A Holistic User Centric Acute Myocardial Infarction Prediction System With Model Evaluation Using Data Mining Techniques

Procheta Nag[*,1], Saikat Mondal[1], Arun More[2]

[1] *Computer Science And Engineering, Khulna University, Khulna-9208, Bangladesh*

[2] *Department of Cardiology, Ter Institute of Rural Health and Research, Murud-413510, India*

A R T I C L E I N F O

A B S T R A C T

Acute Myocardial Infarction (Heart Attack), a Coronary Heart Disease (CHD) is one of the major killers worldwide. Around one thousand data has been collected from AMI patients, people are at risk of maybe a heart attack and individuals with the significant features closely related to heart attack. The sophistication in mobile technology, health care applications offers remarkable opportunities to improve our health, safety and in some sense preparedness to common illnesses. The excess delay time between the onset of a heart attack and seeking treatment is a major issue which may lead to permanent blockage or even die often. So, a proficient mobile application approach is projected in this paper that may predict the possibilities of a attack once an individual is bearing the noticeable symptoms of chest pain. Random forest predicts the result of the user input features and the automated result is shown on the smartphone application. The application categorizes the prediction of the user's input as a heart attack, maybe heart attack and no heart attack. The experimental results showed that the accuracy of the proposed technique is 92%, whereas the precision is 95%, 92%, 87% respectively for heart attack, maybe heart attack and no heart attack prediction. Our research target is to raise heart attack awareness on time in an innovative way through available and accessible medium to mass people.

1 Introduction

Acute myocardial infarction (AMI) commonly referred to as Heart Attack is the most common cause of sudden deaths in city and village areas [1]. Coronary Heart disease (CHD), cancer, chronic respiratory disease, and diabetes are becoming fatal at an alarming rate [2]. Acute myocardial infarction occurs when there is a sudden, complete blockage of a coronary artery that supplies blood to an area of the heart also known as Heart Attack [3]. A blockage can develop due to a buildup of plaque, a substance mostly made of fat, cholesterol and cellular waste products [4]. Due to the insufficient blood supply, some of the heart muscles begin to die.

According to the World Health Organization (WHO) report published in May 2014 coronary heart disease deaths (CHDD) in Bangladesh reached 6.96% of total deaths [2],[5]. Detecting heart attack on time is of paramount importance as delay in detecting may lead to severe damage to heart muscle, called myocardium leading to morbidities and mortalities.

The fast integration of mobile devices into clinical observation has been driven by the rising availability and quality of medical software package applications like mobile applications. There are many Mobile Health (mHealth) tools available to the consumer in the prevention of CHD such as self-monitoring mobile applications. Current science shows the evidence on the use of the vast array of mobile devices such as the use of mobile phones for communication and feedback, smart-phone applications [6]. The visiting fees of doc-

---
[*]Corresponding Author: Procheta Nag, Computer Science And Engineering, Khulna University, Khulna-9208, Bangladesh, +8801747120243, Email:prothoma07p@gmail.com

tors are costly; however, medical applications aim to vary that considerably within the close future. Essentially, it is currently attainable for a smart-phone to interchange the association in nursing in-person doctor consultation, and the virtual appointment with a doctor is apparently less costly than some real-life doctor visits. The mobile applications within the health care business can basically facilitate patients to schedule appointments, monitor the aspect effects of a medicine, prompt them to require bills, analyze health reports, and do a plethora of works. These advance portable wellbeing health applications can alter the way they approach patients and doctors have interaction whereas remodeling the long run traditional way of the medical sector.

The health sector is enriched with data but the major issues with therapeutic information mining are their volume and complexity, destitute numerical categorization and canonical form [7]. Whereas aggregating clinical records and discoveries on paper are not well composed, consequently, numerous information captured in unstructured is troublesome to total and analyze in any steady way. In any case, since it appears to have such an enormous effect, the information has to be exact, comprehensive and convenient. Therapeutic information is time growing and ever-changing, making it outlandish to expect what unused information or modern necessities will see like and how they can fit into a demonstrate. Classification of medical data for idealize supposition is an on the rising field of significance and explore in records expulsion. Whereas creating standard forms such as a portable application that moves forward quality is one of the objectives in wellbeing care.

