Multi Disease Prediction System

Shikhar Suryan, Prashant, Vaibhav Kapila, Gautam Dulgas
[1]*UG Scholar, Dept. of computer applications, MAHARAJA SURAJMAL INSITUTE, JANAKPURI DELHI-110058, GGISPU*
[21]*UG Scholar, Dept. of computer applications, MAHARAJA SURAJMAL INSITUTE, JANAKPURI DELHI-110058, GGISPU*
[3]*UG Scholar, Dept. of computer applications, MAHARAJA SURAJMAL INSITUTE, JANAKPURI DELHI-110058, GGISPU*
[4]*UG Scholar, Dept. of computer applications, MAHARAJA SURAJMAL INSITUTE, JANAKPURI DELHI-110058, GGISPU*

Multi Disease Prediction System

Abstract - Multi Malady Prediction" framework based on prescient modeling predicts the malady of the client on the premise of the side effects that client gives as an input to the framework. The framework analyzes the side effects given by the client as input and gives the likelihood of the illness as an yield Illness Forecast is done by executing the irregular woodland classifier conjointly , we actualize Jungle fever and Pneumonia utilizing Profound Learning Model(CNN) and within the prosed strategy it gives the superior precision and we plan web assignment for expectation system.

Index Terms - Irregular Woodland, Unremitting Malady, CNN, Prescient Modeling.

I.INTRODUCTION

If someone is sick, they need to go to a doctor, but this can be costly and take a while. It is hard for people if they cannot see doctors or go to hospitals when they are sick, because their illness may not be found. If we can use a computer to do the procedure quickly and cheaply, it is good for the person and makes the process easier. There are computer programs that check how dangerous heart disease is for someone by looking at their medical information. Disease Predictor is a tool that can tell you what sickness you might have using the symptoms you tell it. The Disease Prediction system got information from different websites about health. You can find out if you're likely to have a disease by telling a machine your symptoms with a Disease Predictor. People like to learn new things, especially now that more and more people are using the internet. When there is a problem, many people want to search for it on the internet. Doctors and hospitals don't have as much access to the Internet as regular people do. feel well and their body is not healthy.

You have lots of choices. This system can help people. Chronic illness means having a sickness that stays for a long time or takes a long time to get better. Most chronic illnesses cannot be fully cured; however, it is possible to manage them through daily treatment. India, just like other countries, is changing a lot. These changes are leading to many people getting heart disease quickly. A lot of countries, even India, are facing problems with long-term diseases like heart disease and diabetes. This can be bad for the whole world's health, safety, and money. Many cities are growing quickly and there are many jobs and different ways of living now. Many people in all countries are now suffering from chronic diseases. It affects about one-third of the population in each country. It is costly and hard for people with long-term illnesses to receive treatment. In medicine, lots of information about long-term illnesses is collected and analyzed by using data mining to help find diseases early. Some illnesses like heart disease, diabetes, liver problems, Alzheimer's, and Parkinson's are very expensive to treat.

It is difficult for medical and healthcare industries to give good care to all patients. Only people who can pay for it can take advantage of it. There is a lot of healthcare information that people are not using well to make good choices. We need to do a better job of finding important information in that data. We are using computers to find diseases like diabetes, cancer, and heart disease before they become too serious. Machine learning is when computers learn from examples or past data to do a better job. Machine learning is when computers are taught to learn by using data and past experiences. When we use a machine to learn things, there are two parts called

Training and Testing. Doctors use signs and past medical information to guess which disease a patient may have. It has been hard to use machines to do this job for a long time. Machine learning helps doctors to fix health problems quickly and effectively in the medical field.

II. RESEARCH OBJECTIVE

We need to create a system that helps people guess if they have a long-lasting sickness without going to the doctor. To find out different illnesses by looking at how the patients feel and using different ways of teaching computers. There is no correct way to deal with written words and organized information. The new plan will look at both organized and unorganized information. "Using Machine Learning can make predictions more accurate. "

III. LITERATURE REVIEW

The think about for the finest restorative conclusion mining method was performed by K.M. Al-Aidaroos, A.A. Bakar, and Z. Othman. For this ponder, the creators compared Nave Baeyes to five other classifiers: LR, KStar (K*), Choice Tree (DT), Neural Organize (NN), and a fundamental rule-based calculation (ZeroR). The proficiency of all calculations was assessed utilizing 15 real-world therapeutic issues from the UCI machine learning store (Asuncion and Newman, 2007). Within the try, NB beated the other calculations in 8 of the 15 information sets, driving to the conclusion that the prescient exactness comes about in Nave Baeyes are predominant to other methods. Darcy A. Davis, Nitesh

