Data Mining and Knowledge Discovery Tools for Human Microbiome Big Data

Oana Geman
Department of Health and Human Development
"Stefan cel Mare" University of Suceava
Suceava, Romania
oana.geman@usm.ro

Iuliana Chiuchisan
Computers, Electronics and Automation Department
"Stefan cel Mare" University of Suceava
Suceava, Romania
iuliana.chiuchisan@usm.ro

Mihai Covasa
Department of Health and Human Development & Department of Basic Medical Sciences,
"Stefan cel Mare" University of Suceava & Western University of Health Sciences,
Suceava, Romania & Pomona, CA, USA
mcovasa@gmail.com

***Abstract*—The human microbiome is a fundamental component of human physiology, with an estimated one-third of circulating metabolites being a product of the gut microbiota. Changes in the microbiome can trigger changes in human cellular activities, resulting in disease or contribute to its progression. Microbiota is considered to be a virtual "code" or a system emerging, with the properties that must be integrated into the biology and physiology of the human. Unlike the other components, the functions of this "code" are not yet fully understood but can be quite easily disturbed by diet, diseases, and the various treatments. Recently, it becomes more and more clear that these functions can be beneficial both and with the negative impact on the health status of the human host. The recognition of the role of soil microflora in the intestinal diseases and neurological or in metabolic diseases of systemic triggered an avalanche of studies which also were aimed at the elucidation of the human microbiome (Human Microbiome Project), but also a complicated interaction between the bi-directional relationship between the human and the intestinal bacteria. An important step represents the discovery of potential methods of use of human microbiota in prevention and treatment of diseases such as autism, asthma, Parkinson disease, obesity, and diabetes. There is an impressive collection of data (Human Microbiome Big Date) which can be analyzed and classified using the algorithms of Data Mining or Knowledge Discovery Date Tools.**

Index Terms - Data Mining; Knowledge Discovery Date; Classification and Clustering; Big Data; Human Microbiome

I. INTRODUCTION

The microbiome is currently one of the recent issues in biomedical research. For every one human gene we have, there are 100 associated genes within our microbiome. More than 100 trillion microorganisms live in our gut, mouth, skin and other mucosal surfaces of our bodies [1].

These microbes have numerous beneficial functions relevant to supporting life such as digesting food, preventing disease-causing pathogens from invading the body, and synthesizing essential nutrients and vitamins. With the advancement of genomic technologies, the capacity of this "second genome" to influence health and disease can now be harnessed. Only in the last years began to appear new studies, which indicates the important role of bacteria from so-called saprophytic in elementary processes, such as the digestion of nutrients and vitamins, but also in the most complicated processes, such as the synthesis of amino acids. The microbiome is heavily involved in metabolic processes, but has an influence also on the immunity and even on the human behavior. The bacteria forming the microbiome exists in human digestive tract since the first 24 hours from birth and stabilizes around the age of two years, becoming a permanent presence in our lives and forming a real component. Specialists estimate that the thousands of billions of germs that live in our body weighing in total almost 2 kilograms [1] ÷ [6]. Most of them live in the intestines and their destruction may have unpleasant consequences, such as the obesity or the development of diseases of the colon. Knowing the microbes that live in our body, scientists hope to be able to provide a better life through custom diets, faster diagnosis of the problems and customs treatment of diseases. Many of the research conducted in the past few years have shown that the maladjustment of the microbiome may have bad consequences on human health, among them the allergies, rash, asthma and obesity.

Human Microbiome Project (HMP) refers to DNA/ARN mapping of all microorganisms which cohabit in the healthy human body and aim to „characterize microbial communities found at multiple human body sites and to look for correlations between changes in the microbiome and human health" [1].

The latest approaches of biotechnology in medicine and molecular biology have allowed the investigation of the gut microbiota role referring to the development of obesity and metabolic diseases as well as to the susceptibility to infections or the occurrence of several chronic diseases (neurological, cardiovascular), immunological and autoimmune diseases such as Rheumatoid Arthritis (RA), or even behavioral changes (depression, anxiety), autism and the ADHD in children [1] ÷ [15].