When observing patients in the clinic, the queries they are inquired and treatment they are given are all noted down in therapeutic records. Collecting and utilizing data from patients reflect the reality of day to day health condition of AMI patients present time. One of the greatest challenges is to free information from the silos in which it usually remains, a difficulty that influences each patient care and medical analysis. Although this information is private, once anonymized this information holds unlimited potential for public benefit, so long because it is utilized securely and viable. Furthermore, information security is a crucial truth because hackers have made wellbeing care information a significant target nowadays. So, as for individuals to feel comfortable sharing their information on the mobile application and encourage them to use the applications are important.

Individuals who are busy in their homes or offices with their regular works and rural people having no knowledge on the symptoms of heart attack may neglect the chest discomfort. Reducing the delay time between the onset of a heart attack and seeking treatment is a major issue [1]. As the medical diagnosis of heart attack is an important but complicated and costly task, we proposed an automated system for medical diagnosis that would enhance medical care and reduce cost. Our aim is to provide an ubiquitous service that

is both feasible, sustainable and which also help people to assess their risk for heart attack at that point of time or later.

In order to provide the automated result of the user data random forest is used to predict the result, as it is an ensemble method to enhance the accuracy of the large and multi-class dataset [8]. Upon the preprocessed data, the individual decision trees are generated using a random selection of attributes at each node to determine the splits. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [9]–[11]. During classification, each tree votes and the most popular class is returned as the predicted class which is the result of the user input, finally, the developed REST API with Python and Flask shows the automated result on smart-phone application by JSON parsing.

2 Related Works

In [12], the author proposed a system for detecting heart attack by analyzing the number of beats per minute (BPM). They have used the sensor to detect the heart attack and intimate the occurrence of heart attack to the helpline in wireless GSM module. In the system, the heartbeat sensor detects the heartbeat rate from the finger of the user and LCD display is used to display that. Then the output is sent to the microcontroller where the microcontroller runs the heart attack detection algorithm, eventually, display the number of pulses in the LCD display. The pulse rate except for the range (60-90) occurs it is regarded as the indication of abnormal or heart attack. Once if abnormal is detected, then the microcontroller will activate the output to the GSM module and GSM module will send the alert message to mobile numbers already coded in a microcontroller.

In [13], the author proposed a system that can automatically predict heart disease. How data be turned into useful information that can enable healthcare practitioners to make effective clinical decisions, was the main objective of this research. They used Cleveland heart disease dataset which is available in the UCI machine learning repository. They have designed a system using Decision Tree (C4.5) as a method that can efficiently discover the rules to predict the risk level of patients based on the given parameter about their health. The rules can be prioritized based on the users requirement. They had used KEEL (Knowledge Extraction based on Evolutionary Learning) tool for prediction.

In [14], the author proposed a system that they had developed a prototype Heart Attack Prediction System (HAPS) using data mining techniques, namely, Decision Trees, NaÃŕve Bayes, and Neural Network. The main objective of this research is to develop a Web Application using data mining modeling technique calls Naive Bayes. The scope of the project is that integra-

tion of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation and improve patient outcome. The system takes some dataset, generate questions depending on them and creates decision calculating them by Naive Bayes classifier.

In [15], the researcher proposed a mobile phone application that can help victims to identify whether they are having a heart attack or not without going to a specialist in person. They have used a mobile phone application with some questions to analyze, a wearable electrocardiogram (ECG) sensor, blood pressure measurement device. If a person is in danger (cardiac arrest, fall) and unable to call an ambulance, the mobile phone will automatically determine the current location of the person using WiFi, GSM Cell-id or GPS and sends automated voice and text messages to their cardiologist.

In [16], the author analyzed the Cardiovascular Disease (CVD) rate in Singareni coal mining regions in Andhra Pradesh state, India. This study analyzed the Cardiovascular Disease (CVD) rate in Singareni coal mining regions in Andhra Pradesh state, India. The have used Decision Trees, Naïve Bayes and Neural Network as their method and UCI repository dataset for their analysis with 15 attributes to predict the morbidity. They have used data mining techniques: Decision Trees, Naive Bayes and Neural Network as their method and UCI repository dataset for their analysis. Bayesian model (BN) achieved a classification accuracy of 0.82 with a sensitivity of 0.87, The decision trees (C4.5) achieved a classification accuracy of 0.825 with a sensitivity of 0.8717. However, the neural network model (MLP) performed the best of the four models evaluated. MLP achieved a classification accuracy of 0.897 with a sensitivity of 0.9017.