Chawla, Nicholas Blumm, Nicholas Christakis, and Albert-Laszlo Barabasi found that treating inveterate sickness at a worldwide level is not one or the other time nor fetched viable. As a result, the creators performed this ponder in arrange to estimate potential illness hazard. CARE (which employments as it were a patient's therapeutic history and ICD- 9-CM codes to foresee conceivable infection risks) was utilized for this. Based on their claim restorative history which of comparative patients, CARE consolidates collective sifting approaches with clustering to foresee each patient's most prominent malady dangers. ICARE, an iterative form that coordinating outfit standards for progressed effectiveness, has moreover been characterized by the authors.

These cutting-edge frameworks do not require any progressed information and can foresee a wide extend of therapeutic conditions in a single run. ICARE's exceptional potential hazard scope implies more exact early cautions for thousands of ailments, a few a long time ahead of time. When utilized to its full degree, the CARE framework can be utilized to examine a more extensive run of malady foundations, raise already unconsidered questions, and encourage dialogs with respect to early location and prevention.

This term paper was composed by JyotiSoni, Ujma Ansari, Dipesh Sharma, and SunitaSoni to supply a study of existing strategies of data revelation in databases utilizing information mining procedures that are utilized in today's therapeutic inquire about, particularly in Heart Illness Expectation. A number of tests have been carried out to compare the execution of prescient information mining procedures on the same dataset, and the comes about appear that Choice Tree beats, with Bayesian classification having comparable precision to Choice Tree in a few cases, but other predictive approaches such as KNN, Neural Systems, and Classification based on Clustering underperform. Shadab Adam Pattekari and Asma Parveen conducted a consider to anticipate heart infections utilizing the Choice Tree Algorithm, in which the shopper gives information that's compared to a qualified set of values. As a result of this consider, patients were able to supply essential data that was compared to information, and heart malady was anticipated. M.A.NisharaBanu and B. Gomathy examined the different sorts of heart-related issues utilizing restorative information mining methods such as affiliation run the show mining, gathering, and clustering I. The point of a choice tree is to appear any possible outcome of a decision. To realize the leading result, various rules are concocted. The criteria utilized in this think about were age, sex, smoking, being overweight, drinking liquor, blood sugar, heart rate, and blood weight. The chance level for different parameters is spared with their ids extending from 1 to 100. (1-8). The standard level of expectation is spoken to by IDs less than 1, while higher IDs other than 1 speak to higher chance levels. The design within the dataset is considered utilizing the K-means clustering method. The calculation isolates the information into k bunches. The closed cluster is designated to each point within the dataset. Each cluster middle is recalculated as the normal of the cluster's points.

IV PROPOSED SYSTEM

We have mixed structured and unstructured data in the healthcare fields to determine disease risk in this project. The use of a latent factor model to recreate missing data in medical records obtained from online sources. We could also assess the major chronic diseases in a specific area and population using statistical information. We consult hospital experts to learn about useful features when dealing with structured data. In the case of unstructured text files, we use the randrom forest algorithm to automatically select features.

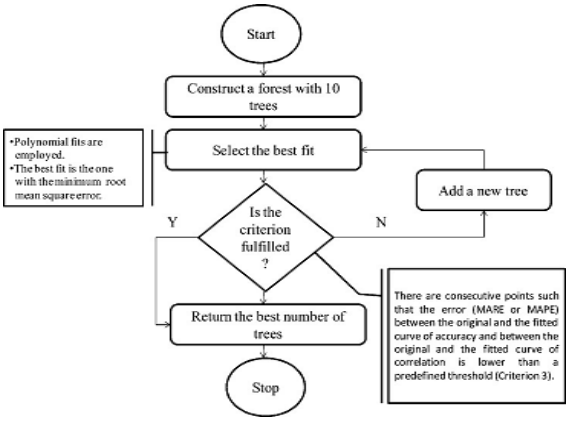

Fig 1: - System Model

4.1 Data collection

Data collection has been done from the internet to identify the disease here the real symptoms of the disease are collected i.e. no dummy values are entered. The symptoms of the disease are collected from different health related websites.

