Cardiovascular Disease prediction using Machine Learning Models

Arnav Gupta	Karan Gupta	Shivesh Gulati	Vishal Singh
2021236	2021258	2021286	2021575

1. Motivation

A timely and correct diagnosis of cardiovascular diseases helps provide a better prognosis and improve the quality of patient care. A diagnosis often involves multiple parameters some of which, the doctors might neglect or whose effects are not yet explored. Machine learning applications can identify complex patterns in the data and give much better, timely and accurate diagnoses.The idea for this topic came up due to the large number of cardiovascular-related mortalities during the COVID pandemic which were often caused due to negligence of subtle signs and symptoms. This prompted us to come up with a machine-learning model for diagnosing cardiovascular diseases.

2. Related Works

[1] Effective Heart disease prediction using machine leaning techniques This study develops a model to predict cardiovascular diseases and reduce related fatalities. Different models like random forest (RF), decision tree (DT), multilayer perceptron (MP), and XGBoost (XGB) were employed. Parameters were optimized using GridSearchCV. The research concludes that cross-validated multilayer perceptron outperformed other algorithms, achieving 87.28% accuracy. [1]

[2] Cardio-Vascular Disease (CVD) and Metabolic Associated Fatty Liver Disease This study aims to establish a relationship between (MAFLD) . To identify the MAFLD patients with the highest risk of CVD, this paper uses techniques such as multiple logistic regression and PCA on the dataset.The paper concluded that people showing various symptoms of MAFLD had an increased chance of CVD. MAFLD is closely linked with symptoms like high lipid profile due to metabolic dysfunction , which is mainly caused due to high cholesterol.[2]

[3] Blood pres-sure variables and cardiovascular risk: new findings from advance. This research paper is aimed at finding out the importance of Blood Pressure indices such as Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Pulse Pressure (PP) (defined as (mean(SBP)-mean(DBP)) and Mean Arterial Pressure (MAP) defined as (DBP$+\frac{1}{3}$(PP)) in predicting the risk of cardiovascular diseases in a patient, using well-established ML Models such as Cox proportional hazard regression models.[3]

3. Timeline

Week 1-2 : Data Cleaning
Week 3 : Pre-processing and Data Visualization.
Week 4 : Feature Extraction and Analysis.
Week 5 : Feature Analysis, Selection, Correlation, Heat-Maps.
Week 6 : Logistic Regression, Support Vector Machines.
Week 7 : Decision Trees, Random Forest.

Week 8 : K- Nearest Neighbours, K- Shortest Path.
Week 9 : Analysis and performance of models.
Week 10 : Hyper-parameter Tuning, Check for model for Overfitting and Under-fitting.
Week 11 : Report Generation.
Week 12 : Buffer.

4. Individual Tasks

Week	Tasks	Team Members
1-2	Data Cleaning	Vishal, Karan
3	Pre-processing and Data Visualization	Arnav, Shivesh
4	Feature Extraction	Vishal, Arnav
5	Analysis of Features (Selection, correlation, etc.)	All Members
6	Logistic Regression and SVMs	Vishal, Shivesh
7	Decision Trees and Random Forest	Arnav, Vishal, Karan
8	K-Nearest Neighbours and K-Shortest Path	Arnav, Karan
9	Performance Monitoring	All members
10	Hyperparameter Tuning	All Members
11	Report Writing	All Members
12	Buffer	All members

5. Final Outcome

The model is expected to achieve an accuracy of atleast 80% on the test dataset. This project aims to create a Machine Learning model with a high accuracy of predicting whether a person has cardiovascular disease or not. The project would use classical machine learning algorithms to find a pattern in the data of patients , thereby developing a model whose data could be easily visualized and improved upon to get much better predictions.

References

[1] Chintan M Bhatt, Parth Patel, Tarang Ghetia, and Pier Luigi Mazzeo. Effective heart disease prediction using machine learning techniques. *Algorithms*, 16(2):88, 2023.

[2] Karolina Drożdż, Katarzyna Nabrdalik, Hanna Kwiendacz, Mirela Hendel, Anna Olejarz, Andrzej Tomasik, Wojciech Bartman, Jakub Nalepa, Janusz Gumprecht, and Gregory YH Lip. Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: a machine learning approach. *Cardiovascular Diabetology*, 21(1):240, 2022.

[3] Andre-Pascal Kengne, Sébastien Czernichow, Rachel Huxley, Diederick Grobbee, Mark Woodward, Bruce Neal, Sophia Zoungas, Mark Cooper, Paul Glasziou, Pavel Hamet, et al. Blood pressure variables and cardiovascular risk: new findings from advance. *Hypertension*, 54(2):399–404, 2009.