There are numerous international projects included in the Human Microbiome Project, and also the European projects, including the MetaHIT [1], that have led to the discovery of the

fact that people differ by bacteria also, not only through the components of the type of the blood group. Scientists have discovered that there are 3 types of microbial populations, named enterotypes.

The researchers have discovered that enterotypes have no connection with age, gender, nationality or diet. These enterotypes have received the name of the dominant microbial group: bacteroides Prevotella, and Ruminococcus. Bacteroides microbes are more efficient in the decomposition of carbohydrates, which could explain why the persons who present this enterotype are more susceptible to problems of weight. Bacteria type Prevotella tends to decompose mucus of the intestine and Ruminococcus facilitates the absorption of the sugar which may promote the increase in weight.

A study conducted last year has shown that Bacteroides enterotype is associated with a diet rich in fat and proteins when Prevotella enterotype is associated with a diet rich in carbohydrates. The researchers have discovered that in 90 percent of cases it can bed deduce that a person is obese or not only by analyzing the microbiome. A study carried out in Denmark has shown that the analysis of the microbiome allows physicians to estimate with accuracy if a patient suffers from diabetes, the gut microbes Bacteroides group multiplying being a mark better than the index of the body mass of the person concerned.

II. Enterotypes and Personalized Medicine

Numerous studies show that the discovery of the enterotypes will lead to the transformation of the medicine as it happened in the case of blood- grouping reagents. Custom diets on the basis of the enterotype and the prescription of medicine suitable for the patient based on the enterotype are the first changes prefigured by specialists [1] ÷ [15].

Also, the studies in this field show that the enterotypes will play an important role in the discovery of alternative treatments to antibiotics, which have become ineffective in recent years, as the bacteria have become more resistant. Thus, instead of trying to abolish all the bacteria from the gut, including those harmful, doctors will be able to stimulate the bacteria usefully, in order to restore the balance of the existing bacterial before the development of the affection. With time, enterotypes will allow the development of personalized medicine in which each of the patients will benefit from a treatment designed according to the needs of its own, identified on the basis of microbiome (Figure 1).

	Human microbiota (10 times more microbial than human cells: 10^{14} vs 10^{13})	
Human microbial habitats	Most represented Phyla and their relative abundance (%)	Number of species
Oral cavity	*Firmicutes* (36.7), *Bacteroidetes* (17.3), *Proteobacteria* (17.1), *Actinobacteria* (11.9), *Fusobacteria* (5.2)	>500
Skin	*Actinobacteria* (52), *Firmicutes* (24.4), *Proteobacteria* (16.5), *Bacteroidetes* (6.3)	~300
Airways	*Actinobacteria* (55), *Firmicutes* (15), *Proteobacteria* (8), *Bacteroidetes* (3)	>500
Gut	*Firmicutes* (38.8), *Bacteroidetes* (27.8), *Actinobacteria* (8.2), *Proteobacteria* (2.1)	>1000
Urogenital tract[a]	*Firmicutes* (83), *Bacteroidetes* (3), *Actinobacteria* (3)	~150

[a] Mainly female.

Fig. 1. Human microbiota composition across the five most extensively studied body sites [1]

Moreover, in the papers [3], [4], professor Mihai Covasa and other researchers focused on the role of gut microbiota in the modulation of physiological, molecular and neuronal signals involved in regulating the energy balance in both, healthy and obesity conditions, using in vivo models and dietary manipulations.

Bacteriotherapy is already a treatment which has begun to be implemented with encouraging results in the case of patients for which the antibiotics and other treatments do not give the results. Ehrlich Dusko, the project coordinator within the framework of the MetaHIT, explains why microbiome will play an important role in the custom medicine: "Our nuclear genome differs by a maximum of 0.1 percent from man to man, but the difference of metagenomes can be at 50 percent" [1]. Therefore, by analysis of genes, clinicians can be identified the unique aspects to each patient which will enable the designing of custom treatments [1]. Also, researchers expect that in the future the children will not be vaccinated only against viruses, but will be subject to detailed analysis of microbiome, to be identified bacteria key missing with a view to reintroducing them.

