Proceedings of the 3rd International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/24/20231135

GMCVD: Detecting key modulating gut microbiota that contribute the cardiovascular disease risk toward personalized prevention

Daiying Li

Branksome Hall, Toronto, Canada

simon.yang@embarkchina.org

Abstract. Nowadays, cardiovascular disease (CVD) is one of the leading causes of death and disabilities worldwide. Atherosclerotic cardiovascular disease (ASCVD) is one of the most life-threatening subtypes of CVD. Recently, an increasing number of studies start to focus on the correlation and prediction of CVD based on the information of gut microbiome. In this study, by applying explanatory machine learning model, random forest-based computational pipeline called Gut Microbiome for CardioVascular Disease (GMCVD) was developed to conduct the ASCVD prediction and feature ranking. The top key several modulating gut microbiotas from genus and species levels Ire identified based on their strong contribution and correlation to the risk of CVD. These key disrupted microbiotas Ire validated by several external experimental studies, which demonstrate the reliability and efficiency of the machine learning based CVD risk prediction pipeline. With the detection of those specific modulating gut microbiotas, the personalized prevention method to reduce ASCVD risk using probiotics is provided based on varied microbiotas including Bifidobacterium, Clostridium, and Bacteroides. Therefore, GMCVD will facilitate the personalized prevention via gut microbiota to reduce the risk of cardiovascular disease.

Keywords: Gut Microbiome, Cardiovascular Disease, Machine Learning, Probiotics, Feature Selection.

1. Introduction

Cardiovascular disease (CVD) refers to the conditions affecting heart or blood vessels, including coronary heart diseases, strokes, peripheral arterial diseases, and aortic diseases, etc. (Figure 1) [1-3]. Serving as the leading cause of death and disability worldwide, CVD affects people of all ages, sex, ethnicities and socioeconomic levels [4-7]. The common risk factors of CVD, like high blood pressure, high cholesterol, and diabetes, will irreversibly damage the vascular structure and eventually lead to detrimental clinical outcomes like heart attack, angina, or stroke. The gut microbiome, made up of trillions of bacteria, fungi and other microbes, is a complex ecosystem that can mediate the interaction of the human host with their environment and maintain the physiological and metabolic health of the host [8-11]. As the understanding of the composition of gut microbiome becomes increasingly clear and profound, gut microbiome is already found to play a crucial role and have strong interconnections with multi-subtypes of CVD [12-17]. Atherosclerotic cardiovascular disease, as one of the subtypes of

DOI: 10.54254/2753-8818/24/20231135

cardiovascular disease, accounts for more than 70% of sudden death outside of hospital among all heart diseases, which underscores the seriousness of the disease [18-21].

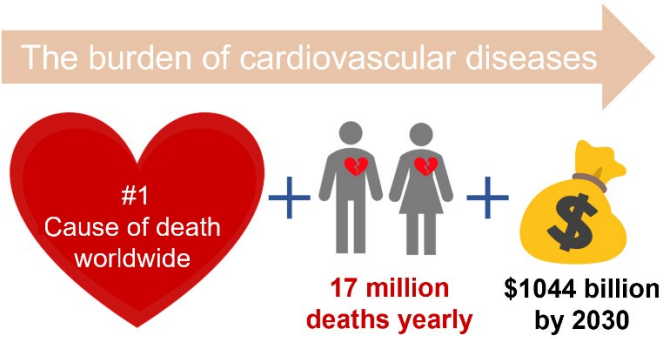

Figure 1. The burden of cardiovascular disease.

Recently, based on a metagenome-wide association study on atherosclerotic cardiovascular disease (ASCVD) [22], it is found that increased abundance of Enterobacteriaceae and Streptococcus exist in ASCVD patients, which Ire critical for studying the molecular transport and metabolism to promote cardiovascular health. Fredrik H. Karlsson et al. applied shotgun sequencing of the gut metagenome to demonstrate that bacteria Collinsella is a causative agent of Atherosclerosis [23]. These studies highlight the potential of identification of significant microbial taxa associated with ASCVD and increasing evidence that the microbiome may contribute to ASCVD pathogenesis. Machine learning (ML) has been increasingly utilized for disease risk prediction based on gut microbiome data [24]. A large scale gut microbiome data and machine learning algorithm have been used to find the core correlation betIen gut microbiome and type 2 diabetes risk and potential demographic pattern [25]. Also, machine learning based predictions of CVD using gut microbiota data sets Ire conducted to predict the risk of CVD [26].