In [6], the group of scientist reviewed how mobile health can play important role in cardiovascular disease prevention. The studies reviewed in this statement targeted the behaviors (ie, smoking, physical activity, healthful eating, and maintaining a healthful weight) and cardiovascular health indicators (ie, blood glucose, lipids, BP, body mass index) as the primary outcomes in the clinical trials testing mobile health (mHealth) interventions. They showed many statistics and gave some idea how mobile health (mHealth) can help to prevent cardiovascular disease.

In [17], the authors described an attempt was taken to find out interesting patterns from data of heart patients by these three algorithms namely, Decision Tree, Neural Network and Naive Bayes in two different scenarios. An on-line available dataset of heart patients with Weka data mining software was used for implementation. The experiments consist of two scenarios, one scenario with all 14 attributes and the other scenario with 8 selected attributes. The Naive Bayes classifier algorithm with all attributes shows the highest accuracy i.e. 82.914% and Naive Bayes with selected attributes is nearest to it with 82.077% accuracy. On the other hand, C4.5 decision tree (un-pruned) with all attributes score the lowest accuracy i.e. 77.219%. So, data mining can be used to predict heart disease efficiently and effectively.

Though the approached systems of these papers possess some advantages, the drawbacks of those systems can not be overlooked. The advantages and drawbacks of these papers are mentioned bellow;

Table 1: Merits and demerits

Paper title	Advantages	Drawbacks
Current science on consumer use of mobile health for cardiovascular disease prevention [5]	Demonstrates the great potential that mobile technologies can have to aid in health care	Lack of evidence to some extent
Embedded based automatic heart attack detector and intimator [12]	Automated message system	Sends alert message based on a fixed pulse rate range without ensuring heart attack
Efficient heart disease prediction system using decision tree [13]	The system has great potential in predicting the heart disease risk level	Can predict the risk of heart diseases but can not predict the risk heart attack. Used only 10 attributes
Heart Attack Prediction System Using Data Mining Techniques [14]	Signifies the computer based patients records for clinical supports	Uses categorical data. For some diagnosis the use of continuous data may be necessary
A Self-test to Detect a Heart Attack Using a Mobile Phone and Wearable Sensors [15]	Automatic messaging system from the patients mobile to the emergency number	Can be costly for mass people as there are few external devices and difficult to use by people who has less tech knowledge

Table 2: Merits and demerits

Paper title	Advantages	Drawbacks
Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques [16]	Shows good analytical results on different algorithms to signify early detection of heart attack	Studied on a particular area (coal mining region). Used only 15 attributes to predict heart attack
Data mining in health care for heart diseases [17]	Uses attribute selection method to increase the classification accuracy and decrease the time and complexity	Though dataset had 76 raw attributes but only 14 of them are actually most important to the related topic

3 Methodology

The total system is organized in three modules: at first, preprocessed the raw data, then applied Random Forest algorithm upon the data to predict the possibilities of heart attack and finally the result is showed through the mobile application.
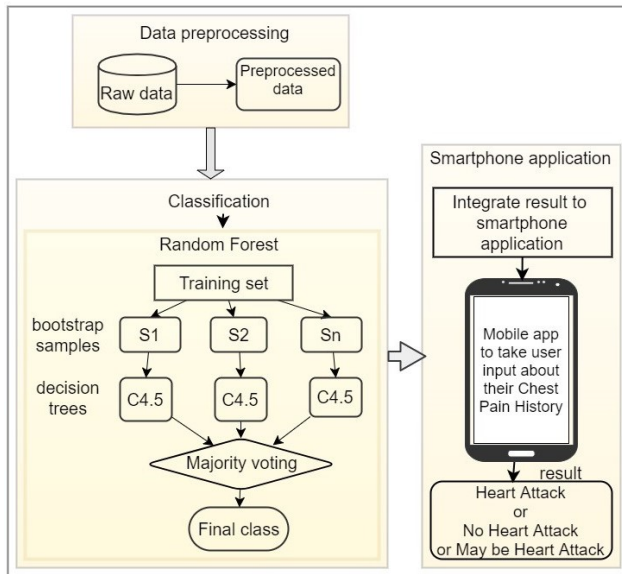

Figure 1: System architecture

The total system is automated, the data mining algorithm is predicting the class of unknown user features on runtime in the back end of the mobile application and predicted class is showed on the application screen.

