Prediction of Cardiac Disease using Machine Learning

PROJECT SYNOPSIS

OF MAJOR PROJECT

BACHELOR OF TECHNOLOGY

CSE

SUBMITTED BY: PCSE25-33

August 2023

KIET Group of Institutions, Delhi-NCR,

Ghaziabad (UP)

Department of Computer Science and Engineering

KIET Group of Institutions, Delhi-NCR,
Ghaziabad (UP)
Department of Computer Science and Engineering

Name of Student	Branch	University Roll no.	Class Roll No.
• **Manav Rohilla**	**CSE**	**2100290100090**	**25**
• **Parkhi Gupta**	**CSE**	**2100290100112**	**46**
• **Sukriti Rai**	**CSE**	**2100290100169**	**36**

Table of Contents:

Introduction

In the times of medical science and technological advancements, the ability to predict cardiac diseases holds immense significance. The project " Prediction of Cardiac Disease using Machine Learning " aims to use the power of modern technology and data analytics to develop a predictive model that assists in identifying individuals at risk of cardiac diseases. This project resides at the intersection of medical diagnostics, data science, and machine learning, promising substantial contributions to healthcare.

<u>Technology Utilized</u>:

The project leverages cutting-edge technologies in the fields of both medicine and data science. It amalgamates medical expertise with computational prowess to yield accurate and reliable predictions. Core technologies include:

1. <u>Data Collection</u>: A comprehensive dataset comprising various cardiac health parameters will be collected. These parameters could encompass medical history, lifestyle choices, biometric measurements, and more.
2. <u>Machine Learning Algorithms</u>: Advanced machine learning algorithms will be employed to analyze the dataset and identify patterns, correlations, and trends that are indicative of potential cardiac diseases.
3. <u>Feature Engineering</u>: Expert medical knowledge will guide the selection and engineering of relevant features from the dataset, enhancing the model's accuracy and interpretability.
4. <u>Data Visualization</u>: Visualization tools will be employed to represent complex data relationships in an understandable manner. This will aid medical professionals and patients in comprehending the risk factors more intuitively.

Specialized Field and Technical Terms:

The project is situated at the convergence of medical science and data science. It requires a deep understanding of cardiovascular health, including the various risk factors associated with cardiac diseases. Some specialized technical terms that might be encountered include:

1. <u>Atherosclerosis</u>: The buildup of plaque inside arteries, leading to restricted blood flow and potential heart problems.
2. <u>Myocardial Infarction</u>: Commonly known as a heart attack, it occurs when there is a blockage in the blood supply to the heart muscle.
3. <u>Hypertension</u>: Also known as high blood pressure, it is a condition where the force of blood against the artery walls is consistently too high.

Rationale

The development of a predictive model for cardiac disease using machine learning holds paramount significance in the times of preventive healthcare. Cardiovascular diseases stand as a leading cause of global mortality, necessitating innovative approaches to address this burgeoning public health concern. The project's need arises from several compelling factors.

Firstly, early detection and intervention are pivotal in reducing the mortality associated with cardiac diseases. By harnessing the power of modern technology and data analysis, the project aims to identify subtle patterns and correlations that might not be evident through conventional diagnostic methods. This early identification could empower medical practitioners to administer timely interventions, preventive measures, and lifestyle adjustments, significantly mitigating the impact of potential heart-related issues.

Secondly, the predictive model aligns with the paradigm shift towards personalized medicine. Each individual's cardiac health is influenced by a complex interplay of genetic predisposition, lifestyle choices, and environmental factors. A tailored predictive model can incorporate these multifaceted variables to offer personalized risk assessments, enabling healthcare professionals to provide patients with targeted advice and strategies for optimal heart health management.

Literature Review Report

Report-1

Image CHD: A 3D Computed Tomography Image Dataset for Classification of Congenital Heart Disease

Author: Xiaowei Xu, Tianchen Wang, Jian Zhuang, Haiyun Yuan,

Published on: 26th January 2021

Published by: Cornell University and Springer Link

Methodology: Image CHD contains 110 3D computed Tomography (CT) images covering most type of CHD.