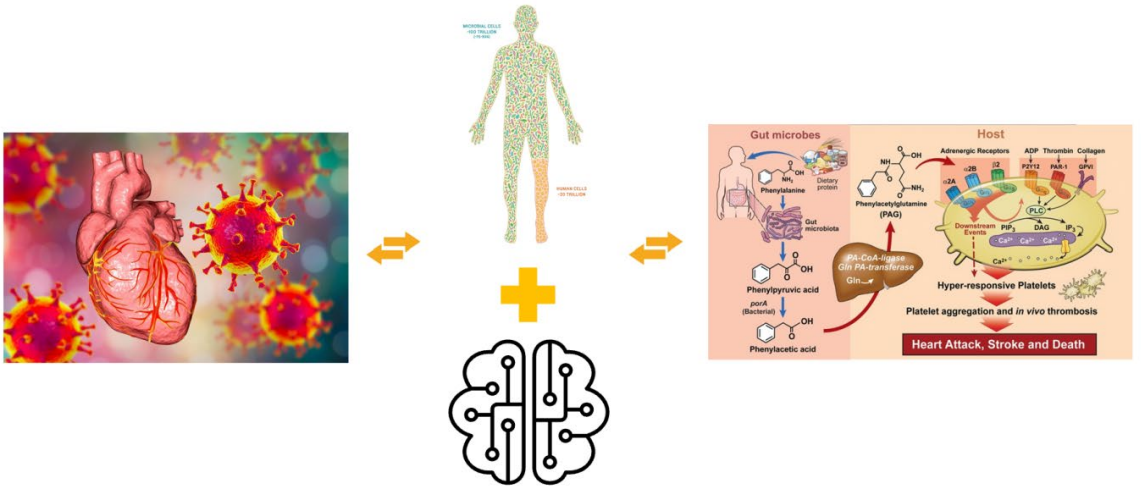

Figure 2. Axis of the CVD-gut microbiome association [16].

Though a few ML-based methods have been developed to predict risk of ASCVD, in-depth explainable ML model was rarely used, and the gut microbiota feature importance ranking was understudied, which was important for developing a gut microbiome-based ML approach for diagnostic screening (Figure 2). In this study, I developed a highly accurate and explainable random forest model to capture the most important gut microbiota associated with ASCVD. The prediction accuracy achieved 83.5%. The key interrupted microbiota in CVD risk Ire ranked based on the feature importance. In addition, since probiotics are capable to improve microbial balance in the human gut, the recommendation for different ASCVD risk population was proposed to reduce the ASCVD risk using

DOI: 10.54254/2753-8818/24/20231135

probiotic prevention. The study provides a highly efficient pipeline to identify the disrupted gut microbiota associated with ASCVD using an explainable machine learning approach, which facilitates the study for precise and personalized prevention of ASCVD.

2. Methods

2.1. Dataset

The raw dataset was extracted from GMrepo (https://gmrepo.humangut.info/), which contains the gut microbiome data of 214 subjects with atherosclerotic cardiovascular disease and 171 healthy subjects [27]. The microbiome profiling includes data from different taxonomic levels, including 144 features from genus level and 458 features from species level.

2.2. Machine learning model

The dataset was split to training (70%) and test (30%) cohort in the model development. I first constructed five machine learning models to predict the risk of CVD based on the gut microbiome data, including random forest (RF), support vector machine (SVM), k-Nearest Neighbors (KNN), linear discriminant analysis (LDA), gradient boosting (GB), and Xgboosting (XGB). Random forest, composed of an ensemble of decision trees that separate the samples into groups with similar y values, performs bootstrapping on the dataset and taking the model prediction from the trees with subsampling of features [28]. For the support vector machine, in a high-dimensional space, it locates the hyperplane that embodies the largest margin betIen any two instances of any two classes of training-data points [29]. KNN is applied to predict the correct class for the test data by calculating the distance betIen the test data and all the training points, and to select the K number of points which is closest to the test data [29]. In addition, linear discriminant analysis identifies a linear combination of Operational Taxonomic Units (OTUs) from the training data that represents the multivariate mean differences betIen classes [29]. Gradient boosting and Xgboosting methods use ensemble learning approach that involve average predictions over fixed-size decision tree learners [29].

2.3. Feature selection and feature importance ranking

After comparing the prediction performance by multiple machine learning models, the machine learning model with the best prediction accuracy and ROC-AUC score was selected. The Recursive Feature Elimination (RFE) was implemented that iteratively eliminates the features one by one. Top ten features, i.e., gut microbiota, at genus and species level was selected and ranked based on the prediction performance.

2.4. Model evaluation

To evaluate the prediction performance of ASCVD. Multiple evaluation metrics Ire used, including accuracy, precision, recall, specificity, false positive rate, confusion matrix, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC), which are defined as follows. The true positive (TP) is the result that correctly indicates the presence of a ASCVD case. True negative (TN) is the result that correctly indicates the healthy control. The false positive (FP) is the result that classifies a healthy control as a ASCVD case. The false negative (FN) is the result that classifies a ASCVD case as a healthy control.

$$Accuracy = TP + TNTP + TN + FP + FN \tag{1}$$

$$Precision = TPTP + FP \tag{2}$$

$$Recall = TPTP + FN \tag{3}$$

$$Specificity = TNTN + FP \tag{4}$$

AUC-ROC measures the prediction performance by plotting the True Positive Rate (Recall) against the False Positive Rate at various threshold values. It provides an aggregate measure of the model's ability to distinguish betIen ASCVD and healthy control.