3.1 Data Preprocessing

Managing the data related to each of the classes and turning it into something usable across a system is one of the major and time-consuming segments for the health data since a total number of 1000 data has been collected from the real world. The datasets consist of different formats (e.g. text, numeric) and sometimes the same data exists in the different dataset and in different formats as they are collected from different sources. Afterward Aggregating this data into a single, central system, inaccurate, incomplete, and inconsistent data may appear as they are commonplace properties of large real-world databases. So, it is obvious to pre-process them in the required format for analysis, hence, it improves the accuracy and efficiency of mining algorithms. Three significant steps of processing dataset are as follow-

1. Trimmed mean is used to handle numerical missing values in the dataset. By deleting a significant portion of outliers, mean is computed using the remaining values, consequently, missing values are filled with the value of the trimmed mean [8].

2. Eliminating redundant features to reduce data size is an emergent step to boost the efficiency of

the training dataset. So attribute subset selection is used to reduces the data set size by removing irrelevant or redundant attributes [8][18].

3. Data are transformed so that the resulting mining process can be more efficient, and the patterns found may be easier to understand. For the shake of correlation, some relevant attributes are added to the original dataset to improve the mining process [8].

3.2 Prediction

After preprocessing, the dataset is trained to the learning model to generate the result that either the chest pain is for heart attack or not or maybe for heart attack. Random Forest, a data mining algorithm [19] [20] is used to predict the classes. There are two stages in the Random Forest algorithm, one is random forest creation, the other is to make a prediction from the random forest classifier created in the first stage. The whole process is shown below-

At first the Random Forest creation:

1. Randomly selected two-third bootstrap sample features from total features where bootstrap sample < total features.

2. Among the sample features, calculated the node d using the best split point.

3. Splited the node into daughter nodes using the best split.

4. Repeated the steps one to three until l number of nodes had been reached.

5. Built forest by repeating steps one to four for maximum number (such as 200 times) times to create a maximum number (such as 200) of trees.
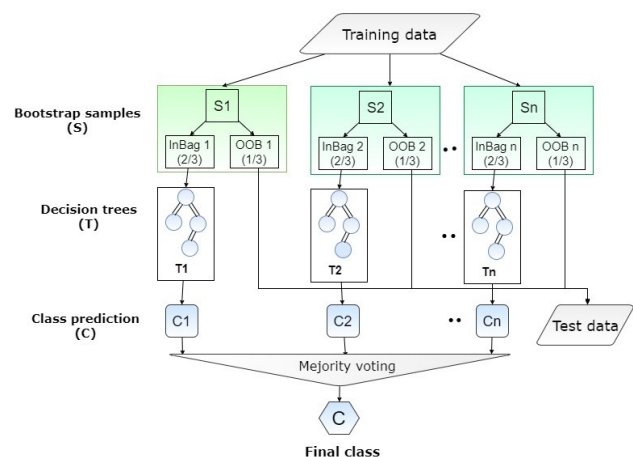

Figure 2: Random Forest algorithm flow chart

In the next stage, while the random forest classifier created, the prediction is made. Prediction procedure by random forest is shown below:

Training data: The collected one thousand data is used as training data of the random forest model.

Test data: User input data from the mobile application is considered as test data which are not classified and unknown because they do not exist in the training dataset. The target is to predict the output class of the test data.

1. Took the test features and use the rules of every randomly created decision tree to predict the outcome and stores the predicted outcome.

2. Calculated the votes for each predicted class.

3. Considered the predicted class which got the majority votes as the final prediction from the random forest algorithm.

Random forest is used for predicting the result because-

Over-fitting reduction: There are considerably lower hazards of over-fitting by averaging some trees.

Handle vast dataset: It has the facility to handle the large information set with higher spatial property and maintain the accuracy of an over sized proportion of knowledge.