It uses Baseline Method. The baseline method employs either central space discretization combined with artificial viscosity or upwind discretization. Integration in time is done using explicit multistage time-stepping schemes. For steady calculations convergence is accelerated by implicit residual smoothing, local time stepping and multigrid. Due to the lack of baseline method for CHD classification, along with the dataset we establish one as shown in Fig. 3, which modifies and extends the whole heart and great artery segmentation method in CHD [18]. It includes two subtasks:

Segmentation based connection analysis and Similarity based shape analysis.

Approach: With the extracted connection and shape features, the classification can be finally determined using a rule-based automatic approach.

Specifically, ASD and VSD have unexpected connection between LA and RA,

and LV and RV, respectively.

Key words: Congenital Heart Disease, Automatic Diagnosis, Computed Tomography.

Limitations: Test images are classified as uncertain due to segmentation error. For example - The test image given in this paper has very low contrast, and its blood pool and boundary are not clear compared with other areas, resulting in segmentation error: compared with the ground truth is only RA and part of the initial parts of great arteries are segmented.

Report-2
Deep Learning for Cardiac Segmentation

Author: Chen Chen

Publication on: 5th March 2020

Published by: Frontiers in Cardiovascular Medicine

Methodology: Deep learning has become the most widely used approach for cardiac image segmentation in recent years We include some fundamentals of deep learning that as neural network CNNs, FCNs, GAN etc. It also involves the methodology of cardiac MR image segmentation. Cardiac MRI is a non-invasive imaging technique that can visualize the structures within and around the hearts. Compared to CT, it does not require ionizing radiations.

Approach: It also involves the advanced building blocks for improve segmentation. Medical image segmentation as an important step for quantitative analysis and clinical research it requires high pixel images that's why over the past years many researchers have Developed advance building blocks to learn robust representative features for precise segmentation. These techniques have inevitably applied to stage of the art neural network to improve cardiac image segmentation.

Key terms: Artificial intelligence, deep learning neural networks cardiac image segmentation, cardiac image analysis CT, ultrasound.

Report 3

Automated coronary artery atherosclerosis detection and weakly supervised localization on coronary CT angiography with a deep 3-dimensional convolutional neural network

Author: Sema Candemir

Published on: 30th March 2020

Published by: Elsevier

Methodology: We propose a 3D CNN based supervised system which: (1) processes coronary artery volumes extracted from CCTA examinations (2) characterizes the pathological structures and lesion and (3) automatically locates the regions for providing visual clues for atherosclerosis. The proposed system has four stages: (i) extracting coronary arteries and branches from CCTA image datasets; (ii) pre-processing coronary artery volumes; (iii) classifying the vessels with the 3D CNN; and (iv) localizing the discriminative region for abnormal cases

Approach: The study involves a multi-stage approach for cardiovascular image analysis. In Stage 1, the focus is on extracting coronary arteries from Coronary Computed Tomography Angiography (CCTA) scans, a crucial step in isolating the relevant structures. Stage 2 emphasizes pre-processing the Multi-Planar Reconstruction (MPR) volumes to prepare them for training, which likely includes tasks such as image enhancement, noise reduction, and data augmentation. In Stage 3, a 3D Convolutional Neural Network (CNN) model is configured, indicating the use of deep learning techniques for feature extraction and analysis of the processed volumes. Finally, in Stage 4, the study employs a weakly supervised approach for abnormality localization, implying that the model is trained to identify anomalies or diseases within the cardiovascular images without access to precise pixel-level annotations, which is a common challenge in medical image analysis. This multi-stage methodology likely enhances the accuracy and efficiency of cardiovascular image analysis, contributing to advancements in this critical field.