4.2 Data Preprocessing

Before feeding the data into the Prediction model, following data cleaning and preprocessing steps are performed

- Checking null values and filling using forward fill method
- Converting data into different cases
- Standardizing the data using mean and standard deviation
- Splitting the dataset into training and testing sets

4.3 Building Model

Many methods are used to perform data mining. Machine learning is one of the approaches. Random forest Machine learning strategies include grouping, clustering, summarization, and many others. Since classification techniques are used in this project, classification is one of the data mining processes in this phase of categorical data classification. And this step is divided into two phases: training and testing. In the training phase, predetermined data and associated class labels are used for classification. The training stage is often referred to as supervised learning. The preparation and testing phases of the classification process are depicted in the diagram. In the training process, training tuples are used, and in the test data phase, test data tuples are used, and the classification rule's accuracy is calculated. Assume that the classification rule's accuracy on testing data is sufficient for the rule to be used for classification of unmined data.

4.4 Prediction:

Prediction using Random Forest: -

Prediction done by Random Forest Model using Flask framework model trained by training chronic disease dataset

4.5Algorithm

4.4.1 Random Forest Algorithm

Input: Dataset

Output: Predicted class label

Step 1 : Set Number of classes = N, Number of features =M

Step 2 : Let „m" determine the number of features at a node of decision tree, (m < M)

Step 3 : For each decision tree do

Select randomly: a subset (with replacement) of training data that represents the N classes and use the rest of data to measure the error of the tree

Step 4 : For Each node of this tree do Select randomly: m features to determine the decision at this node and calculate the best split accordingly.

Step 5: End for

Step6 : End For

4.2.2 CNN(CONVOLUTIONAL NEURAL NETWORK):

A convolutional neural network is a feed-forward neural network that is generally used to analyze visual images by processing data with grid-like topology. It's also known as a ConvNet. A convolutional neural network is used to detect and classify objects in an image. A convolution neural network has multiple hidden layers that help in extracting information from an image. The four important layers in CNN are:

1. Convolution layer
2. ReLU layer
3. Pooling layer
4. Fully connected layer Convolution Layer

This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values

4.2.2.1 ReLU layer

ReLU stands for the rectified linear unit. Once the feature maps are extracted, the next step is tomove them to a ReLU layer. ReLU performs an element-wise operation and sets all the negative pixels to 0. It introduces non-linearity to the network, and the generated output is a rectified feature map.

4.2.2.2 Pooling Layer

Pooling is a down-sampling operation that reduces the dimensionality of the feature map. The rectified feature map now goes through a pooling layer to generate a pooled feature map.

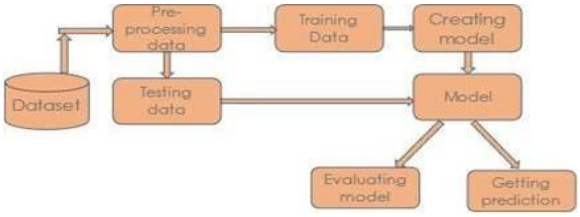

FIG 2: SYSTEM ARCHITECTURE

V.RESULTS AND CONCLUSION

Model	Accuracy
Diabetes Model	98.25%
Breast Cancer Model	98.25%
Heart Disease Model	85.25%
Kidney Disease Model	99%
Liver Disease Model	78%
Malaria model (CNN)	96%
Pneumonia model (CNN)	95%