III. Human Microbiome Project and Obesity

Humanity is faced with a devastating epidemic of metabolic syndrome associated with a dramatic increase in the whole world of obesity and diabetes. According to estimates of recent International Federation of diabetes, there are 387 million people with diabetes at world level, and their number is expected to grow at 592 million by 2035 [17].

In the EU there are about 60 million people with diabetes [18], out of which over 1.5 million of diabetics from Romania in 2014 [19]. Almost 85 percent to 95 percent of persons with diabetes have diabetes of type 2. To the development of diabetes type 2 contribute a set of risk factors such as genetic predisposition, age, excessive weight or obesity, and an unhealthy lifestyle. In addition to these factors, recent research has suggested that the intestinal microbiota plays an important role in the pathogenesis of diabetes of type 2 [20]. A human adult is settled of about 100 trillion bacteria, which represents ten times more than the number of total cells from the human body [21].

The human intestinal tract contains a unique group of micro-organisms which is represented by the microbiota, consists of numerous bacteria, Archaea (Archaea bacteria) and viruses. All these microorganisms generate a biomass of over 1.5 kg and their combined genomes (microbiome) exceed the human genome of more than 100 [22] ÷ [26]. In general, these genes have the main functions the involved processes, digestion of the complex carbohydrates and the development of the immunity [27]. Recent studies suggest that the microbiota may have key functions in the regulation of metabolic pathways in both the health status and the disease.

The process of dysbiosis of the intestinal microbiota as a result of the diet with a high content of fat content or high heat

has a key role in the development of obesity, resistance to insulin, hyperglycemia, and hyperlipidemia, all associated with the development of diabetes [28] ÷ [29]. Indeed, discount numerical beneficial bacteria (e.g. Bifidobacterium) and increase of bacteria pro-inflammatory or pathogenic organisms are consistently associated with the development of or the emergence of the obesity, diabetes, adipose tissue and systemic self and metabolic causes both humans [30] ÷ [33] and animals [24].

The Human Microbiome Project (HMP) Data Analysis and Coordinating Center (DACC) Data Portal provides access to all publicly available HMP datasets [1]. The portal is a comprehensive collection of databases and libraries, serving as a useful gateway for access to microbiome data [1].

IV. Classification of the Data Mining Algorithms and Knowledge Discovery Tools

With the rapid growth of the number and size of the databases as well as of the applications of databases in the field of Human Microbiome, it is necessary to examine the automatic extraction of knowledge from large databases. The discovery of knowledge from the databases (Knowledge Discovery in databases - KDD) or extraction of data (Data Mining - DM) represents the effort to understand, analysis and possibly to use a huge amount of available data specific to Human Microbiome Project (Figure 2).

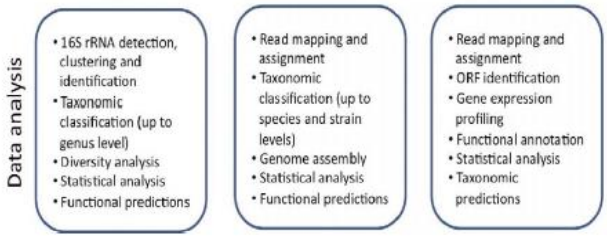

Fig. 2. Specific Data Mining Tools pipelines used for Data Analysis [1]

The main function of the Data Mining is to extract the models of knowledge from data. To do this, DM uses a variety of algorithms such as statistics, recognition of forms, classification, fuzzy logic, automatic learning (machine learning), genetic algorithms, neural networks, data view, etc. The variety of algorithms may be grouped in the main components of the DM. The number of components differs from an author to another, being:

- The choice of significant data from the database - which, like any computer model, it means by a function in space or multidimensional (a combination of functions), depending on the parameters. It can be represented either as a linear function of parameters, either as a function of the probability (e.g. normal), either as a function fuzzy, etc. obtaining the model is done by different algorithms, such as those of the classification and the pool (clusterization);