Keywords: Coronary artery computed tomography, Angiography, Coronary artery disease, Stenosis classification, 3D convolutional neural networks, Weakly supervised localization

Report 4

Machine learning cardiac-MRI features predict mortality in newly diagnosed pulmonary arterial hypertension

Author: Samer Alabed

Published on: 2nd May, 2022

Published by: European Heart Journal

Methodology and Approach: This research study employs a systematic methodology encompassing several key components. Firstly, a well-defined study population is chosen to target specific medical conditions or demographics, ensuring the data's relevance to the research goals. Standardized Magnetic Resonance (MR) imaging protocols are rigorously followed to acquire consistent and high-quality images from the selected population. Subsequently, the acquired MR images undergo rigorous preprocessing, including noise reduction, motion correction, registration, and intensity normalization, aimed at enhancing data quality. The core of the analysis lies in the Multilinear Principal Component Analysis (MPCA) pipeline, which efficiently extracts complex multi-dimensional patterns from the high-dimensional imaging data. Visualizing the tensor features using techniques like t-SNE or PCA provides insights into data structures, facilitating exploratory analysis. Additionally, the study incorporates clinical and mortality data, allowing researchers to correlate imaging findings with patient outcomes and medical histories. Statistical analysis methods, such as hypothesis testing and regression, are then employed to establish the significance of identified patterns and their clinical implications, contributing to a comprehensive understanding of the research objectives. This comprehensive methodology ensures robust data collection, feature extraction, and analysis, making it a valuable approach for addressing complex medical research questions.

Keywords: Machine learning, Artificial Intelligence, Cardiac MRI, Prognosis, Mortality, Pulmonary hypertension

Feasibility Study

The feasibility study is the initial step in the software engineering process, evaluating the viability and potential success of the " Prediction of Cardiac Disease using Machine Learning " project.

Technical Feasibility:

The project capitalizes on well-established technologies such as machine learning, data analytics, and programming languages like Python. The availability of open-source libraries and frameworks for data manipulation and model development ensures a strong technical foundation.

Economic Feasibility:

The economic feasibility evaluates whether the project's benefits outweigh its costs. Developing the predictive model has the potential to significantly reduce healthcare costs by enabling early detection and prevention of cardiac diseases.

Operational Feasibility:

The operational feasibility assesses whether the project aligns with the organization's goals and objectives. Given the global prevalence of cardiac diseases and the growing importance of preventive healthcare, the project's objective is well-aligned with the broader healthcare sector.

Scheduling Feasibility:

The scheduling feasibility determines whether the project can be completed within a reasonable timeframe. With a clear roadmap and well-defined milestones, the project's development timeline can be effectively managed.

Need and Significance:

The need for accurate cardiac disease prediction is paramount due to the rising global burden of cardiovascular illnesses. Traditional diagnostic methods often identify diseases at advanced stages, limiting treatment options and increasing healthcare costs. The predictive model proposed in this project addresses this gap by enabling early identification of risk factors, leading to timely interventions and improved patient outcomes. As a result, the project's significance lies in its potential to revolutionize cardiac disease management, saving lives, and positively impacting public health on a global scale.

Methodology/ Planning of work

1. Data Collection:

   Gather a comprehensive dataset that includes relevant cardiac health parameters. This data can be obtained from hospitals or research institutions.

2. Feature Extraction:

   Extract meaningful features from the collected data that are indicative of cardiac health.

3. Feature Selection:

   Choose the most relevant features for your predictive model. Feature selection is crucial for model efficiency and interpretability.

4. Machine Learning Model Selection:

   Experiment with a variety of machine learning algorithms suitable for classification tasks. This may include logistic regression, decision trees etc.

5. Model Training:

   Use a portion of dataset to train the selected machine learning models. Ensure that the data is split into training and testing sets for model evaluation.

6. Model Evaluation:

   Assess the performance of each model using appropriate evaluation metrics like accuracy, precision, recall, F1-score, and ROC curves.

7. Comparison:

   Compare the performance of different machine learning models to identify which one performs the best for your specific dataset and task.

8. Validation and Testing:

   Validate your models on external datasets if available, to ensure their generalizability. Perform thorough testing to confirm that the models work reliably in real-world scenarios.

9. Interpretability and Visualization:

   Utilize techniques like SHAP values or feature importance scores to interpret and explain the predictions made by your models. Visualize key insights and findings to aid in understanding and communication.

10. Fine-Tuning and Optimization:

    Fine-tune the hyperparameters of your selected model to further improve performance. This step may involve grid search or random search.

11. Documentation and Reporting:

    Documenting methodology, findings, and results thoroughly. Preparing a comprehensive report that includes the dataset description, feature selection rationale, model details, and performance metrics

Facilities required for proposed work

The successful development of the "Prediction of Cardiac Disease using Machine Learning " project requires a combination of software tools and hardware resources to ensure accurate data analysis, model development, and testing.

Software:

1. Programming Languages: Proficiency in programming languages such as Python is essential. Python offers a rich ecosystem of libraries and frameworks for data manipulation, machine learning, and data visualization.

2. Machine Learning Libraries: Libraries like Scikit-Learn, TensorFlow, and PyTorch provide pre-built algorithms and tools for developing machine learning models.

3. Data Analysis Tools: Utilize tools like Pandas and NumPy for data manipulation, exploration, and preprocessing.

4. Data Visualization Tools: Matplotlib, Seaborn, and Plotly are valuable tools for creating insightful visualizations to interpret and present data trends.

Hardware:

1. Computing Resources: A capable computer with sufficient processing power and memory to handle data analysis, model training, and experimentation.

2. Storage: Adequate storage space to manage and store the dataset, codebase, and trained models.

3. Graphics Processing Unit (GPU): For complex machine learning tasks, having access to a GPU can significantly accelerate model training times.

4. Web Hosting: If the project involves deploying a web application, a web hosting service or cloud platform like AWS, Azure, or Heroku may be needed.

Data:

1. Dataset: A comprehensive and well-curated dataset containing relevant cardiac health parameters is essential for training and testing the predictive model.

Expected Outcome

The expected outcomes of the " Prediction of Cardiac Disease using Machine Learning " project encompass a range of significant contributions to the field of healthcare. Foremost, the project aims to deliver a robust predictive model capable of accurately identifying individuals at risk of cardiac diseases at an early stage. This will empower medical professionals to administer timely interventions, personalized treatment plans, and lifestyle recommendations, ultimately leading to improved patient outcomes. Moreover, the project's success could foster a shift towards proactive healthcare strategies, reducing the burden on healthcare systems and enhancing overall public health. By leveraging the synergy between medical expertise and technological innovation, the project aspires to be a transformative force in cardiac disease management, potentially saving lives and improving the quality of healthcare on a global scale.

References

1. https://www.frontiersin.org/articles/10.3389/fcvm.2020.00001/full
2. Automated coronary artery atherosclerosis detection and weakly supervised localization on coronary CT angiography with a deep 3-dimensional convolutional neural network - ScienceDirect
3. Machine learning cardiac-MRI features predict mortality in newly diagnosed pulmonary arterial hypertension | European Heart Journal - Digital Health | Oxford Academic (oup.com)
4. Application of artificial intelligence in cardiac CT: From basics to clinical practice - European Journal of Radiology (ejradiology.com)
5. https://www.researchgate.net/publication/348802958_ImageCHD_A_3D_Computed_Tomography_Image_Dataset_for_Classification_of_Congenital_Heart_Disease?enrichId=rgreq-211f84e2f20b68e11a2393eb21288a3f-XXX&enrichSource=Y292ZXJQYWdlOzM0ODgwMjk1ODtBUzoxMDIxNjAyODc5Nzg3MDA5QDE2MjA1ODA1MTUxNTY%3D&el=1_x_2&_esc=publicationCoverPdf
6. State-of-the-Art Deep Learning in Cardiovascular Image Analysis - ScienceDirect
7. Frontiers in Cardiovascular Medicine