Less variance: By utilizing various trees, it reduces the chance of staggering across a classifier that will not perform well owing to the connection between the train and test data.

Feature importance: The Random Forest provides the distinctive vital features from the dataset, in alternative words, feature importance.

3.3 Smart-phone Application

A user input form with heart attack prediction questionnaire which is mostly related to the symptoms of heart attack is made to take input from the users. Whenever a user gives input to the input boxes, the data is sent to the server and generate an automated result from the learning model on the back-end. New data will be saved to the server continuously in order to keep the user records. The resulting process is an off-line and automated system and predicts heart attack on the basis of input that will be provided by the user.
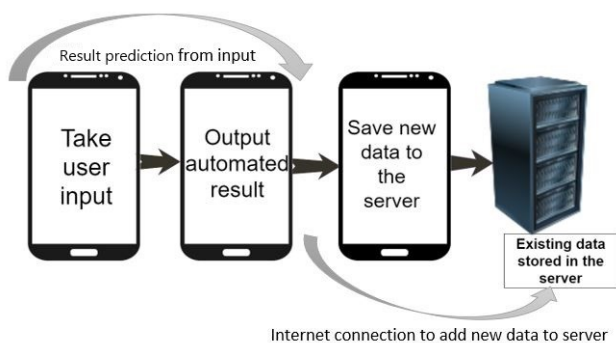
Figure 3: Data flow on smartphone application

4 Implementation and Results Analysis

4.1 Dataset

The model is trained on total 1000 data including 550 AMI patients admitted in three different cardiology specialized hospitals, 280 individuals data who are in risk of heart attack and 170 general people data have been collected and analyzed with 22 attributes which are closely connected to the heart attack symptoms. Age, gender, hypertension, diabetics, cholesterol, smoking, family history, chest pain, without chest pain, chest pain time, chest pain location, chest pain type, chest pain mark, chest pain going, chest pain association, chest pain duration, chest pain subsided, chest pain relieve, past similar pain, doing while it started-are the major features of the model. As a consequence, users input data is also referred to add with the trained dataset.

4.2 Integration

Developed REST APIs with Python and Flask to show the generated result on smart-phone application by JSON parsing.
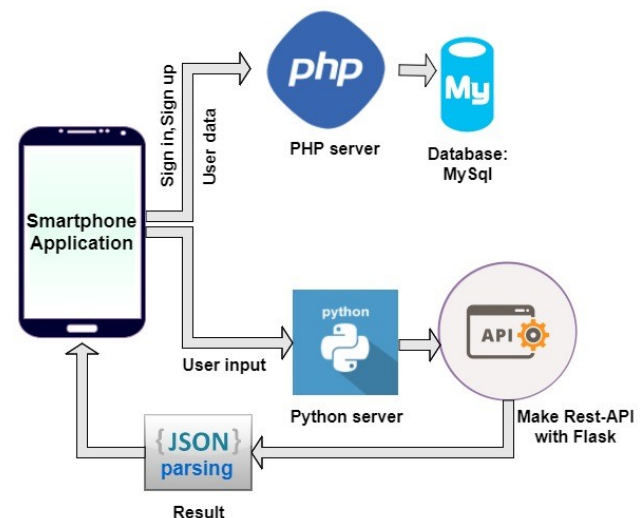
Figure 4: JSON parsing and API creation with Flask

Figure 4.6 illustrates the architecture upon which the mobile application prediction system is built. It shows the flow of how the Mobile application connects to the server and how the server sends particular results according to the request sent by the application. When a user fills up the heart attack prediction questionnaire the inputs are converted to a comma separated string with preprocessing and sent to the server. In the server, the python code takes it as an input (test data) and returns a JSON formatted result. To make this REST API we have used the Python Flask. The JSON data are then parsed inside the application to show the final result (Heart Attack, No Hear Attack,

Maybe Heart Attack). Similarly, for user login and registration, data are sent to the server, executed there and the server sent JSON response which is used by the application to perform specific tasks (Sign in, Sign up). New data will be saved to the server each time after submitting the user form.