- The criteria of preference, preprocessing of data which are significant in the current search - which may be of a different nature, some of them based on the sort, some in the interpolation or the best approximation;

- The algorithms of selection, which are the algorithms for the determination of the desired results on the basis of data selected which lead to the selection of three important elements which appear in the DM: the model which shall be selected on the basis of models and the data which shall be selected from the database and constitutes the parameters and the criterion or the criteria of preferences that are selected on the basis of criteria;

- The establishment of deviation - consisting in general of algorithms for the determination of the deviation and the stability, a specific category of such algorithms are those statistics, which lay down the model deviations from the ideal.

Among the classical methods used for data mining include statistical methods, the methods which use information from the neighbors and those of clustering.

The algorithm closer neighbor's *k* (*k*-nearest neighbors or *k*-NN) is a predictive technique suitable for the classification of the patterns. Technical *k-NN* assumes that the entire set of drive shaft includes both the data and the desired classifications for each item. As a result of the application of the technique, the data become drive model. *K* represents the number of similar cases or the number of items in the group. Upon the occurrence of an event or a new example in the model, the algorithm to verify all the data to find a subset which is closest to the new one and offers this subset as a result, by placing it in the class which contains the most items manufacturer of this set of the closer *k* items manufacturer. The algorithm *k-NN* has two main parameters: the number of the nearest (similar) cases *k* and a meter to measure the similarity to each use of the algorithm *k-NN* it is necessary to specify a positive integer for value of *k*. This value represents the number of existing cases that are being analyzed to the specification of a new case. The algorithm *k-NN* shall decide what class to place a new case by examination of a number *k* cases (neighbors) with a high degree of resemblance.

The method of classification of the data is a statistical method used to group multi-dimensional data specifications (i.e. "points" representing the cases or observations) in groups (clusters) defined algorithmically. This method is useful for summarization of large amounts of information, each group representing several points with the same characteristics. Separate clusters do not overlap. In fact, the analysis of the classification consists of a collection of algorithms exploiting the fundamental heuristics, mainly based on the experience of our "visual inspection" in the group of points in the "clouds of points". In general, to be able to use an algorithm for the classification, there is a need to specify:

- The type of the distance between the points of a multidimensional space;

- The strategy for the choice of the representative point (i.e. the 'Center') for any group of points. Most people have a tendency to choose the arithmetic average (i.e. "center of gravity");

- The type of the distance between the two groups of points. The most frequently used distances also take into account the distance between points are chosen previously.

Among the Data Mining modern technologies it can be mentioned the rules of the association. These are defined as follows: $I = \{i_1, i_2,..., i_m\}$ a number of symbols, called elements. It is considered *D* a set of transactions in which each transaction *T* constitutes a subset of *I*, where *T* is a lot included or equal to the *I*. Shall only be taken into account the representation (coded in binary) of the elements in the transaction and shall not be considered other qualitative or quantitative characteristics. Each transaction has an associated identifier. The key steps in the framework of the extracting the rules of association are the support and confidence. The support relates to the proportion in which a relationship appears in the database. The confidence or trust of the rules of association refers to the probability to find a prerequisite having a consequence. The determination of the rules shall be carried out in two steps:

1) Determination of sets of frequent items, those which have sufficient support;

2) The determination of the rules of the association between these sets, the determination of rules countries.

V. Applications of Data Mining Methods in Human Microbiome Database

On HMP Database, there are available: 2471 Complete Genomes; 5543 Draft Bacteria and Archaea Genomes; 2399 Complete Virus Genomes; 26 Complete Fungi Genomes; 309 HMP Eukaryote Reference Genomes; Total 10,741 genomes, approximately 30 GB of sequences. With this increased technological means, the HMP aim to provide a catalog of microbes living in the human body and to establish their functions.

Some studies have demonstrated a clear association between the metabolic diseases and changes in the composition of the intestinal microbiota, evidenced by a low weighted of Firmicutes and a high proportion of Bacteroidetes and Proteobacteria in patients with diabetes of type 2, comparing the profile of the intestinal microbiota in persons with non-diabetes [3], [4].