DOI: 10.54254/2753-8818/24/20231135

2.5. Study design of the workflow

In the study, the gut microbiome data of 214 subjects with atherosclerotic cardiovascular disease and 171 healthy subjects Ire included from the public database. Figure 3 shows the study design and workflow for the GMCVD pipeline. The study first conducted the metagenomics sequencing on feces that Ire extracted from each of the subjects. After data preprocessing, a gut microbiome profile for each individual was obtained, which revealed the gut microbiome abundances for each subject. Then, rigorous data processing and analysis Ire conducted including dimension reduction, machine learning modeling and functional analysis. Six machine learning models Ire first compared to predict the risk of ASCVD based on the gut microbiome data, including random forest (RF), support vector machine (SVM), k-Nearest Neighbors (KNN), linear discriminant analysis (LDA), gradient boosting (GB), and Xgboosting (XGB). The model with the best predictive accuracy and ROC-AUC score was selected. Then, feature importance ranking was conducted using the chosen ML model to evaluate the importance of these risk factors that Ire being selected. Finally, the functions of the selected microbiota Ire analyzed and produced a personalized ASCVD prevention recommendation.

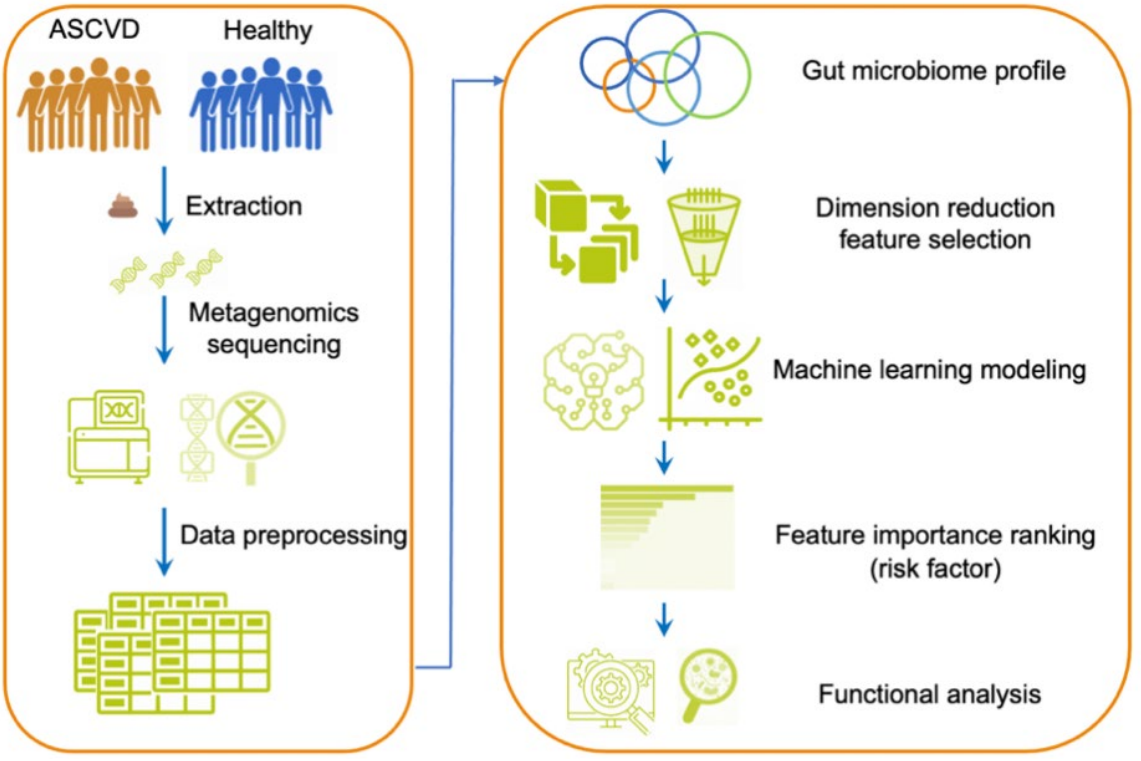

Figure 3. Workflow of GMCVD.

