the input BMI patient and the F__f_Desulfovibrionaceae microbiome is shown, as shown by the total number of samples (19), because training is 25% of the 76 samples. For

F__f_Desulfovibrionaceae, lean patients' taxa (0),but (1) for obese patients' taxa.

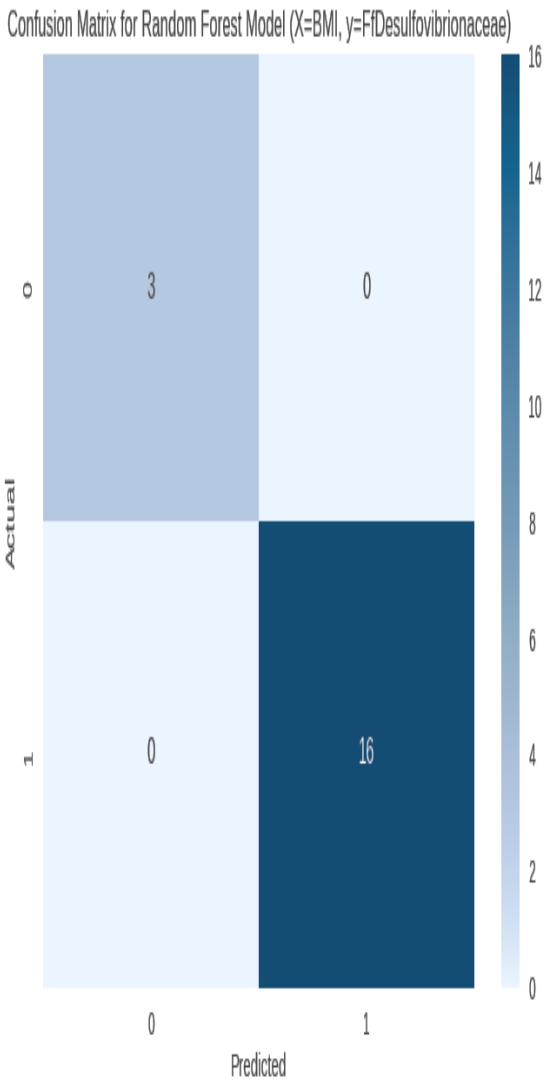

Fig. (3). Confusion matrix between (BMI, F__f_Desulfovibrionaceae) prediction RF model.

- **Scenario 2: In the case of an NAFLD patient, his BMI is unknown.**

The same trial was done for all gut microbiomes, and the same results were obtained, with an accuracy of 99%. In this result, the x, the input value for the classifier RF, represents the NAFLD data [33] only, and the y, which represents the output value, represents the gut microbiome in blood or fecal, as mentioned in Table 2. In the next Figure 4, the confusion matrix between predicted and actual values of the input NAFLD (lean obese) patient and F__f_Desulfovibrionaceae microbiome is shown, as shown by the total number of samples (19), because training is 25% of the 76 samples. For F__f_Desulfovibrionaceae lean patients, the taxa are (0), but for obese patients, the taxa are (1).

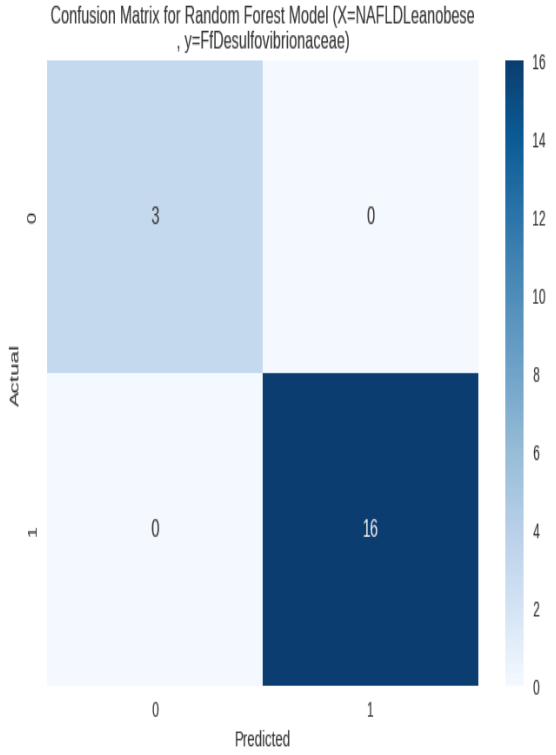

Fig. (4). Confusion matrix between (BMI, F__f_Desulfovibrionaceae) prediction RF model.

