3\_background\_and\_literature\_review

Before the advent of ubiquitous application of machine learning and other modern data science methods, it had been long practised according to the belief that the risk of cardiovascular diseases is based on linear relationship with countable factors[Lloyd-Jones 2006], which was based on the limitations of data collection and prediction tools and has been already proved to be biased[Weng2017].

Some attempts [Wang T.j 2006] use biomarkers of large cohort that are difficult to collect and the interpretation of which is restricted to certain professionals. This kind of methods has been expected to be at least partly replaced by more simple and easier prediction models like those that are based on more understandable and available attributes such like age, blood pressure and alcoholism.[Muthuvel 2018]

Muthuvel etc [ ] summarised the researches of heart disease prediction using Machine Learning and other data analytics approaches. It is seen that in recent 5 years the main paradigm of research of this problem has been shifted to common attributes-based as mentioned beforehand.

The techniques in use are common ones like Multiple Linear Regression and Logistic Regression, Decision Tree, Naïve Bayes, Support Vector, Artificial Neural Network (ANN). The combination of at least techniques enhances the accuracy from some 60% (the case of J48, a decision tree method) to more than 80%, for example [Jaymin Patel 2016]. However the sensitivity (recall) is generally much higher than the precision (positive predictive value). This is typical in Weng[] where sensitivity is near 70% after improvement of techniques but precision is still lower than 20%. It is worth being noticed that the attributes used in recent years’ researches usually include medical diagnostic attributes such like electrocardiogram, serum and lipid contents, heart beats which are available after some instrumental diagnosis and also key factors in early stages of diagnosing heart diseases. These more professional attributes indicate the common practice of leveraging the prediction of risks based on “better safe than sorry” principle and this results in high false negative rate in machine learning.

Lloyd-Jones, D.M., Leip, E.P., Larson, M.G., d’Agostino, R.B., Beiser, A., Wilson, P.W., Wolf, P.A. and Levy, D., 2006. Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. Circulation, 113(6), pp.791-798.

Wang, T.J., Gona, P., Larson, M.G., Tofler, G.H., Levy, D., Newton-Cheh, C., Jacques, P.F., Rifai, N., Selhub, J., Robins, S.J. and Benjamin, E.J., 2006. Multiple biomarkers for the prediction of first major cardiovascular events and death. New England Journal of Medicine, 355(25), pp.2631-2639.

Weng, S.F., Reps, J., Kai, J., Garibaldi, J.M. and Qureshi, N., 2017. Can machine-learning improve cardiovascular risk prediction using routine clinical data?. PloS one, 12(4), p.e0174944.

Muthuvel, Marimuthu & Abinaya, M & Hariesh, K & Madhankumar, K & Pavithra, V. (2018). A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach. International Journal of Computer Applications. 181. 975-8887. 10.5120/ijca2018917863.

Bostjan Kaluza Machine Learning in Java 2015

in Weng et al , only systolic pressure is used to evaluate the situations of blood pressure.

Jaymin Patel, Prof. Tejal Upadhyay, Dr.Samir Patel,“Heart Disease Prediction using Machine Learning and Data Mining Technique”, International Journal of Computer Science and Communication, September 2015-March 2016, pp.129-137.