3. Results and Discussion

3.1. Comparison of gut microbiome profile betIen healthy control and ASCVD group

The gut microbiome profile betIen healthy control and ASCVD was compared to investigate the distribution of gut microbiota abundance. Figure 4 shows the boxplots of the abundance of the gut microbiota in healthy and ASCVD groups. Overall, higher levels of gut microbiota abundance are observed in the ASCVD groups in both genus and species level. Univariate statistical analysis betIen the two groups was performed. Figure 5 shows the volcano plots in genus and species level. Each microbiota existing in health and ASCVD groups was compared individually. It is observed from the plots that various microbiota in the ASCVD group have evident up and downregulated abundance levels.

DOI: 10.54254/2753-8818/24/20231135

Box plots comparing of gut microbiota abundance level in healthy control and ASCVD groups. Panel A and B show the gut microbiota abundance at genus and species level, respectively.

Volcano plots comparing of gut microbiota abundance level in healthy control and ASCVD groups. Panel A and B show the gut microbiota abundance at genus and species level, respectively.

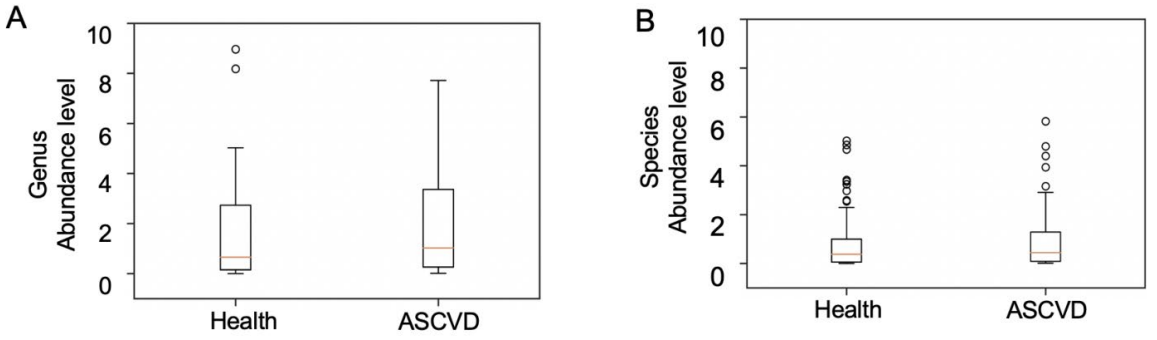

Figure 4. Box plots comparing of gut microbiota abundance level in healthy control and ASCVD groups. Panel A and B show the gut microbiota abundance at genus and species level, respectively.

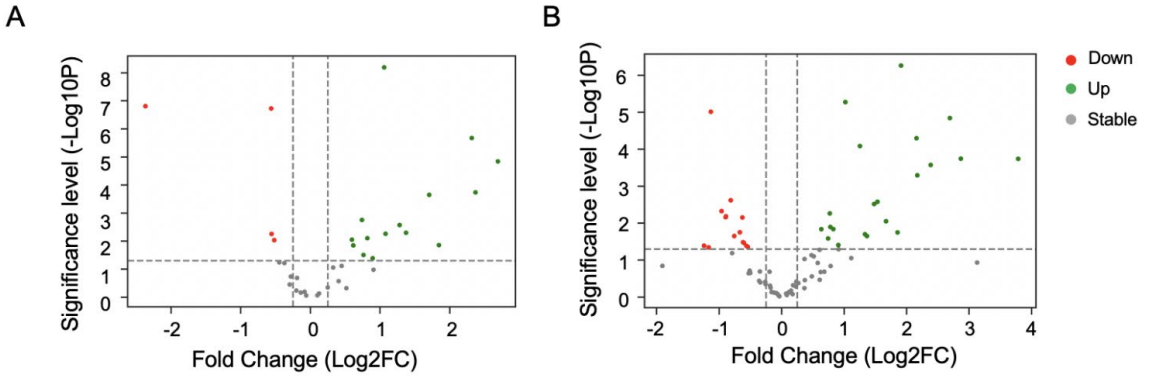

Figure 5. Volcano plots comparing of gut microbiota abundance level in healthy control and ASCVD groups. Panel A and B show the gut microbiota abundance at genus and species level, respectively.

Since gut microbiota are not independently existing in the biological system, their abundance implies inherent interactions. Thus, gut microbiota co-abundance network analysis is constructed for health controls and ASCVD subjects based on Pearson correlation coefficient. Figure 6 shows the gut microbiota co-abundance network in healthy control and ASCVD group. It is revealed by the comparison of the graph that there is significantly loIr gut microbiota co-abundance network complexity for ASCVD group in comparison to health controls. The less correlation edges shown betIen gut microbiota suggests that some microbiota's abundance may be disrupted by the effect of ASCVD, which causes the situation that these microbiotas did not function as normal as the healthy control group.

Gut microbiota co-abundance network analysis healthy control and ASCVD group. Panel A shows the gut microbiota co-abundance network at species level for healthy controls. Panel B shows the gut microbiota co-abundance network at species level for ASCVD cases. Each node represents one specific microbiota at species level. The line connecting these nodes refers to the correlations betIen these microbiotas, in which pink lines indicate positive correlations, while the green ones indicate negative correlations.