4. Discussion

In this approach, the desired system was proposed to predict the abundance of gut microbiome in NAFLD patients according to BMI for each patient. In case of knowing the weight of the patient.

To achieve this purpose, I benefited from a preliminary related study that contained data on patients with fatty liver disease and medical tests for patients of different weights [1].

It is clear from the last section, approach 1, that it can be estimated that the abundance of gut microbiome, according to Table 2, provides a basis for lean and obese patients.

Also in the last section, approach 2, the random forest algorithm has obtained excellent predictions for gut microbiome abundance in the case of known or unknown BMI of patients.

The methods that are implemented are very useful in the study of the gut microbiome in case of poor funding, and also achieve high accuracy equal to 99% by applying RF in gut microbiome prediction, compared to other related work [24,26].

5. Conclusion and Future Work

Through this study, a proposal was presented to find the relationship between the abundance of the gut microbiome and its effect on a person suffering from nonalcoholic fatty liver disease, according to the effect

of BMI variation. It became clear from the study that the body mass index factor of a person affects the proportion of the gut microbiome. And because of the high cost of testing gut microbe profiles in fecal and blood for one patient, this study helps in using artificial intelligence techniques, represented by one of the machine learning algorithms, which is a decision tree, in making predictions in this way. Gut microbiome in an infected person based on the body mass index factor, which helps to minimize the cost of testing gut microbiome profiles.

In future studies, the relationship between the gut microbiome and people suffering from pancreatic and heart diseases will be expanded, and a more accurate perception of the microbiome's shape will be created based on these studies.

Acknowledgements

We announce that this research proposal is among the studies that will be conducted at the Excellence Research Center for Liver, Heart, and Pancreatic Diseases at Benha University, code (CEPHEI 05145), focusing on the effect of fatty liver on the microbiome in the human body.

Conflicts of interest

"The authors declare that they have no conflicts of interest.".

References

[1] Yun, Yeojun, et al. "Fecal and blood microbiota profiles and presence of nonalcoholic fatty liver disease in obese versus lean subjects." *PloS one* 14.3 , 2019.

[2] Vernon, G., A. Baranova, and Z. M. Younossi. "Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults." Alimentary pharmacology & therapeutics 34.3, 2011.

[3] European Association for the Study of The Liver, and European Association for the Study of Diabetes (EASD. "EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease." Obesity Facts 9.2,65-90,2016.

[4] Younossi, Zobair M., et al. "Nonalcoholic fatty liver disease in lean individuals in the United States." Medicine 91.6, 319-327, 2012.

[5] Riazi K, Azhari H, Charette JH, Underwood FE, King JA, Afshar EE, et al. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol,7:851–861,2022.

[6] Kleiner DE, Brunt EM, Van NM, Behling C, Contos MJ, Cummings OW, et al. "Design and validation of a histological scoring system for nonalcoholic fatty liver disease". Hepatology. ,41(6):1313–1321,2005.

[7] Savage, Dwayne C. "Microbial ecology of the gastrointestinal tract." Annual review of microbiology 31.1 (1977): 107-133.

[8] Vrieze, Anne, et al. "Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome." Gastroenterology ,913-916.2012.

[9] Schnabl, Bernd, and David A. Brenner. "Interactions between the intestinal microbiome and liver diseases." Gastroenterology 146.6, 1513-1524,2014.

