## 1.INTRODUCTION

## Brief Information about the Project

Cardiovascular Diseases (CVDs) are the most common cause of death globally representing 32% of all global deaths with about 17.9 million people being affected by it. Out of these, two most common CVDs are in the form of heart attack and heart stroke accounting 85% of the total people. Heart attack is caused due to blockage of oxygen or blood supply to the heart muscle while heart stroke is caused when there is blockage of the vessel feeding the brain. Although both of the diseases are different from each other, the risk factors contributing to them are quite similar. The risk factors include unhealthy diet, tobacco use, diabetes, sedentary lifestyle, unhealthy use of alcohol, high blood pressure and family history. Detecting heart stroke and taking medical action immediately can not only prolong life but also help to prevent heart disease in the future.

Machine learning has become one of the most demanding field in modern technology. It is a form of artificial intelligence where the model can analyze the data, identify patterns and predict the outcome with minimal human intervention. Heart stroke prediction in adults can be done by using various machine learning algorithms. It has become an intrigued research problem as there are various factors or parameters that can influence the outcome. The factors include work type, gender, residence type, age, average glucose level, body mass index, smoking status of the individual and any previous heart disease.

The proposed model predicts heart stroke prediction of several individuals using various machine learning algorithms like Random Forest, K-Nearest Neighbors, Decision Tree Classifier, Support Vector Machine, Logistic Regression and Naïve Bayes based on these input factors which have been taken from the dataset on which the model has been trained.

### Motivation and Contribution of the Project

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems.

Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped. This raises an important question: "How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions”.

### Objective of the Project

### Heart disease and strokes have rapidly increased globally even at juvenile ages. Stroke prediction is a complex task requiring huge amount of data pre-processing and there is a need to automate the prediction process for the early detection of symptoms related to stroke so that it can be prevented at an early stage. In the proposed model, heart stroke prediction is performed on a dataset collected from Kaggle. The model predicts the chances a person will have stroke based on symptoms like age, gender, average glucose level, smoking status, body mass index, work type and residence type. It classifies the persons risk level by implementing various machine learning algorithms like Random Forest, Naive bayes, Logistic Regression, K-Nearest Neighbor (KNN), Decision Tree and Support Vector Machine (SVM). Thus, a comparative analysis is shown between the various algorithms and the most efficient one is obtained. Decision Tree algorithm was found out to be the most effective one with an accuracy of 100%.

### Organization of the Project

* + - **Chapter2: Literature Survey:** This chapter consists of background of the project, possible approaches, introduction and comparison of technologies.
    - **Chapter3: System Analysis:** This chapter consists of description of current system, proposed system, algorithms and requirement specifications.
    - **Chapter4: System Design:** This chapter mainly consists of modules description and unified modeling language diagrams: use case diagrams, class diagrams, sequence diagrams, collaboration diagrams, and activity diagrams.
    - **Chapter 5: Technology Description:** This chapter mainly consists of technology used in this project.
    - **Chapter 6: Sample code:** This chapter mainly consists of sample code for few modules.
    - **Chapter 7: Testing:** This chapter mainly consists of testing techniques and test cases for modules.
    - **Chapter 8: Screenshots:** This chapter mainly consists of output screens of this project.

**Conclusion:** Main conclusion of this project.

# CHAPTER-2

# LITERATURE SURVEY

## LITERATURE SURVEY

In [1], heart disease prediction is done using Naive Bayes and Genetic algorithms. The model has been trained on a UCI dataset with attributes like gender, age, resting blood pressure, cholesterol, fasting blood sugar, old peak, etc. It is a web-based machine learning application where the user inputs his medical details based on these attributes to predict his heart disease. The algorithm calculates the probability of having a heart disease and the result is displayed on the web page itself.

In [2], various classification algorithms are studied and the most accurate model is obtained for predicting the heart disease in the patient. It was found that Random Forest and XGBoost were the most efficient algorithms while K- Nearest Neighbor was found to be the most ineffective one.

In [3], a novel heart attack prediction mechanism is proposed mainly using Decision Tree Classifier algorithm. The model first learns the deep features based on the attributes provided in the dataset and then trains on the learned features to obtain the outcome or prediction.

In [4], a survey is proposed on the various machine learning algorithms that could be used for the heart disease prediction. The authors have summarized the various algorithms and then worked towards finding the best algorithm by analysing the various features.

In [5], heart disease prediction has been performed using the four algorithms- Logistic Regression, Naive Bayes, Random Forest and Decision Tree. The objective is to effectively study whether the patient has any heart disease. The health professional enters the input values from the patients health report. The data is then fed into the machine learning model which provides the probability of having the heart disease.

In [6], heart stroke prediction is analysed using various machine learning algorithms and the Receiver Operating Curve (ROC) is obtained for each algorithm. It has been implemented using Apache Spark and it shows that the Gradient Boosting Algorithm gives the highest ROC score of 0.90. The analysis of the features has been done by using univariate and multivariate plots to obtain the correlation between the several features.