ueg week

TU WIEN BiomeDx

October 8–11
ueg.eu

Analysis of colorectal cancer and adenoma microbiome signatures and the application of machine learning classification as a potential screening tool

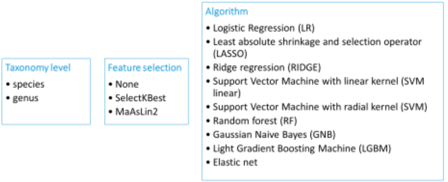
K. Priselac[1], C. Jansen[2], C. Pacífico[2], B. Sladek[2], N. Gasche[2]
[1]TU Wien, Faculty of Technical Chemistry, Institute of Chemical, Environmental and Bioscience Engineering, Vienna, Austria
[2]Biome Diagnostics GmbH, Vienna, Austria.

Introduction

- Colorectal cancer (CRC) is the 3rd most common and the 2nd most fatal cancer in the world.
- Gut microbiota were shown to be dysregulated in gut diseases, including CRC. Thus, microbiome composition has a high potential for employment in the diagnosis and treatment of CRC
- The aims of this study were to identify microbial signatures of CRC and colorectal adenoma and to optimise machine learning (ML) algorithms for the early screening of CRC based on the stool microbiome.

Methods

- Meta-analysis dataset was used consisting of amplicon sequence reads from 1786 samples from 4 datasets
- Differential abundance analysis (DAA) was conducted using the MaAsLin2 package to identify microbial biomarkers of CRC and colorectal adenoma
- Machine learning (ML) pipeline was written in Python using the sklearn library
- Following parameters have been evaluated to find the best performing model:



Taxonomy level
- species
- genus

Feature selection
- None
- SelectKBest
- MaAsLin2

Algorithm
- Logistic Regression (LR)
- Least absolute shrinkage and selection operator (LASSO)
- Ridge regression (RIDGE)
- Support Vector Machine with linear kernel (SVM linear)
- Support Vector Machine with radial kernel (SVM)
- Random forest (RF)
- Gaussian Naive Bayes (GNB)
- Light Gradient Boosting Machine (LGBM)
- Elastic net

Results

- DAA of CRC vs. healthy identified 40 differentially abundant taxa ($q < 0.05$, abs(coeff) > 1.5) with most of the taxa commonly detected as intestinal biomarkers of CRC in the literature
- DAA of adenoma vs. healthy identified only one bacterial family, enriched in adenoma samples. This taxon has not been detected in adenoma samples in previous studies

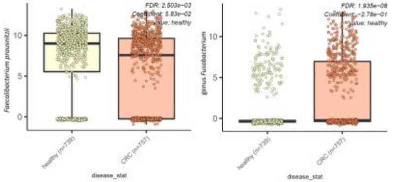


Figure 1: MaAsLin2 plots for two detected differentially abundant taxa; *Faecalibacterium prausnitzii* is enriched in healthy samples and *Fusobacterium* is enriched in CRC samples

- The best performing model for CRC vs. healthy classification was the SVM model on a genus level with MaAsLin2 feature selection method. It resulted in the AUC of 0.843, sensitivity of 0.724 and specificity of 0.824

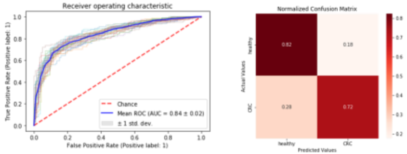


Figure 2: Receiver operating characteristic curve and confusion matrix for the best performing CRC vs. healthy model

- The best performing model for adenoma vs. healthy classification was the LGBM model on a species level with SelectKBest feature selection method. It resulted in the AUC of 0.853, sensitivity of 0.556 and specificity of 0.852

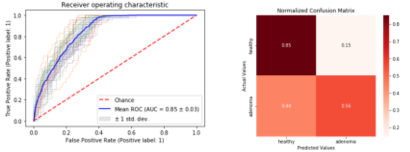


Figure 3: Receiver operating characteristic curve and confusion matrix for the best performing adenoma vs. healthy model

Conclusion

- When compared to other non-invasive methods, the developed ML models are superior to gFOBT at both CRC and adenoma detection and superior to FIT at adenoma detection. Detection at an adenoma stage is crucial for increasing the success rate of the treatment

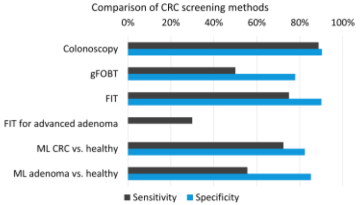


Figure 4: Comparison of sensitivity and specificity between the most frequently used CRC screening methods and the developed ML models

- For the first time, a large meta-analysis dataset has been used successfully to demonstrate the suitability of machine learning algorithms for the development of microbiome-based solutions for the non-invasive, early screening of CRC and colorectal adenoma

Sources: Martín-López, J. E., Beltrán-Calvo, C., Rodríguez-López, R. & Molina-López, T. Comparison of the accuracy of CT colonography and colonoscopy in the diagnosis of colorectal cancer (2014); Elsafi, S. H., Alqahtani, N. I., Zakary, N. Y. & al Zahrani, E. M. The sensitivity, specificity, predictive values, and likelihood ratios of fecal occult blood test for the detection of colorectal cancer in hospital settings (2015).

Contact: e11724446@student.tuwien.ac.at

All authors have declared no conflict of interest