**A STUDY ON IDENTIFICATION OF FACTORS AFFECTING FOLLOW UP PARTICIPATION BASED ON RISK CATEGORY OF THE PERSON**

Cardiovascular disease (CVD) is globally considered as the leading cause of death with 80% of CVD related deaths being reported from low and middle income countries like India. According to a report of World Health Organization (WHO) in 2005, cardiovascular mortality in India will reach with a projected escalation to 57.2 million in 2025.With current pandemic around health has taken a center stage all over again and this makes the detection of heart risk on time even for important than ever before. Conventional cardiovascular risk is attributed to lifestyle changes and altered metabolic activity. This forms the basis of a 10-year risk prediction score inspired by the Framingham study. Many people think CVD is an inevitable part of life as cancer or any other chronic diseases but cause of CVD is mainly attributed to modifiable risk factors that can be controlled easily with some lifestyle changes.

• Obesity

• Unhealthy diet

• Lack of physical activity

• High BP

• Diabetes

• Smoking

• Cholesterol

 An alternative approach might be to use non-invasive, routinely available data to stratify the population according to CVD risk and, then, only invite those at highest risk for further assessment.

The major concern in the healthcare sector for a country like India is affordability and accessibility. Also, lack of awareness among people and poor infrastructure makes the detection of the disease at preliminary stages very difficult.  Machine learning models can provide a knowledge rich environment to significantly improve the clinical decision. Such information, if provided in advance, can give important insights to doctors who can then adapt their diagnosis and treat patients accordingly. This can save up a lot of effort and cost/resources both for the population and the healthcare system.

Through this study we aim to accomplish the following:

* Understanding the factors influencing the CVD Risk
* Segmentation of persons from screening data and follow up data in various categories.
* Using Machine learning classification model on the profile output obtained from segmentation to eventually categorize the persons as Healthy, moderately healthy, Risky and High-Risk based on Probability scores.
* Correlating the labeled data thus obtained with the follow up categories and understanding the parameters/factors that influences people turning up for follow up in future.
* Eventually we try to find the probability of a person to turn up for future follow ups, given the risk category of a person