Risk Prediction for Cardiovascular Disease using the Framingham Heart Study Dataset

Akshat Gupta [1,4] , Swati Sharma [2,4] , Rajorshi Mondal [3,4]
[1] akshat21515@iiitd.ac.in, [2] swati21568@iiitd.ac.in, [3] rajorshi21187@iiitd.ac.in
[4] Indraprastha Institute of Information Technology, Delhi

1. Motivation

Cardiovascular Disease (CVD) remains one of the leading causes of mortality worldwide. One of the most important aspects of managing CVD is identifying people with high associated risk, and diverting more attention towards this population, which necessitates the need for a precise and effective risk scoring system. While existing risk systems exist, there remains significant room for enhancing their performance.

This inspiration prompted us to contemplate and potentially develop a novel system capable of forecasting individual patient risks related to CVD. Such an advancement would facilitate the identification of individuals at heightened risk, necessitating tailored interventions and precise treatment strategies.

2. Related Work

2.1. DeepHit [1]

This paper describes a deep learning-based approach to predict the survival times associated with a set of patients, with a consideration of *competing risks*, i.e. settings in which there is more than one possible event of interest. It also describes the relationship between patient covariates and the associated risk.

2.2. Machine Learning Techniques for Heart Disease Prediction [2]

This article describes the use of Machine Learning (ML) techniques for predicting the possibility of CVD (classification problem). It analyzes data visually to establish a set of characteristics closely related to Heart Disease and builds ML models around these characteristics.

2.3. Cardiovascular Disease Risk Prediction using automated Machine Learning [3]

This paper describes the use of an automated ML technique *AutoPrognosis* for the prediction of risk associated with CVD. It also considers the use of non-traditional covariates to check for increased accuracy of the risk prediction system.

3. Timeline

Following is a description of the expected timeline:

Task	Estimated Duration
Data Retrieval and Analysis	End of August
Statistical Analysis of Features and Feature Selection; Data Preprocessing based on results from Statistical Analysis	September
Testing traditional/conventional models on the pre-processed data and tuning for best possible results	September
Testing unconventional and/or ensemble learning methods on the data and tuning for best possible results	Mid September - Mid October
Determining the models with best results and defining the corresponding best split-set and hyper-parameters	Mid October - End of October
Documenting the results as a research paper with fully described results and corresponding methodologies	End of October

4. Individual Tasks

Each member of the group will be contributing in their best possible capacities to each of the aforementioned tasks, and the expected contribution of each member will be equal for all the tasks.

5. Final Outcome

We expect that the project will result in a robust risk scoring system capable of precisely predicting individual patients' risk scores. Ideally, the system should also be able to perform as good as, if not better than, traditional scoring systems in terms of its performance metrics that include but are not limited to accuracy, precision, recall, etc.

References

[1] Changhee Lee, William Zame, Jinsung Yoon, and Mihaela Van Der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, number 1, 2018.

[2] Yang Li and Zhaoxun Li. Heart disease prediction based on machine learning methods. In *ICBDT '22: Proceedings of the 5th International Conference on Big Data Technologies*, number 2, 2022.

[3] Ryan Poplin, Avinash V Varadarajan, Katy Blumer, Yun Liu, Michael V McConnell, Greg S Corrado, Lily Peng, and Dale R Webster. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature biomedical engineering*, (3), 2018.