The selective changing in intestinal microbiota, through the supplementation with prebiotics or probiotics has demonstrated that this method drops the metabolic endotoxemia induced by a diet with high levels of fat, decreases the permeability of the intestinal flora, improves the control of blood sugar and prevents diabetes of type 2 with the intention of maximizing insulin resistance [17]. However, these findings and the results are promising, very few studies have been carried out on the man, in order to prevent and control the diabetes of type 2. Also, clinical studies of randomization on large samples are missing and the results are contradictory.

The algorithms of Data Mining and Knowledge Discovery Tools are useful in finding these correlations between the intestinal microbiota and the presence or absence of diabetes. It is possible to do tests "in silico" in order to predict the metabolic diseases and in particular of diabetes on the population in order to be healthy.

For exemplification, we used the database on the Human Microbiome Project with a number of 124 obese persons who have diabetes of type 2 [1] and we used the WEKA environment, a collection of algorithms for Data Mining learning. The WEKA algorithms were used directly on the data sets using the tools for the pre-processing the data, classification, regression, rules of association and viewing of data. The WEKA environment is an open source under the General Public License GNU [34].

In Figure 3, with *A* is noted the class of patients with diabetes of type 2 on which are present Firmicutes micro-organisms, *B* is the class of patients with diabetes of type 2 on which are present Bacteroidetes micro-organisms and *C* is the class of patients with diabetes of type 2 on which are present Proteobacteria micro-organisms.

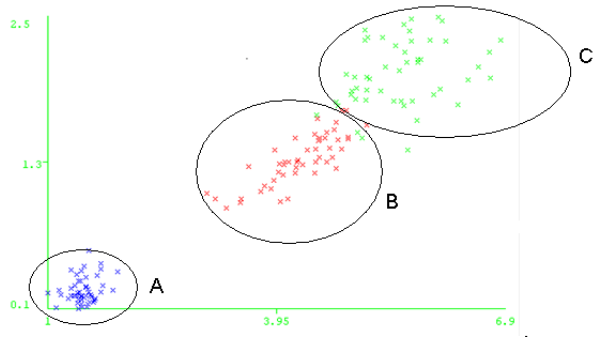

Fig. 3. The WEKA classification using SVM (Support Vector Machine) for A, B, C classes

The best results were obtained with an SVM classifier (Support Vector Machine) and with the artificial neural networks (ANN - Artificial Neural Networks) (Table I and Table II).

TABLE I. Results for the data sets from the patients with diabetes of type 2 for firmicutes class (A)

WEKA ML Classifier	**Results**		
	Mean abs. error/ Mean sqr. error	***Classification***	***Time***
ZeroR	0.476/0.482	65.39%	0.65
DecisionStump	0.231/0.334	93.23%	0,52
DecisionTable	0.089/0,234	95.16%	4.76
IB1	0.067/0.234	94.37%	0.12
J48.J48	0.545/0.124	93.11%	3.42
kStar.kStar	0.056/0.243	94.11%	0.66
Logistic Regress	0.054/0.122	93.32%	1.07
Naïve Bayes	0.035/0.143	94.45%	1.53
SMOptimization	0.013/0.112	96.99%	12.45
Voted Perceptron	0.198/0.344	89.98%	1.08
Neural Network	0.023/0.123	97.12%	12
Voting Feature	0.435/0.34	93.33%	0.45
Ada BoostM1	0.111/0.122	94.76%	7.23

TABLE II. RESULTS FOR THE DATA SETS FROM THE PATIENTS WITH DIABETES OF TYPE 2 FOR BACTEROIDETES CLASS (B)

