A Prediction Model for Detecting Colorectal Cancer and Identifying Biomarkers From the Gut Microbiome

Ganapathy, Ananya (School: The International School of Bangalore (TISB))

Colorectal cancer (CRC) ranks as the second leading cause of cancer-related deaths worldwide. Although CRC incidence rates have traditionally been the highest in developed nations, they are now increasing in developing countries as well. Recent research presents a positive correlation between gut microbiome dysbiosis and CRC development, with certain bacteria being identified as potential oncomicrobes. These oncomicrobes have been associated with DNA damage, inflammation, and the production of metabolites potentially leading to tumour growth, resulting in the development of CRC. While studies relating the gut microbiome to CRC are available for the Western populations, there is limited data from developing countries like India. Hence, this study presents a machine learning model to analyse microbiome datasets of the Indian population. As a part of this, I initially performed normalisation and dimensionality reduction, followed by diversity analyses, then training the dataset using the Random Forest classification model. The model was evaluated on several metrics, where it achieved an AUC value of 0.992 and an accuracy score of 0.947. Further, the diversity analyses also revealed the potential of certain microbes like Ruminococcaceae UCG-002, Christensenellaceae R-7 group, Streptococcus, Prevotella-2 and Escherichia Shigella to serve as biomarkers for the early detection of CRC on an Indian population. In conclusion, this study proposes a machine learning model that can effectively distinguish between CRC and non-CRC samples. Additionally, it also identifies microbial genera associated with CRC patients, which could serve as diagnostic biomarkers or be used in therapeutic strategies following large-scale validation.

Awards Won: