ABESEC Ghaziabad

Department of Computer Science & Engineering

SYNOPSIS REPORT

(Session 2022-23)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Project Title:  Prediction of heart disease | | |  |  |
| Project Type (application, product, research, review etc.) | |  |  |  |
|  | Name | Roll Number | Section | Signature |
| Group member (1) | DHEERAJ KUMAR | 2100320100064 | A |  |
| Project Guide | MR. DEVENDRA KUMAR MISHRA | Remarks: |  |  |
| Signature |  |
| Date of submission |  |

**Contents of Synopsis Report:**

Defining what is needed is the first step toward establishing a project timeline, setting of project goals and allocating project resources. These steps will help you to define the work that needs to be done - or in other words, define the scope of the project.

# **Problem Introduction**

Heart disease, alternatively known as cardiovascular disease, encases various conditions that impact the heart and is the primary basis of death worldwide over the span of the past few decades. It associates many risk factors in heart disease and a need of the time to get accurate, reliable, and sensible approaches to make an early diagnosis to achieve prompt management of the disease. Data mining is a commonly used technique for processing enormous data in the healthcare domain. Researchers apply several data mining and machine learning techniques to analyse huge complex medical data, helping healthcare professionals to predict heart disease

**1.1.2. Project Objective**

To predict the heart disease with the previous database.

**1.1.3. Scope of the Project**

Heart disease affects millions of people, and it remains the chief cause of death in the world. Medical diagnosis should be proficient, reliable, and aided with computer techniques to reduce the effective cost for diagnostic tests. Data mining is a software technology that helps computers to build and classify various attributes. This research paper uses classification techniques to predict heart disease. This section gives a portrayal of the related subjects like machine learning and its methods with brief descriptions, data pre-processing, evaluation measurements and description of the dataset used in this research.

# **1.2. Related Previous Work**

(It briefly includes previous work carried out in the suggested field, researching the problem studied, summarization of the results obtained etc.)

**1.3 Software and Hardware requirements**

## Machine Learning

Machine learning is an emerging subdivision of artificial intelligence. Its primary focus is to design systems, allow them to learn and make predictions based on the experience. It trains machine learning algorithms using a training dataset to create a model.

### Supervised Learning

The model is trained on a dataset that is labelled. It has input data and its outcomes. Data are classified and split into training and test dataset.

### Reinforcement Learning

It does not use labelled dataset nor the results are associated with data, thus model learns from the experience. In this technique, the model improves its presentation based on its association with environment and figures out how to discuss its faults and to get the right outcome through assessment and testing various prospects.

**1.4 Proposed Method**

This research aims to foresee the odds of having heart disease as probable cause of computerized prediction of heart disease that is helpful in the medical field for clinicians paper additionally depicts which attributes contribute more than the others to anticipation of higher precision. This may spare the expense of different trials of a patient, as all the attributes may not contribute such a substantial amount to expect the outcome .

1.5 Deliverables

The overall aim is to define various data mining techniques useful in effective heart disease prediction. Efficient and accurate prediction with a lesser number of attributes and tests is our goal. In this study, I consider only 14 essential attributes. I applied four data mining classification techniques, K-nearest neighbor, Naive Bayes, decision tree, and random forest. The data were pre-processed and then used in the model. K-nearest neighbor, Naïve Bayes, and random forest are the algorithms showing the best results in this model. I found the accuracy after implementing four algorithms to be highest in K-nearest neighbors (*k* = 7). We can further expand this research incorporating other data mining techniques such as time series, clustering and association rules, support vector machine, and genetic algorithm. Considering the limitations of this study, there is a need to implement more complex and combination of models to get higher accuracy for early prediction of heart disease.

**1.8 References**

1. Seckeler MD, Hoke TR. The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease. Clin Epidemiol. 2011;3:67.
2. Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic of coronary heart disease in low-and middleincome countries. Curr Probl Cardiol. 2010;35(2):72–115.
3. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machinelearning improve cardiovascular risk prediction using routine clinical data? PLoS ONE. 2017;12(4):e0174944.
4. Ramalingam VV, Dandapath A, Raja MK. Heart disease prediction using machine learning techniques: a survey. Int J Eng Technol. 2018;7(2.8):684–7.
5. Patel J, TejalUpadhyay D, Patel S. Heart disease prediction using machine learning and data mining technique. Heart Dis. 2015;7(1):129–37.
6. Fatima M, Pasha M. Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl. 2017;9:1–16. [https](https://doi.org/10.4236/jilsa.2017.91001) [://doi.org/10.4236/jilsa .2017.91001](https://doi.org/10.4236/jilsa.2017.91001) .
7. Pahwa K, Kumar R. Prediction of heart disease using hybrid technique for selecting features. In: 2017 4th IEEE Uttar Pradesh section international conference on electrical, computer and electronics (UPCON). IEEE. p. 500–504.