WEKA ML Classifier	Results		
	Mean abs. error/ Mean sqr. error	*Classification*	*Time*
ZeroR	0.482/0.594	75.19%	0.52
DecisionStump	0.131/0.324	68.25%	0,46
DecisionTable	0.081/0,256	77.17%	4.68
IB1	0.076/0.212	84.31%	0.09
J48.J48	0.505/0.101	84.15%	3.47
kStar.kStar	0.023/0.223	84.11%	0.57
Logistic Regress	0.024/0.114	92.35%	1.11
Naïve Bayes	0.027/0.157	95.27%	1.49
SMOptimization	0.003/0.111	92.93%	10.42
Voted Perceptron	0.228/0.222	91.98%	2.05
Neural Network	0.033/0.101	98.21%	11.09
Voting Feature	0.355/0.224	94.23%	0.65
Ada BoostM1	0.021/0.247	92.02%	8.23

We selected the algorithms of classification depending on the shafts of the decision, and also the algorithms rely on Bayesian models. In order to assess the discriminative power of the used WEKA classification algorithms, we computed the overall recognition rate (accuracy), their sensitivity and specificity, as well as the standard deviation of the classification results. We used:

- Sensitivity = (number of correctly classified "A") / (number of "A" + number of false "B");
- Specificity = (number of correctly classified "B") / (number of "B" + number of false "A");
- Total classification accuracy = (number of correctly classified welfare) / (number of total welfare).

The accuracy of the classifier is the percentage of correctly classified instances in the test set, measuring how well the classifier recognizes instances of the various classes […].

For class A, we obtained very good results using SMO and ANN, and for class B and C we obtained similar results in favor of the use of Naïve Bayes and ANN.

VI. CONCLUSIONS

Research on microbiome is still in an incipient phase and a better understanding may turn into an ally in the fight against diseases. Therefore, the bacteria play an essential role in ensuring our health. The specialists involved in the study of the human microbiome have proven potential benefits of its understanding. The effort for a better understanding of the microbiome represents the highest imported scientific project of all times.

The simulations made using the tools WEKA Machine Learning can help us to select suitable Data Mining algorithms according to each database. We can find connections between different pathologies and Human Microbiome by using Data Mining applications.

Human Microbiome Project remains a global challenge priority since it involves large communities of researchers and requires the latest approaches in biotechnology and bioinformatics. Understanding how these microorganisms interact with their human hosts could explain different aspects of many complex disorders from obesity, diabetes and metabolic diseases to autoimmune diseases and neurological disorders.

Within the project "Analysis of novel risk factors influencing control of food intake and regulation of body weight" coordinated by professor Mihai Covasa we found this concern related to microbiota and obesity, diabetes and other metabolic diseases [35]. Concerning the challenges of the most recent research carried out on the microbiome-gut-brain axis, recent studies provide a significant contribution because it could lead to the discovery of alternative ways to treat the patients, especially those with brain dysfunctions that do not always respond to usually prescribed medication [36].

Probably, in the near future, the prevention of such diseases, healing and maintaining health should involve no gene manipulation but rather a greater manipulation of microbiome through major changes in “modern” nutrition (the excess of sugars and carbohydrates). Awarding much more attention to prescribing certain drugs in excess (antibiotics, antidepressants) will be also one of the basic directions.

The next step will be the extraction of rules and finding the rules of association or fuzzy rules. We will use the special algorithms designed for the extraction of rules, such as a priori, FPGrowth or Deep Learning algorithms [37]÷[40]. The purpose of seeking improved alternatives to these rules of association is to be able to develop a system based on the knowledge which will make the connection between the Human Microbiome and diseases such as Parkinson's disease, diabetes or cancer.

ACKNOWLEDGMENT

This work was supported by the Romanian National Program PN-II-ID-PCE-2012-4-0608 no. 48/02.09.2013, "Analysis of novel risk factors influencing control of food intake and regulation of body weight" [35].

REFERENCES

[1] Human Microbiome Project, commonfund.nih.gov/hmp/overview.

[2] B.S. Samuel, A. Shaito, T. Motoike, F.E. Rey, F. Backhed, J. K. Manchester, R.E. Hammer, S.C. Williams, J. Crowley, M. Yanagisawa, J.I. Gordon, “Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41,” Proceedings of the National Academy of Sciences, Vol. 105, pp.16767-16772, 2008.

