

Exploring the Primary Risk Factors Driving Coronary Artery Disease

Data science project Proposal



Names:

Nir Binyamin – 209367184

Ilay Eldar - 206564437

Introduction

Machine learning

Machine Learning (ML), a subset of artificial intelligence (AI), allows computers to learn from data and make decisions. Unlike traditional programming, ML algorithms learn patterns from data instead of relying on explicit logic and rules. With more data, these algorithms improve their accuracy over time. ML involves feeding algorithms large datasets so they can adjust and enhance their performance without explicit programming for specific tasks. This learning process can be supervised, unsupervised, or semi-supervised, depending on the data and desired outcomes.

Significance of Machine Learning in Healthcare

The significance of ML in healthcare is profound. ML algorithms can process, analyze, and interpret vast amounts of data far more quickly and accurately than human capabilities allow. This capability is crucial in healthcare, where timely and accurate decisions can significantly impact patient outcomes. ML contributes to healthcare by:

Disease Prediction and Diagnosis:

ML models predict diseases accurately using historical patient data and interpret medical images for diagnosis, aiding in identifying conditions like tumors or fractures.

Treatment Personalization:

ML enables personalized medicine by analyzing patient data and disease patterns, tailoring treatment plans based on individual genetic makeup, lifestyle, and health history.

Operational Efficiency: Hospitals and clinics use ML to optimize scheduling, manage patient flow, and predict admission rates, improving service delivery and reducing wait times.

Coronary Artery Disease (CAD)

CAD, also known as coronary heart disease, involves the narrowing or blockage of the coronary arteries, the blood vessels that supply blood and oxygen to the heart muscle itself. This condition is primarily caused by atherosclerosis, a process in which plaques (composed of fat, cholesterol, and other substances) build up on the artery walls, reducing blood flow to the heart muscle. The reduced blood flow can lead to chest pain (angina), shortness of breath, or other symptoms, and it significantly increases the risk of heart attacks and heart failure.

Connection Between CAD and Ischemic heart diseases

Ischemic heart disease refers to heart weakening caused by reduced blood flow to your heart. Typically, this reduced blood flow is the result of coronary artery disease, a condition that occurs when your coronary arteries narrow. Ischemic heart disease may also be called cardiac ischemia or ischemic cardiomyopathy.

As your heart weakens, it must work harder to send blood to the rest of your body. This can increase your risk of blood clots, heart valve disease, heart failure, abnormal heart rhythms (arrhythmia) and other problems.

Summary

Our project aims to identify key risk factors for Coronary Artery Disease (CAD) using machine learning models and the UK Biobank dataset. By analyzing a comprehensive set of patient data, we seek to uncover novel insights into the complex interplay of factors contributing to CAD development. Our methodology involves dataset exploration, feature selection, data cleaning, model selection, hyperparameter tuning, and result analysis. Through this process, we aim to develop a predictive tool with high accuracy in CAD risk assessment. The project's outcomes will contribute to advancing personalized medicine and public health strategies, ultimately improving CAD prevention and treatment approaches.

What are the symptoms of ischemic heart disease?

Like other types of coronary artery disease, you may not have any symptoms at first. For some people, the first sign of ischemic heart disease is a heart attack.

As ischemic heart disease worsens, you may experience:

- Chest pain
- Dizziness or fainting
- Shortness of breath.
- Swelling in your feet or ankles.

Project stages:



Solution Description

The proposed solution aims to leverage machine learning (ML) techniques to identify and analyze the most influential risk factors for coronary artery disease (CAD). This project will utilize a comprehensive dataset comprising patient demographics, medical history and lifestyle factors to train and validate predictive models.

Finding the Most Influential Risk Factors for CAD

Objective: To utilize ML algorithms to sift through complex datasets and identify the key factors that significantly contribute to the risk of CAD. This includes analyzing a wide range of variables.

Approach: Implement various feature selection and importance techniques inherent in ML models to rank and highlight the variables most strongly associated with CAD risk.

Sub goal – to determine what can reduce the chances of CAD.

Project Timeline

24.2	Proposal submission
9.6.24	Submission of the project document
13-27.6.2024	Presentation to an external examiner
8/31/24	Submitting the documents on the project's website in preparation for the exhibition
10.9.24	"Looking to the future" - exhibition

Tools

UK Biobank

The UK Biobank is a large-scale biomedical database and research resource containing in-depth genetic and health information from half a million UK participants. Its extensive data repository includes genomic data, medical records, and various physical and cognitive measurements. This resource is invaluable for researchers aiming to understand the determinants of diseases like coronary artery disease. Utilizing the UK Biobank, we can access a wealth of data to identify potential risk factors and explore the interactions between lifestyle choices and CAD outcomes.

Python and Machine Learning Libraries

Python is a versatile programming language favored for its readability, and robust ecosystem, including an extensive range of libraries for data analysis, machine learning, and scientific computing. In the context of this project, Python serves as the foundational tool for data processing, statistical analysis, and the development of predictive models. Key libraries such as NumPy and pandas facilitate efficient data manipulation and analysis, while libraries like scikit-learn, TensorFlow, and PyTorch offer advanced machine learning algorithms for classification, regression, and neural network models.

GitHub

GitHub enhances our CAD risk factor project by streamlining collaboration, version control, and documentation, making your research more efficient and accessible.

[Link to the project repository](#)

Bibliography

1. [Coronary Artery Disease \(Atherosclerosis\)](#)
2. [Coronary artery disease program](#)
3. [Ischemic heart disease](#)
4. [Coronary Artery Disease \(CAD\)](#)
5. [UkBioBank showcase](#)