Discovery of Biomarkers for Coronary Microvascular Disease

Saumya Agrawal, Alicia Arredondo Eve, Justina Zurauskiene, and Zeynep Madak-Erdogan

Student Pushing Innovation Program, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign

INTRODUCTION

Coronary microvascular disease (CMD) is a heart disease that affects the walls and inner lining of tiny coronary artery blood vessels that branch off from the larger coronary arteries. This research is vital because although CMD is vastly common in women, the only method of identification we have is the elimination of the other possibilities such as coronary artery disease.

Biomarkers

The term “biomarker, refers to a broad subcategory of medical signs – that is, objective indications of medical state observed from outside the patient which can be measured accurately and reproducibly.

This research is vital because currently, the identification of CMD is subject to the elimination of the other similar possibilities such as Coronary Artery Disease (CAD).

With definite biomarkers for CMD, we could eliminate the need for a stress test, or a coronary angiography by optimizing testing and treatment.

AIM

DATA OVERVIEW

The raw data consisted of 71 patients and 410 features. These features were in the form of targeted and untargeted metabolites, and general health data.

RESULT

After the pre-processing, the data set with 40% missing data imputed had 265 features, and the data set with 20% missing data had 245 features.

After the difference in the difference in means t test, we found more than 60 potential biomarkers, including total cholesterol, choline, indole, valine, and creatinine.

CONCLUSIONS

More data needs to be collected for further testing and validation of these biomarkers. The machine learning accuracy was a lot higher than it should be for a data set this small. One sided tests also need to be performed to identify what biomarkers lean to what level for each group. The same code can be used for the discovery of biomarkers for other diseases.

ACKNOWLEDGEMENTS

We would like to acknowledge the National Center for Supercomputing Applications and their Students Pushing Innovation Program.