[3] T.D. Swartz, F.A. Duca, T. de Wouters, Y. Sakar, M. Covasa, “Upregulation of intestinal type 1 taste receptor 3 and sodium glucose luminal transporter-1 expression and increased sucrose intake in mice lacking gut microbiota,” Br. J. Nutr., Vol. 107, pp. 621-630, 2012.

[4] F.A. Duca, M. Covasa, “Current and emerging concepts on the role of peripheral signals in the control of food intake and development of obesity,” Br. J. Nutr., pp.1-16, 2012.

[5] J.M. Fettweis, J.P. Brooks, M.G. Serrano, N.U. Sheth, J. Vladimir, Microbiology (Reading, England), 2014.

[6] A. Eren, A. Murat, G.G. Borisy, S.M. Huse, J. Mark Welch, “Oligotyping analysis of the human oral microbiome,” Proceedings of the National Academy of Sciences of the United States of America, 2014.

[7] L. Ma, Kim Jungwoo, R. Hatzenpichler, M.A. Karymov, N. Hubert, I.M. Hanan, E.B. Chang, R.F. Ismagilov, “Gene-targeted microfluidic cultivation validated by isolation of a gut bacterium listed in Human Microbiome Project's Most Wanted taxa,” Proceedings of the National Academy of Sciences of the United States of America, pp. 9768-73, 2014.

[8] M.G. Rooks, P. Veiga, L.H. Wardwell-Scott, T. Tickle, N. Segata, M. Michaud, C.A. Gallini, C. Beal, J.E.T. van Hylckama-Vlieg, S.A. Ballal, X.C. Morgan, J.N. Glickman, D.G. Huttenhow, “Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission,” The ISME journal, pp. 1403-17, 2014.

[9] J.F. Spinler, A. Sontakke, E.B. Hollister, S.F. Venable, L. Phaik, M.A. Balderas, D. M. A. Saulnier, T.A. Mistretta, S. Devaraj, J. Walter, J. Versalovic, S. K. Highlander, “From prediction to function using evolutionary genomics: human-specific ecotypes of Lactobacillus reuteri have diverse probiotic functions,” Genome biology and evolution, pp.1772-89, 2014.

[10] L. Ma, S.S. Datta, M. A. Karymov, Q. Pan, S. Begolo, R.F. Ismagilov, “Individually addressable arrays of replica microbial cultures enabled by splitting SlipChips,” Integrative biology: quantitative biosciences from nano to macro, 2014.

[11] C. Huttenhower, A.D. Kostic, R. Xavier, J. “Immunity, Inflammatory bowel disease as a model for translating the microbiome,” pp. 843-54, 2014.

[12] E.A. Franzosa, X.C. Xuan Zhang, “The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment,” Nature Medicine, 2015.

[13] V. D’Argenio, Francesco Salvatore, “The role of the gut microbiome in the healthy adult status,” Clinica Chimica Acta, 2015.

[14] S. Mandal, et al., “Analysis of composition of microbiomes: a novel method for studying microbial composition,” Microbial Ecology in Health and Disease, Vol. 26, pp. 1 – 7, 2015.

[15] G. Musso, R. Gambino, M. Cassader, “Gut microbiota as a regulator of energy homeostasis and ectopic fat deposition: mechanisms and implications for metabolic disorders,” Curr. Opin. Lipidol., Vol. 21, pp. 76-83, 2010.

[16] World wide web: who.int/diabetes/eng.

[17] World wide web: euro.who.int/en/health-topics/ noncommunicable-diseases/diabetes/data-and-statistics.

[18] World wide web: www.societate-diabet.ro.

[19] N. Larsen, F.K. Vogensen van den Berg, et al., “Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults,” PLoS One 5;5(2):e9085, 2010.

[20] P.J. Turnbaugh, R.E. Ley, et al., ”The human microbiome project,” Nature 449, pp. 804-810, 2007.

[21] C.L. Maynard, C.O. Elsom, R.D. Hatton, et al., “Reciprocal interactions of the intestinal microbiota and immune system,” Nature 489, pp. 231-241, 2012.