[10] Boursier, Jérôme, et al. "The severity of nonalcoholic fatty liver disease is associated with gut dysbiosis and shift in the metabolic function of the gut microbiota." Hepatology 63.3 ,764-775,2016.

[11] Ley, Ruth E., et al. "Human gut microbes associated with obesity." Nature 444.7122,2006.

[12] Huang, Jui-Chan, et al. "Application and comparison of several machine learning algorithms and their integration models in regression problems." Neural Computing and Applications 32.10,5461-5469,2020.

[13] Huang, Jui-Chan, et al. "Application and comparison of several machine learning algorithms and their integration models in regression problems." Neural Computing and Applications 32.10 ,5461-5469,2020.

[14] Hernández Medina, Ricardo, et al. "Machine learning and deep learning applications in microbiome research." ISME communications 2.1,2022.

[15] Marcos-Zambrano LJ, Karaduzovic-Hadziabdic K, Loncar Turukalo T, Przymus P, Trajkovik V, Aasmets O, et al. " Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment. Front Microbiol.,2021.

[16] Moitinho-Silva L, Steinert G, Nielsen S, Hardoim CCP, Wu YC, McCormack GP, et al. Predicting the HMA-LMA status in marine sponges by machine learning. Front Microbiol. 8:752,2017.

[17] Chang H-X, Haudenshield JS, Bowen CR, Hartman GL. Metagenome-Wide Association Study and Machine Learning Prediction of Bulk Soil Microbiome and Crop Productivity. Front Microbiol.8:519.2017.

[18] Metwally AA, Yu PS, Reiman D, Dai Y, Finn PW, Perkins DL. Utilizing "longitudinal microbiome taxonomic profiles to predict food allergy via Long Short-Term Memory Networks. PLoS Comput Biol,15:e1006693,2019.

[19] Wirbel J, Zych K, Essex M, Karcher N, Kartal E, Salazar G, et al. "Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. Genome Biol.22:93,2021.

[20] Sharma D, Paterson AD, Xu W. TaxoNN: "Ensemble of neural networks on stratified microbiome data for disease prediction. Bioinformatics.36:4544–50,2020.

[21] Reiman D, Metwally AA, Dai Y. "Using convolutional neural networks to explore the microbiome. Conf Proc IEEE Eng Med Biol Soc. 4269–72,2017.

[22] Reiman D, Metwally AA, Sun J, Dai Y. "PopPhy-CNN: A Phylogenetic Tree Embedded Architecture for Convolutional Neural Networks to Predict Host Phenotype from Metagenomic Data. IEEE J Biomed Health Inform.24:2993–3001,2020.

[23] García-Jiménez B, Muñoz J, Cabello S, Medina J, Wilkinson MD. Predicting microbiomes through a deep latent space. Bioinformatics. 37:1444–51,2021.

[24] Le V, Quinn TP, Tran T, Venkatesh S. " Deep in the Bowel: Highly Interpretable Neural Encoder-Decoder Networks Predict Gut Metabolites from Gut Microbiome. BMC Genomics.21:256,2020.

[25] Morton JT, Aksenov AA, Nothias LF, Foulds JR, Quinn RA, Badri MH, et al. Learning "representations of microbe-metabolite interactions. Nat Methods; 16:1306 14,2019.

[26] Tataru CA, David MM. "Decoding the language of microbiomes using word-embedding techniques, and applications in inflammatory bowel disease. PLoS Comput Biol.16: e1007859,2020.

[27] Reiman D, Dai Y. "Using Autoencoders for Predicting Latent Microbiome Community Shifts Responding to Dietary Changes. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp 1884–91,2019.

[28]Qi, Yanjun. "Random forest for bioinformatics." Ensemble machine learning: Methods and yer, H. Husslein, N. Concin, A. Ridder, M. Musielak, C. Pfeifer, et al. Fetal weight estimation at ter[1] Yun, Yeojun, et al. "Fecal and blood microbiota profiles and presence of nonalcoholic fatty liver disease in obese versus lean subjects." PloS one 14.3,2019.