Table 1: - shows the accuracy achieved using random forest for each disease

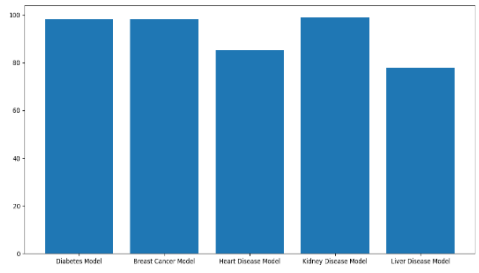

Fig. 3 shows the accuracy or each model using Random forest classifier

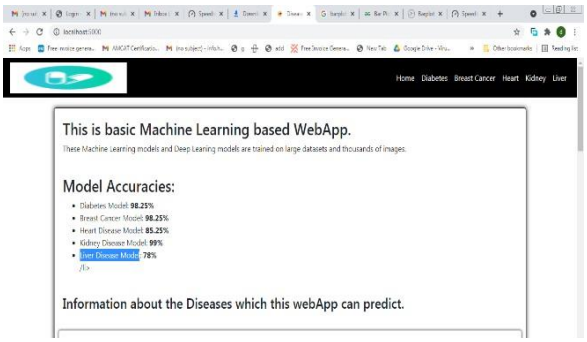

Fig. 4 Home screen of the prediction system

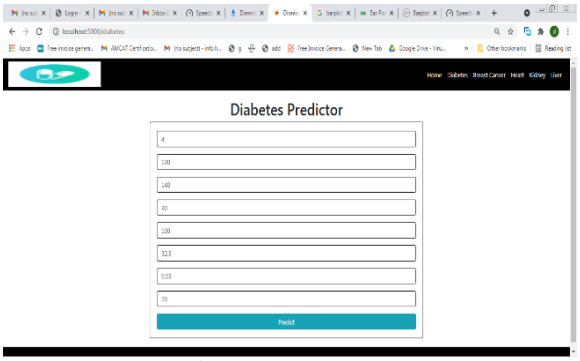

Fig. 5 :- Diabetes Prediction entry form

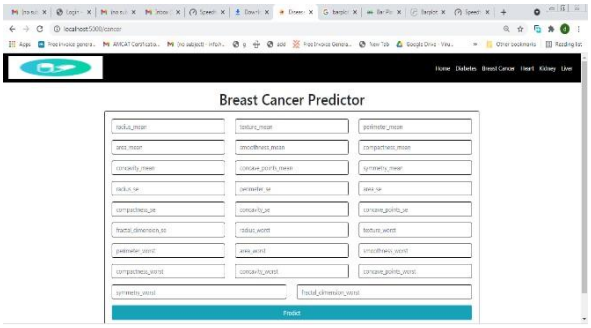

Fig. 6 :- Breast cancer Prediction entry form

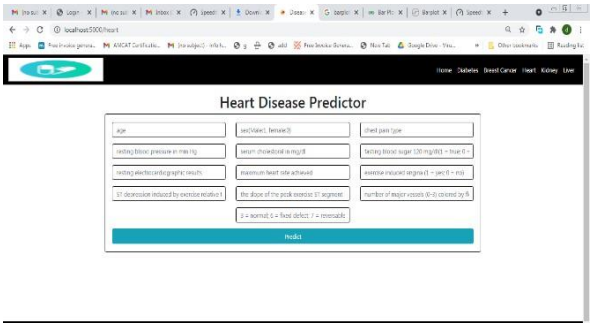

Fig. 7 :- Heart Disease Prediction entry form

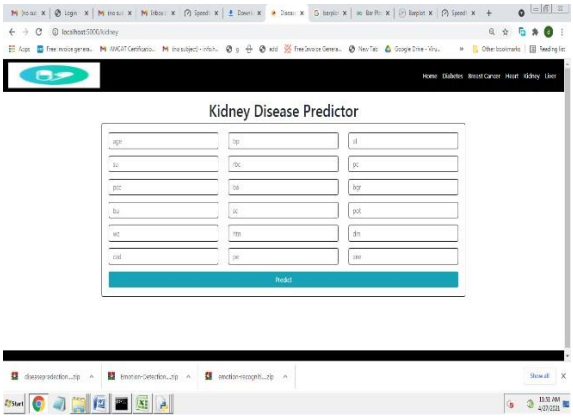

Fig. 8:- Kidney Disease Prediction entry form

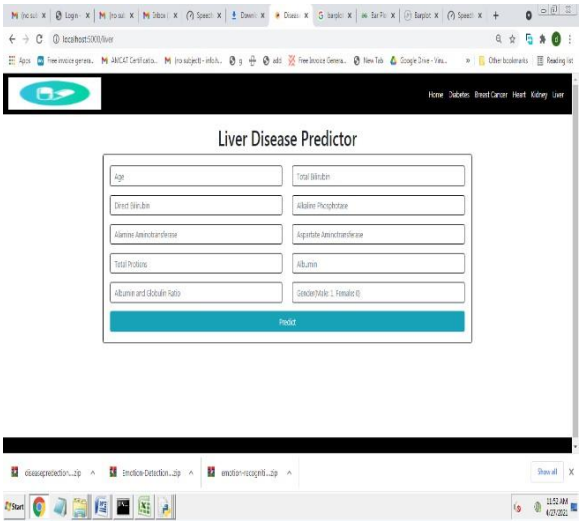

Fig. 9:- Liver Disease Prediction entry form

```
Epoch 7/50
345/345 [==============================] - 130s 377ms/step - loss:
0.2068 - accuracy: 0.9295 - val_loss: 0.2069 - val_accuracy: 0.9249
Epoch 8/50
345/345 [==============================] - 130s 376ms/step - loss:
0.1960 - accuracy: 0.9372 - val_loss: 0.1835 - val_accuracy: 0.9289
Epoch 9/50
345/345 [==============================] - 130s 378ms/step - loss:
0.1950 - accuracy: 0.9356 - val_loss: 0.1897 - val_accuracy: 0.9249
Epoch 10/50
345/345 [==============================] - 130s 378ms/step - loss:
0.1763 - accuracy: 0.9433 - val_loss: 0.1872 - val_accuracy: 0.9372
Epoch 11/50
345/345 [==============================] - 129s 375ms/step - loss:
0.1588 - accuracy: 0.9477 - val_loss: 0.1844 - val_accuracy: 0.9352
Epoch 12/50
345/345 [==============================] - 130s 376ms/step - loss:
0.1636 - accuracy: 0.9455 - val_loss: 0.1869 - val_accuracy: 0.9365
Epoch 13/50