4.2.1 Smart-phone Application Functionality

Since it is a cross-platform application it can run any type of smart system as well as smart-phone. The general pages with functionalities of the application are given below

1. A new user has to Sign Up for creating an account whereas an existing user can log in to existing account.

2. This is the Registration page for creating account.

3. After login to check the history of previous analysis result with time click on History. And in order to fill up the questionnaire click on New Entry.

4. A user starts answering the required questions for prediction.

5. After answering all the questions the user may press Calculate Risk button to see what his selected symptoms mean. At the same time, input data will be saved to the server dataset on the presence of Internet connection.

6. In the result, screen user can see the predicting result of with suggestion related to that. Calculating the input data the random forest algorithm predicted the result as Heart Attack along with the necessary suggestion. However, the result may differ according to the user input data.

   If the result shows Maybe heart attack which means his chest pain related symptoms can be for heart attack or lead to heart attack, so he must be aware of that as well as may consult with a cardiologist.

   On the contrary, if the result shows no heart attack that means his chest pain related symptoms are not for heart attack.

7. The user can see his previously checked result history.

After implementing the whole process the accuracy has come with 92%. Based on symptoms of the user input the result has been calculated within three types: heart attack may be a heart attack or no heart attack. Along with accuracy other result measurement criteria are as follow.

4.3 Confusion matrix

The values in the diagonal would always be the true positives (TP) which are indicated by blue color in the table.
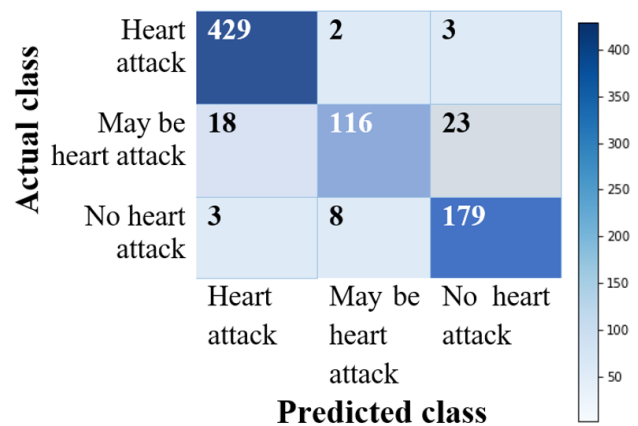
Figure 5: Confusion matrix (without normalization)

429 users data are correctly predicted as heart attack who are actually suffering from AMI and 179 users data are correctly predicted who has no possibility if heart attack apparently. Alongside, 116 data are referred as may be heart attack which actually possesses the symptoms of heart attack.
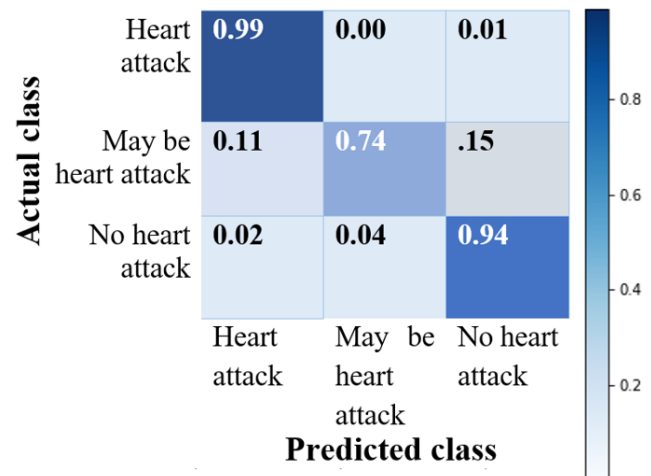

Figure 6: Normalized confusion matrix

Normalizing the confusion matrix it is depicted that the Heart attack prediction rate is very efficient (0.99 out of 1) by the model and also the people who are not suffering from heart attack are almost correctly identified. However a tiny fraction of data has exhibited incorrect prediction.