DOI: 10.54254/2753-8818/24/20231135

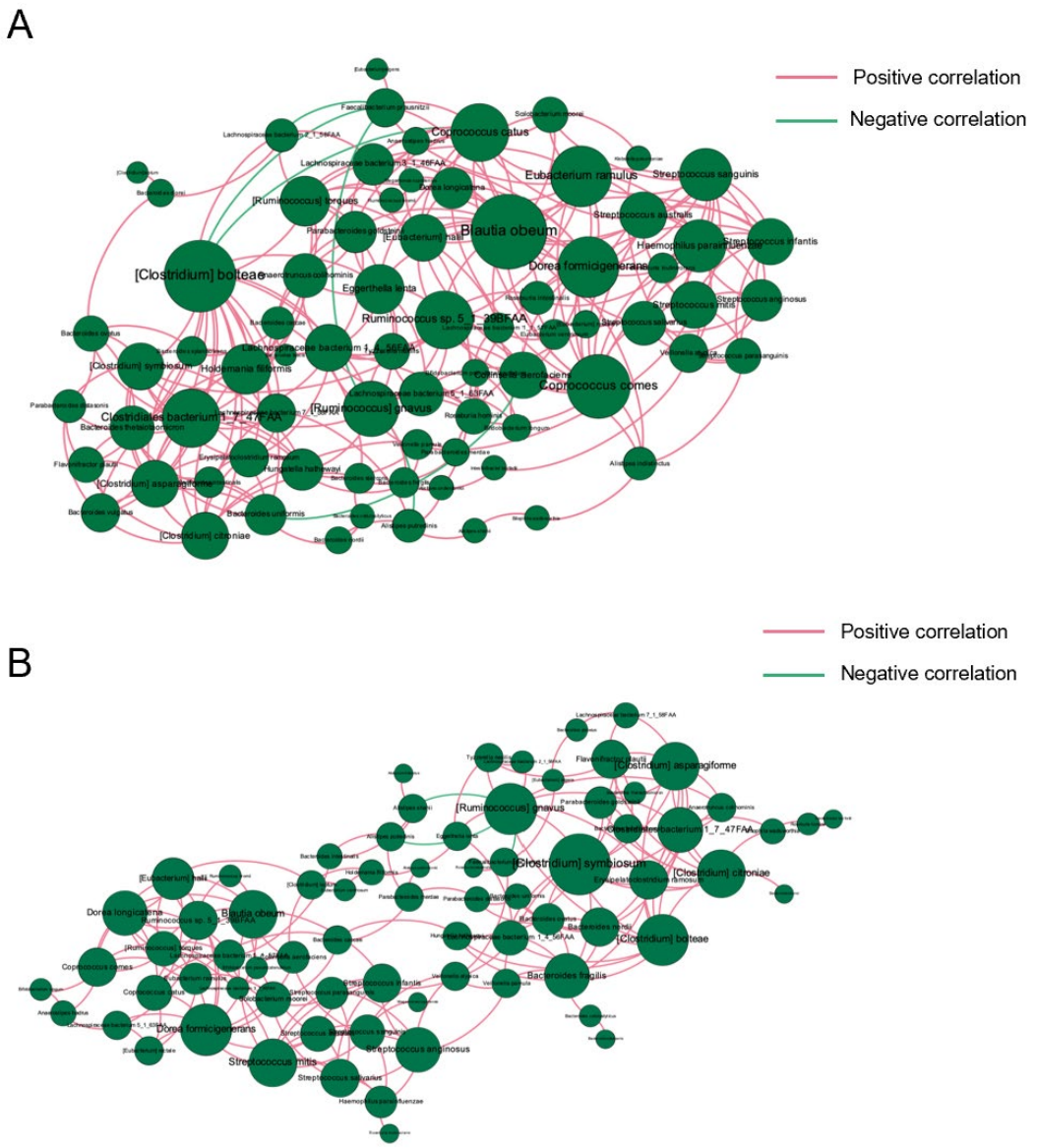

Figure 6. Gut microbiota co-abundance network analysis healthy control and ASCVD group. Panel A shows the gut microbiota co-abundance network at species level for healthy controls. Panel B shows the gut microbiota co-abundance network at species level for ASCVD cases. Each node represents one specific microbiota at species level. The line connecting these nodes refers to the correlations betIen these microbiotas, in which pink lines indicate positive correlations, while the green ones indicate negative correlations.

3.2. Machine learning model performance for prediction of ASCVD risk

I have evaluated and compared six machine learning models as described in the Methods section. The performance of different ML models is presented with their prediction accuracy and ROC-AUC score in Figure 7 and Table 1. According to the prediction results, random forest has the highest ROC-AUC score of 0.802 (genus) and 0.825 (species), indicating its potential of being selected as an accurate disease prediction ML model.

