

➤ **Cover:**

Assalamu Alaikum. I'm Tanjina Rahman. I stand here to present my thesis work : Heart Disease Prediction at Early Stage using Data Mining Technique.

I specially thank my supervisor Dr. AHM Sarowar Sattar, Professor, Department of CSE, RUET, for his help and support.

➤ **00:** I stand here to talk about our work, motivation, methodology and result evaluation of our work.

➤ **01:**

At first, I should tell about Heart and then comes its Diseases and types. Now, what is the heart? We all know that the heart is one of the most important organs of our body. Which is pumping all the time and remaining us alive. What will happen if it stops working? We will die within a moment. Some conditions may affect our heart. I repeat, some conditions.

➤ **02:**

A common mistake is we think heart disease is a single condition. But no, it's a group of conditions. We even never heard of terms like AF, IRDs, etc.

➤ **03:**

In today's world, when medical science is doing outstanding works, in this time about 17.3 million people die from heart disease! Statistics say it's 19% of all deaths. What if we can predict heart disease earlier? Then people will have time to keep his heart healthy. But how? Here comes Data mining.

➤ **04:**

Data mining finds an unknown pattern or hidden information from data. Medical science has a huge dataset but those are not mined. We can find out hidden knowledge from data and predict heart disease using the data mining technique. A question may arise on your mind. Why we have to predict when there are doctors?

➤ **05:**

Well, clinical decisions are made based on the doctor's experience and intuition. Inexperienced, even experienced doctors may do mistakes. And the poor medical decision means a waste of money, a waste of time and then? Death will knock at your door.

➤ **06:**

Researchers are already working on this field. In 2008, researchers developed an Intelligent heart disease prediction model. They used Naïve Bayes, Neural Network and decision tree. And they decided that Naïve Bayes is the best. Another team found better accuracy for Naïve Bayes. Before then, in 2005, another team proved that Tree Augmented Naïve Bayes is better than Naïve Bayes. I think, for better performance we should apply more algorithms.

➤ **07:**

In order to find better performance it's important to focus on feature selection. Researchers also said about its importance and they found better performance.

➤ **08:**

Our main goal is to predict heart disease which will reduce our costs and money and will save our life. Due to do this we will perform several algorithms and will compare their performance and find the best one.

➤ **09:**

We can divide our work into 4 steps. Data collection, Preprocessing, Algorithm apply and Decision.

Data collection means collecting related datasets.

And then we will process our dataset according to our work and then apply algorithms on it and after then we will take the decision like which one gives better performance.

➤ **10:**

We have collected 5 datasets from UCI machine learning repository. 4 of them are raw dataset and those have 76 attributes. And they have missing values. One of them is a processed dataset which has 14 attributes and 270 instances.

➤ **11:**

As our most of the datasets are not processed so, we had to focus on data preprocessing. Data preprocessing has some different sectors like Data cleaning, Data Integration etc.

➤ **12:**

At first we worked on Cleveland dataset which has 303 records and 76 attributes. 21 records have garbage values. So we have removed them. Then removed some attributes and then removed missing values.

➤ **13:**

Our dataset was in .data format. We had to convert it to csv format to load in WEKA. We had to save this file in .arff format to apply filter. Then we applied ReplaceMissingValue filter and missing values were replaced by their attribute mean. And then we converted it to csv format again.

➤ **14:**

And after preprocessing our dataset has 282 instances, 38 attributes and no missing values.

➤ **15:**

In future we will apply feature selection algorithms to improve performance. We will combine our dataset. And we will apply ML algorithms.

➤ **16:**

We will continue our work in this procedure. We will apply our algorithms both on single datasets and combined datasets. And we will compare their performance.

➤ **17:**

I hope, after completing our experiment we will be able to improve performance for heart disease prediction and able to predict heart disease accurately. I hope, this work will be very effective for medical science.

➤ **I-III:**

Here are the literature references for the information I present here.

➤ Thank you, everyone.

