**iii Background Study**

According to the World Health Organization, 15 million people suffer stroke worldwide each year. Of these, 5 million dies and another 5 million are permanently disabled. High blood pressure contributes to more than 12.7 million strokes worldwide. Europe averages approximately 650,000 stroke deaths each year. In developed countries, the incidence of stroke is declining, largely due to efforts to lower blood pressure and reduce smoking. However, the overall rate of stroke remains high due to the aging of the population. [2]

1. **Data mining**

Data mining is a process of collecting or gathering hidden data from large data sets where data set means a collection of data. The data mining process is used for finding a pattern or similarity and relationship among large data set. We are living in an age where data is constantly being collected. Our required and unnecessary data is being submitted too. So we need to get help with data mining if we need to find our required data. In the data mining process, Flat Files, Rational Databases, Datawarehouse, Transactional Databases, Multimedia Databases, Spatial Databases, Time Series Databases and World Wide Web (WWW) are different sources are used. Classifications, regression, clustering analysis, associations are different functionalities of data mining. Data mining techniques have great success in search engine, health care, education and business intelligence etc. It is also a very helpful technique for the researcher. So we wanted to use it in the healthcare sector to predict the risk of stroke.

Data mining techniques have great success in research and development. Data mining process can be thought of as a genuine appraisement of information technologies. There are many different issues in research which can be implemented by using data mining process and techniques. [3] In the health care sector, a huge amount of data are adding day by day, those data can be mined for different research work. So we preferred to use data mining techniques in our thesis work.

1. **Knowledge Discovery in Databases (KDD)**

As we know in medical sectors there are many data sets are being collected and we do not need all of those data sets. So we have to find out our necessary data sets among them. Knowledge Discovery Database can help us in that situation. Knowledge Discovery Database is the name of a process that can recognize our required data from the large data sets. It is a very helpful process for finding a pattern form a large amount of data.

Data mining is a part of KDD. In KDD process at first, data that we have collected through the database from the user interface will be stored in the data warehouse. In the warehouse, data will be checked or tested that, whether this data is good or not for the user or the operation. Then data will be going through many different phases namely; data mining, pattern finding, related data or pattern finding, representation of data. KDD starts its work through data cleaning, data integration, data selection.

**Data Cleaning** is a task performed in KDD for removing noise, handling missing data, finding the error, inconsistent data etc. For example data parsing in which data will be checked whether the data is acceptable or not.

**Data Integration** is a task used for integrating or collecting together all of those data which are coming from different sources for storing that's data set in the data warehouse.

**Data Selection -** KDD will retrieve all of the data which are related to our required data sets from the database for our analyzation.

After this KDD complete the rest four steps. They are transformation, mining, interpretation or evaluation and knowledge.

**Transformation** is a step of KDD where all data which are cleaned, integrated and selected will be being consolidated or transformed from one format to another format. With the help of this step, the result of data mining in KDD will be more accurate and data will be more easier to understand. For transforming data into a form KDD follows some steps those are smoothing, attribute constructing, aggregation, normalization, discretization, hierarchy generation for nominal data etc. It is a very important step.

**Data Mining** is the most important step in the whole process. In there data will be mined to find a pattern or similarity. After finding the pattern, it will be evaluated and interpreted in **Data interpretation or evaluation** steps for making the pattern more appropriate. Then KDD will show or represent the data sets to the user which is known as knowledge representation it can be as like image, graph, tree, chart, text etc. If the represented knowledge is matched with the user required data then KDD will store it as **knowledge**.

1. **Classification**