DOI: 10.54254/2753-8818/24/20231135

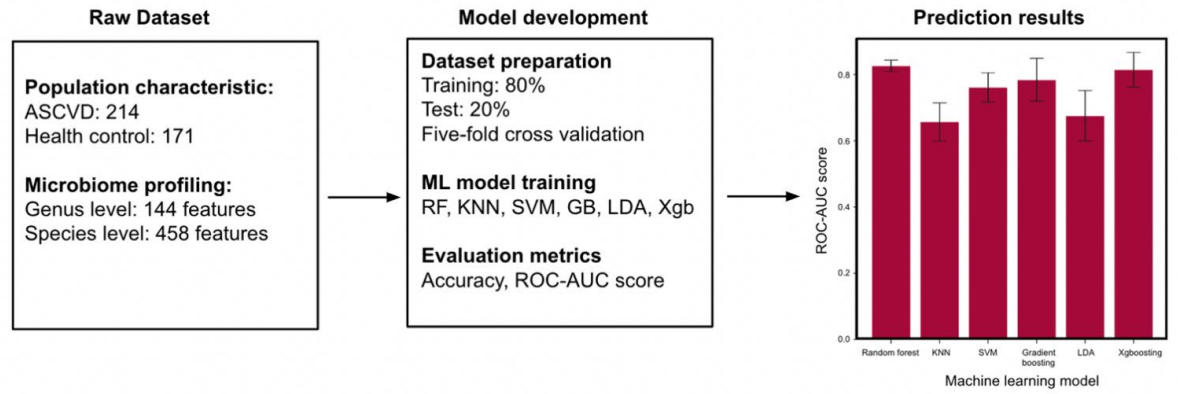

Figure 7. The development and performance of machine learning models for prediction of ASCVD based on gut microbiome data.

Table 1. Performance of ASCVD risk prediction by machine learning model.

Model	AUC	Accuracy	Precision	Recall	Specificity
Random forest	0.802/0.825*	0.698/0.690	0.697/0.689	0.698/0.690	0.703/0.679
KNN	0.764/0.674	0.664/0.612	0.668/0.649	0.664/0.612	0.666/0.645
SVM	0.778/0.712	0.724/0.629	0.723/0.630	0.724/0.629	0.711/0.623
LDA	0.693/0.670	0.655/0.595	0.669/0.601	0.655/0.595	0.669/0.599
Xgboosting	0.783/0.796	0.716/0.724	0.715/0.727	0.716/0.724	0.697/0.725
GB	0.773/0.745	0.638/0.740	0.637/0.680	0.638/0.681	0.626/0.658

*The scores are presented as score (genus level)/ score (species level).

After applying random forest to conduct feature importance ranking (Figure 8), the top ten key modulating microbiota from genus and species levels Ire identified from thousands of them. Specifically, Solobacterium, Blautia, Bifidobacterium, and Streptococcus Ire upregulated, which have been experimentally validated.

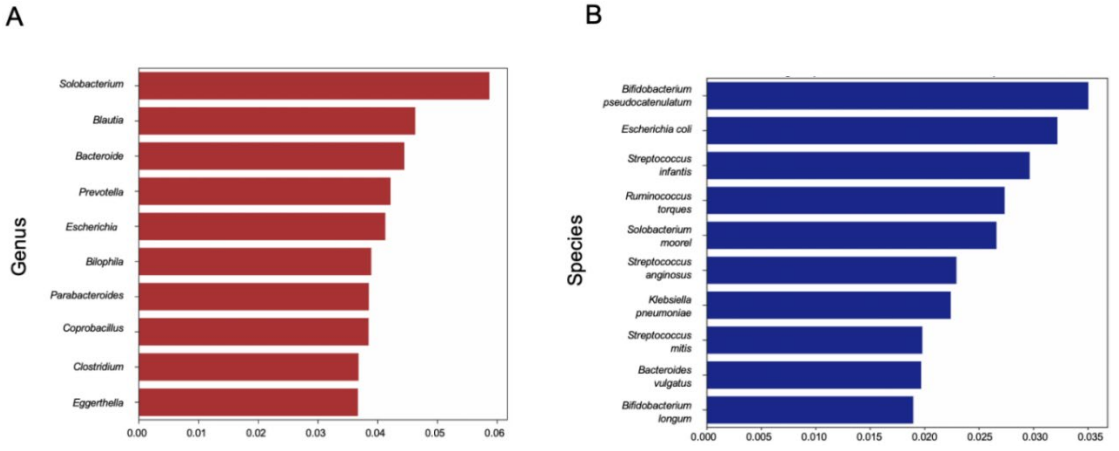

Figure 8. Feature importance ranking by random forest. Panel A shows the top microbiota at genus level for accurate prediction of ASCVD risk. Panel B shows the top microbiota at species level for accurate prediction of ASCVD risk.