345/345 [==============================] - 130s 377ms/step - loss:
0.1553 - accuracy: 0.9497 - val_loss: 0.2017 - val_accuracy: 0.9399
Epoch 00013: early stopping
```

Fig 10 :- this graph represents the training and validation accuracy of malaria disease where the training accuracy is 94.7 at 13 epochs and Val accuracy is 93.99

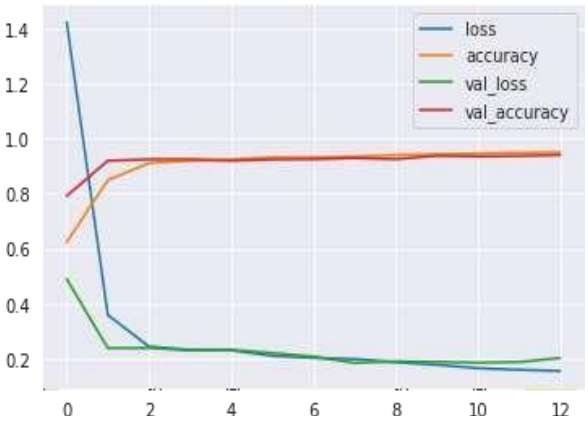

Fig 11 :-this graph represents the Malaria disease training and validation accuracy and training and validation loss

```
0.0150 - accuracy: 0.9948 - val_loss: 2.3504 - val_accuracy: 0.5000
Epoch 24/30
163/163 [==============================] - 75s 458ms/step - loss:
0.0381 - accuracy: 0.9870 - val_loss: 2.1119 - val_accuracy: 0.6875
Epoch 25/30
163/163 [==============================] - 76s 467ms/step - loss:
0.0214 - accuracy: 0.9925 - val_loss: 1.4086 - val_accuracy: 0.7500
Epoch 26/30
163/163 [==============================] - 75s 457ms/step - loss:
0.0143 - accuracy: 0.9938 - val_loss: 1.5934 - val_accuracy: 0.6875
Epoch 27/30
163/163 [==============================] - 75s 463ms/step - loss:
0.0062 - accuracy: 0.9974 - val_loss: 0.1150 - val_accuracy: 1.0000
Epoch 28/30
163/163 [==============================] - 76s 468ms/step - loss:
0.0019 - accuracy: 0.9994 - val_loss: 0.3739 - val_accuracy: 0.8125
Epoch 29/30
163/163 [==============================] - 75s 461ms/step - loss:
0.0044 - accuracy: 0.9988 - val_loss: 1.5382 - val_accuracy: 0.7500
Epoch 30/30
```

Fig 12: - Pneumonia disease accuracy here we can see that training accuracy is 99.88 and validation accuracy is 75.00 after 30epochs and training loss is 0.0044 and validation loss is 1.5382

