Detecting sleep apnea from the metabolomic profile

Sleep apnea is a serious disease resulting in sleep-deprived individuals and complications in breathing while sleeping. Both being highly dangerous. Sleep apnea may result in heart attack, stroke, diabetes, heart failure, irregular heartbeat, obesity, motor vehicle collisions. At the same time, cardiovascular disease is associated with high blood pressure, smoking, diabetes, lack of exercise, obesity, high blood cholesterol, poor diet, and excessive alcohol consumption. Therefore, it is of interest if these diseases can be associated with the metabolic profile of the patient.

Recruitment of subjects took place from the clinic population in collaboration with the Sleep Center X. Patients newly diagnosed with sleep apnea were solicited via an overnight sleep study, which is general monitoring of sleep and a variety of body functions during sleep. This includes the brain (EEG), eye (EOG), and muscle activity, and heart rhythm. The diagnosis of sleep apnea was established by a full sleep study.

The information about height, weight, medical history, demographics, gender, BMI, habits of physical exercises, smoking, age, medications (name/duration/frequency), and presence of comorbidities (e.g. hypertension, CVD defined by doctor diagnosis of myocardial infarction, stroke, or heart failure) have been recorded but are currently not available for data protection reasons.

For the metabolic markers, 9 ml blood was drawn into 3 EDTA tubes (3 ml each) and gently centrifuged at 4˚C for 10 min; plasma aliquots were immediately frozen and stored at -80 ˚C; were stored at -80°C in the CTRC until all samples are collected and then sent to LC-MS/MS analysis for simultaneous detection and quantification of metabolites for analyses.

The information about the analysis:

Analysis type	MS
Chromatography type	HILIC
Chromatography system	Cohesive TX2
Column	Altma HP HILIC, 2.1x150mm, 5um
MS Type	ESI
MS instrument type	Triple quadrupole
MS instrument name	ABI Sciex 6500 QTrap
Ion Mode	POSITIVE
Units	micromolar

You are given the concentrations of betaine, carnitine, choline, tma, tmao determined for 18 sleep apnea, and 18 cardiovascular disease patients. Your task is to help the doctors decide if a metabolic profile can be used to differentiate between these diseases?

Tasks

We will develop a decision tree to decide based on the metabolomic profile if the patient suffers from sleep apnea or cardiovascular disease.

(1) Develop a decision tree model to predict the metabolic profile. Visualize the decision tree. Explain what you see.

(2) Use your model to predict the disease from the metabolic profile. Use probabilistic predictions.
(3) Plot the ROC curve and evaluate if the model is performing better than a random guess.
(4) Analyze, what is more important: to avoid false positives or false negatives. How costly is either? Now, choose the suitable cut-off value based on this information.
(5) Develop also a random forest model to predict the disease. Investigate from the caret page which hyperparameters can be optimized within train function. Prepare a manual grid of hyperparameter values that you would like the train function to test and use it within the train function.
(6) Analyze the result with the ROC curve, compare with the previous model.
(7) Random forest is not one algorithm but a collection of multiple algorithms (trees). Therefore, you can not deduce an analytical model. However, you can investigate the importance of each of the features. Evaluate manually if the importance value smake sense.

Anneli