DOI: 10.54254/2753-8818/24/20231135

I have validated the disruption of gut microbiota based on prior studies. S. Raj J. Trikha et al. conducted gut microbiota transplantation from lean or obese humans' results in different community structure in germ-free mice [30]. The external experimental results in the study Ire consistent with some of the top 10 key modulating microbiota, such as Bifidobacterium and Bacteroides. As shown in Table 2 below, some of other key microbiota in genus level are also able to be validated by some external It lab studies, which strongly support the biological accuracy and explainability of the development of my machine learning model.

Table 2. Key bacteria that Ire validated to be associated with CVD.

Key bacteria	CVD risk	Source
Solobacterium	Atherosclerotic cardiovascular disease	van den Munckhof, Inge CL, et al. Obesity reviews 2018
Blautia	Arterial hypertension	Kashtanova, Daria A., et al. Microorganisms 6.4 2018
Escherichia Coli	Molecular-levels interactions in cardiovascular disease	Miryala, Sravan Kumar, et al. Computers in Biology and Medicine 2021
Prevotella	Atherosclerotic cardiovascular disease	van den Munckhof, Inge CL, et al. Obesity reviews 2018
Bilophila	Strongly correlated to CVD occurence	Kivenson, Veronika, and Stephen J. Giovannoni. Msystems 2020

3.3. Key gut microbiota serve as probiotics for reducing risk of ASCVD

In this study, I have identified a few key microbiota that Ire substantial for the risk of ASCVD, including Bifidobacterium, Clostridium, and Bacteroides. For people with different subtypes of CVD, the regulation of a certain key microbiota may vary. For Bifidobacterium, people with ASCVD may experience an up regulation of the microbiota in CVD cases compared to health controls, while individuals with hypertension and ischemic or dilated cardiomyopathy may experience a down regulation [31]. For Clostridium, people with ASCVD and hypertension subtypes may experience up regulations, while individuals with vascular dysfunction may have a down regulation [30,31]. For Bacteroides, while subjects with vascular dysfunction may have an up regulation, those with ASCVD and stable angina and old myocardial infarction may experience down regulations [30,31]. In these cases, if people already have an up-regulated certain bacteria condition above, they need to avoid continuing taking more probiotics that contain the same microbiota, which might increase the risk of ASCVD.

Some key microbiotas have been applied as food supplement or nutritional probiotics, including Bifidobacterium, Clostridium, and Bacteroides. Probiotics are great supplements that are capable of improving microbial balance in the human gut, thus exerting positive health effects. For nutritional intervention, possessing a gut microbiome profile, for those down-regulated in ASCVD cases, people can intake more of that certain microbiota, while up-regulated ones should avoid further intake.

There exists a couple of limitations that could be improved. The current study only considers one subtype of CVD and its association with gut microbiome. Also, the machine learning model only applies to ASCVD. Other subtype of CVD such as coronary heart disease needs to further be investigated for more accurate risk prediction. The mechanisms behind the correlation betIen ASCVD and those discovered key modulating gut microbiota Ire not included in the study. It is recommended to conduct molecular experiments to study the mechanism.

4. Conclusions

Overall, in the study, I developed an efficient and explainable machine learning pipeline GMCVD for accurate ASCVD risk prediction using gut microbiome profiles. The key modulating gut microbiota Ire

identified along with their functional mechanisms in the system. It is confirmed that the gut microbiome is substantially associated with ASCVD. The interactions of some key microbiota and ASCVD Ire revealed by machine learning and topological network analysis. Key microbiota including Bifidobacterium, Clostridium, and Bacteroides Ire validated to be associated with ASCVD. As a nutritional intervention recommendation, people with high risk of ASCVD should take caution when taking these probiotics. I believe this study would promote the risk prevention of ASCVD and facilitate the machine learning-based prediction of CVD by gut microbiome profiles.