[22] H.J. Flint, K.P. Scott, S.H. Duncan, “The role of gut microbiota in nutrition and health,” Nature Reviews Gastroenterology and Hepatology Vol. 9, pp. 577-589, 2012.

[23] H.Y. Lee, J.H. Park, S.H. Seok, M.W. Baek, D.J. Kim, K.E. Lee, K.S. Paek, Y. Lee, “Human originated bacteria, Lactobacillus rhamnosus PL60, produce conjugated linoleic acid and show anti-obesity effects in diet-induced obese mice,” Biochim Biophys Acta. 761(7), pp. 736-44, 2006.

[24] P.J. Turnbaugh, R.E. Ley, M.A. Mahowald, V. Magrini, E.R. Mardis, J.I. Gordon, “An obesity-associated gut microbiome with increased capacity for energy harvest,” Nature 444, pp. 1027-1031, 2006.

[25] J. Qin, Y. Li, Z. Cai, et al., “A metagenome-wide association study of gut microbiota in type 2 diabetes,” Nature, Vol. 490, Issue 55, 2012.

[26] A. Vrieze, E. Van Nood, F. Holleman, et al., “Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome,” Gastroenterology, Vol. 143, pp. 913-916, 2012.

[27] X. Zhang, Y. Zhao, Y. Zhang, et al., “Structural changes of gut microbiota during berberine-mediated prevention of obesity and insulin resistance in high-fat diet-fed rats,” PLoS One, Vol. 7, e42529, 2012.

[28] C. He, Y. Shan, W. Song, “Targeting gut microbiota as a possible therapy,” Nutrition Research, Vol. 35, pp. 361-367, 2015.

[29] M. Le Barz, F.E. Anhe, T. Varin, et al., “Probiotics as complementary treatment for metabolic Disorders,” Diabetes and Metabolism Journal, Vol. 39, pp. 291-303, 2015.

[30] Y. Wu, Y. Ding, Y. Tanaka, “Risk Factors Contributing to Type 2 Diabetes and Recent Advances in the Treatment and Prevention,” Int J Med Sci, Vol.11, pp. 1185-1200, 2014.

[31] V. Tremaroli, F. Backhed, “Functional interactions between the gut microbiota and host metabolism,” Nature, Vol. 489, pp. 242-249, 2012.

[32] A. Everard, P. Cani, “Diabetes, obesity and gut microbiota,” Best Practices and Research Clinical Gastroenerology, Vol. 27, pp. 73-83, 2013.

[33] C. Apostolescu, R. Moroti, V. Molagic, et al., “Gut microbiota and its complex role. The experience of the National Institute for Infectious Diseases “Prof. Dr. Matei Balş” in fecal bacteriotherapy for Clostridium difficile infection,” BMC Infect Dis.13 (Suppl 1): O19, 2013.

[34] WEKA, Weka 3: Data Mining Software in Java, cs.waikato.ac.nz/ml/weka, 2016.

[35] Project: "Analysis of novel risk factors influencing control of food intake and regulation of body weight,” PN-II-ID-PCE-2012-4-0608 no. 48/02.09.2013, (eed.usv.ro/idei_48).

[36] M.M. Lungu, A. Bosancu, O. Geman, “Mini-review: Human Microbiome Project – Recent Trends and Future Challenges,” The 5th IEEE International Conference on E-Health and Bioengineering - EHB 2015, pp.1-4, 2015.

[37] J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural Nets., Vol. 61, pp. 85-117, 2015.

[38] Y. LeCun, Y. Bengio, G. Hinton, “Deep learning,” Nature, Vol. 521, No. 7553, pp. 436-444, 2015.

[39] O. Geman, “Data Mining Tools Used in Deep Brain Stimulation– Analysis Results,” Artificial Intelligence Applications and Innovations, pp. 259-264, Editor Springer Berlin Heidelberg, 2011.

[40] O. Geman, C. Turcu, “Partitioning methods used in DBS treatments analysis results,” Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA, July 31 – August 5, 2011, pp. 1788-1793, 2011.