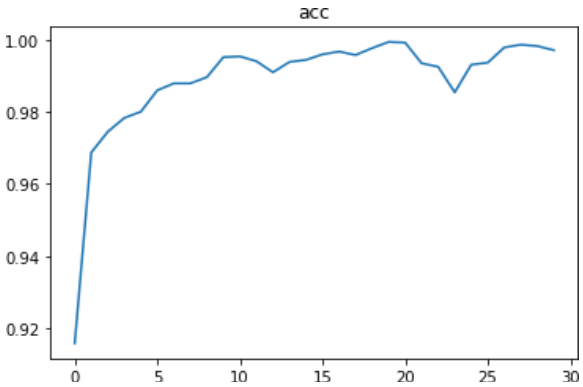

Fig 13: - this figure represents the Pneumonia Training accuracy graph the accuracy of the model is 99.88

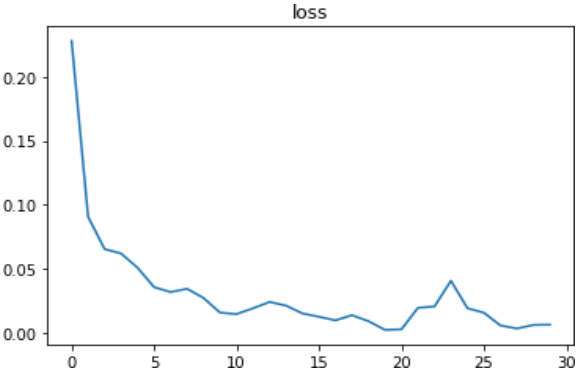

Fig 14: - this figure represents the Pneumonia Training loss graph and the loss of the training data is 0.0044

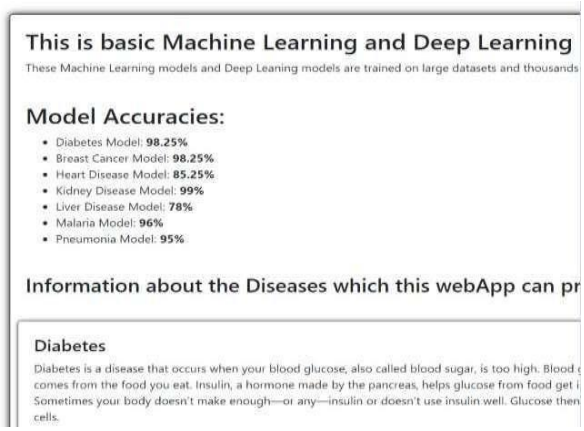

Fig 12 this figure represents home page of the web site Which gives the different discerption of the diseases and accuracy obtained from each disease

VII. CONCLUSION

The aim of this project is to predict disease based on symptoms. The project is set up in such a way that the device takes the user's symptoms as input and generates an output, which is disease prediction. A prediction accuracy probability of 95% is obtained on average. The grails system was used to successfully incorporate Disease Predictor.

REFERENCE

[1] A.Davis, D., V.Chawla, N., Blumm, N., Christakis, N., & Barbasi, A. L. (2008). Predicting Individual Disease Risk Based on Medical History.

[2] Adam, S., & Parveen, A. (2012). Prediction System for Heart Disease Using Naive Bayes.

[3] Al-Aidaroos, K., Bakar, A., & Othman, Z. (2012). Medical Data Classification with Naive Bayes Approach. *Information Technology Journal* .

[4] Darcy A. Davis, N. V.-L. (2008). Predicting Individual Disease Risk Based on Medical History.

[5] JyotiSoni, Ansari, U., Sharma, D., & Soni, S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction.

[6] K.M. Al-Aidaroos, A. B. (n.d.). K.M. Al-Aidaroos, A. B. (n.d.). 2012. *Medical Data Classification with Naive Bayes Approach* .

[7] NisharBanu, MA; Gomathy, B.; (2013). Disease Predicting System Using Data Mining Techniques.

[1] Al-Aidaroos, K., Bakar, A., & Othman, Z. (2012). Medical Data Classification with Naive Bayes Approach. *Information Technology Journal* .

[2] Darcy A. Davis, N. V.-L. (2008). Predicting Individual Disease Risk Based on Medical History.

[3] JyotiSoni, Ansari, U., Sharma, D., & Soni, S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction.

[4] K.M. Al-Aidaroos, A. B. (n.d.). K.M. Al-Aidaroos, A. B. (n.d.). 2012. *Medical Data Classification with Naive Bayes Approach* .

[5] NisharBanu, MA; Gomathy, B.; (2013). Disease Predicting System Using Data Mining Techniques.