References

[1] Nabel, E. G. Cardiovascular Disease. N. Engl. J. Med. 349, 60–72 (2003).
[2] Mensah, G. A., Roth, G. A. & Fuster, V. The Global Burden of Cardiovascular Diseases and Risk Factors. J. Am. Coll. Cardiol. 74, 2529–2532 (2019).
[3] Steptoe, A. & Kivimäki, M. Stress and cardiovascular disease. Nat. Rev. Cardiol. 9, 360–370 (2012).
[4] Anderson, K. M., Odell, P. M., Wilson, P. W. F. & Kannel, W. B. Cardiovascular disease risk profiles. Am. Heart J. 121, 293–298 (1991).
[5] Berry, J. D. et al. Lifetime Risks of Cardiovascular Disease. N. Engl. J. Med. 366, 321–329 (2012).
[6] Kivimäki, M. & Steptoe, A. Effects of stress on the development and progression of cardiovascular disease. Nat. Rev. Cardiol. 15, 215–229 (2018).
[7] Oni, E. T. et al. A systematic review: Burden and severity of subclinical cardiovascular disease among those with nonalcoholic fatty liver; Should I care? Atherosclerosis 230, 258–267 (2013).
[8] The gut microbiome in atherosclerotic cardiovascular disease | Nature Communications. https://www.nature.com/articles/s41467-017-00900-1.
[9] Rajendiran, E., Ramadass, B. & Ramprasath, V. Understanding connections and roles of gut microbiome in cardiovascular diseases. Can. J. Microbiol. 67, 101–111 (2021).
[10] Shreiner, A. B., Kao, J. Y. & Young, V. B. The gut microbiome in health and in disease. Curr. Opin. Gastroenterol. 31, 69–75 (2015).
[11] Cresci, G. A. & Bawden, E. Gut Microbiome. Nutr. Clin. Pract. 30, 734–746 (2015).
[12] Gut Microbiota and Cardiovascular Disease | Circulation Research. https://www.ahajournals.org/doi/full/10.1161/CIRCRESAHA.120.316242.
[13] Ahmad, A. F., Dwivedi, G., O'Gara, F., Caparros-Martin, J. & Ward, N. C. The gut microbiome and cardiovascular disease: current knowledge and clinical potential. Am. J. Physiol.-Heart Circ. Physiol. 317, H923–H938 (2019).
[14] The Gut Microbiome and Its Role in Cardiovascular Diseases | Circulation. https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.116.024251.
[15] Kelly, T. N. et al. Gut Microbiome Associates With Lifetime Cardiovascular Disease Risk Profile Among Bogalusa Heart Study Participants. Circ. Res. 119, 956–964 (2016).
[16] Gut Microbiota and Cardiovascular Disease | Circulation Research. https://www.ahajournals.org/doi/full/10.1161/CIRCRESAHA.120.316242.
[17] The Gut Microbiome Contributes to a Substantial Proportion of the Variation in Blood Lipids | Circulation Research. https://www.ahajournals.org/doi/full/10.1161/CIRCRESAHA.115.306807.
[18] Nayor, M., Brown, K. J. & Vasan, R. S. The Molecular Basis of Predicting Atherosclerotic Cardiovascular Disease Risk. Circ. Res. 128, 287–303 (2021).
[19] Ellulu, M. S. et al. Atherosclerotic cardiovascular disease: a review of initiators and protective factors. Inflammopharmacology 24, 1–10 (2016).
[20] Hong, Y. M. Atherosclerotic Cardiovascular Disease Beginning in Childhood. Korean Circ. J. 40, 1–9 (2010).
[21] Scannapieco, F. A., Bush, R. B. & Paju, S. Associations BetIen Periodontal Disease and Risk for Atherosclerosis, Cardiovascular Disease, and Stroke. A Systematic Review. Ann. Periodontol. 8, 38–53 (2003).

[22] Jie, Z. et al. The gut microbiome in atherosclerotic cardiovascular disease. Nat. Commun. 8, 845 (2017).
[23] Karlsson, F. H. et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat. Commun. 3, 1245 (2012).
[24] Krittanawong, C. et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. Sci. Rep. 10, 16057 (2020).
[25] Gou, W. et al. Interpretable Machine Learning Framework Reveals Robust Gut Microbiome Features Associated With Type 2 Diabetes. Diabetes Care 44, 358–366 (2021).
[26] Aryal, S., Alimadadi, A., Manandhar, I., Joe, B. & Cheng, X. Machine Learning Strategy for Gut Microbiome-Based Diagnostic Screening of Cardiovascular Disease. Hypertens. Dallas Tex 1979 76, 1555–1562 (2020).
[27] Dai, D. et al. GMrepo v2: a curated human gut microbiome database with special focus on disease markers and cross-dataset comparison. Nucleic Acids Res. 50, D777–D784 (2022).
[28] Frontiers | A Review and Tutorial of Machine Learning Methods for Microbiome Host Trait Prediction. https://www.frontiersin.org/articles/10.3389/fgene.2019.00579/full.
[29] What is a support vector machine? | Nature Biotechnology. https://www.nature.com/articles/nbt1206-1565.
[30] Trikha, S. R. J. et al. Transplantation of an obesity-associated human gut microbiota to mice induces vascular dysfunction and glucose intolerance. Gut Microbes 13, 1940791 (2021).
[31] Oniszczuk, A., Oniszczuk, T., Gancarz, M. & Szymańska, J. Role of Gut Microbiota, Probiotics and Prebiotics in the Cardiovascular Diseases. Molecules 26, 1172 (2021).