SYNOPSYS CHAMPIONSHIP 2017 PROJECT ABSTRACT

THE ABSTRACT IS A REQUIRED PART OF YOUR PROJECT

Bring your abstract with you to project check-in on Wednesday, March 22, 2017.

PROJECT NUMBER: 112-H43-21_____ STUDENT NAME(s): Kris Satya_____

You should bring at least ten copies of your abstract with you when you come to the Championship. One copy should remain on display with your project during the Championship. You will want others to give to the judges. Your abstract should be written after you finish your research and experimentation and should include:

- Your project title, the full name(s) of all team members, and your school (all centered).
- The purpose of your project
- Your hypothesis or evaluation criteria
- A brief statement about the procedures and equipment you used
- Your results (analysis of data)
- Your conclusions

Type or print neatly using 10- or 12- point black type. Single space throughout.
Center your project title, your name(s), and school.
You must use this form. Your abstract should be less than 500 words, and it should fit within the lines on this form.

Improved Classification of Coronary Artery Disease through a Machine Learning Algorithm

Lynbrook High School
Kris Satya

Problem

Coronary artery disease (CAD) is attributable to over 50% of heart attacks in the United States. The two main methods of diagnosis are the electrocardiogram (ECG) and the coronary angiogram. While the ECG is non-intrusive and widely available, it has a low true positive (specificity) and true negative (sensitivity), so patients must often undergo more procedures. The coronary angiogram is the gold standard for diagnosing the disease but requires X-ray dyes to be injected into the arteries. This makes the angiogram intrusive and inconvenient.

Objective

This project proposes the use of a machine learning algorithm to classify patients with CAD non-intrusively but more accurately than the ECG. The output is logistic where a value of 1 indicates greater than 50% vessel narrowing and 0 indicates less than 50% vessel narrowing.

Methods

The chi-square test of independence was applied to determine the four attributes most correlated to CAD — age, gender, heart rate, and ECG ST segment slope. The resulting matrix contained data for 1941 patients and was compressed with principal component analysis (PCA). A support vector machine (SVM) algorithm was then applied to model the data because of its computational advantages for logistic output. The generated model was implemented in a real-time data processing system.

Results

The algorithm performed with a specificity of 91.8% and sensitivity of 92.3%, compared to 51.5% and 66.1% for the ECG. The k-fold cross validation score for ten partitions was 80.5%, which proved that the model was not overfit. The current most accurate machine learning model for the same database obtained a 78.9% accuracy with a conceptual clustering algorithm.

Discussion/Conclusion

Overall, the SVM was more accurate than the current non-intrusive ECG method. Additionally, since ECG measurements can be obtained with a smartphone attachment and the other attributes can be measured by a smartphone, the diagnosis does not require a medical professional. Patients can directly use this method as a more convenient and accurate diagnostic method. In the future, the process of computing ST segment slope from raw ECG data can be further improved in the real-time data processing system.