Accuracy: Accuracy refers that how much a classifier correctly classify the test set that is given to the classifier [8][18].

$$accuracy = (TP + TN)/(P + N) \qquad (1)$$

Table 3: Confusion matrix analysis

Class	Precision	Recall	F1-score	Support
Heart attack	.95	.99	.97	434
May be heart attack	.92	.74	.82	157
No heart attack	.87	.94	.91	190
Avg/Total	.93	.93	.92	781

Table 4: Feature importance

Rank	Serial no	Feature name	Value
1	17	Chestpain_location	0.145674
2	0	Age	0.138248
3	36	Persistence	0.11042
4	20	Chestpain_time	0.094065
5	35	Subsided	0.079373
6	19	Pain_type	0.071126
7	34	Pain_relieved	0.047263
8	33	Doing_while_started	0.039261
9	18	Chestpain_mark	0.035455

The accuracy measure is not appropriate always because it does not check the possibility of tuples belonging to more than one class [8],[18].

Precision: Precision refers to the percentage of correct positive tuples that are labeled as positive [8],[18].

$$precision = (TP/(TP + FP)) \qquad (2)$$

Recall: Recall refers to the percentage of positive tuples which are actually positive. Recall is also known as sensitivity or the true positive rate [8],[18].

$$recall = (TP/(TP + FN)) = TP/P \qquad (3)$$

F1 score: F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account [14]. F1 is usually more useful than accuracy because accuracy works best if false positives and false negatives have a similar cost. If the cost of false positives and false negatives are very different, it is better to look at both Precision and Recall.

$$F1 Score = 2 * (Recall * Precision)/(Recall + Precision) \qquad (4)$$

Around 781 data showed true prediction among the total dataset as support is the number of occurrences of each label in predictive true values. Precision, recall, f1-score, and support have been calculated from the confusion matrix. From precision 93% correct tuples that are predicted as positive, similarly from recall 93% data are predicted as heart attack, maybe a heart attack and no heart attack which are actually that.

4.4 Feature Importance

The major benefit of using ensembles of decision tree methods like Random Forest is that they can automatically provide estimates of feature importance from a trained predictive model. It is a technique to evaluate the importance of features of the model. The result of feature importance with ranking the features on their feature importance of our dataset is given below-

Table 5: Feature importance (cont...)

Rank	Serial no	Feature name	Value
10	30	Association_(Sweating)	0.027388
11	3	Hypertension_medicine_years	0.025257
12	21	Pain_going_(Left arm)	0.01841
13	8	Smoking	0.016303
14	5	Diabetes_medicine_years	0.015461
15	1	Gender	0.015165
16	10	Chestpain	0.014167
17	4	Diabetes	0.014047
18	9	Family_history	0.013066
19	37	Similar_pain/chest_discomfort_in _past	0.01179
20	23	Pain_going_(Back)	0.009163
21	29	Association_(Palpitation)	0.008682
22	2	Hypertension	0.00836
23	28	Association_(Nausea)	0.008006
24	14	Symptoms_Except_Chestpain_(Sweating)	0.007676
25	6	Cholesterol	0.005911
26	32	Association_(Vomiting)	0.005026
27	7	Cholesterol_medicine_years	0.004365
28	22	Pain_going_(Right arm)	0.00368
29	12	Symptoms_Except_Chestpain_(Giddiness)	0.002808
30	26	Association_(Dyspnea)	0.001804
31	11	Symptoms_Except_Chestpain_(Dyspnea)	0.001221
32	24	Pain_going_(Upper jaw)	0.001037

The graph exhibits the sequence of important features to calculate the result in ascending order.

5 Conclusion

Our research was focused on the use of data mining techniques interpreting with a mobile application in health care specifically in identifying acute myocardial infarction. We used training data of heart patients and individuals by collecting from the real-world as well as hospitals whereas test data is received through the mobile application from the users. The mobile application is used to take user input and show the result which is predicted by the random forest algorithm. Also provides the sequence of the importance of the features to check which one has the higher impact of a heart attack. To evaluate the performance of the model prediction different performance metrics were considered where the precision is 93% and recall is 93% in total for heart attack, maybe a heart attack and no heart attack prediction. Finally, the user can check their condition from the application result and be aware of taking the necessary steps.

Table 6: Feature importance (cont...)

Rank	Serial no	Feature name	Value
33	16	Symptoms_Except_Chestpain_(Vomiting)	0.000243
34	15	Symptoms_Except_Chestpain_(Syncope)	0.00008
35	27	Association_(Giddiness)	0
36	13	Symptoms_Except_Chestpain_(Nausea)	0
37	31	Association_(Syncope)	0
38	25	Pain_going_(No_movement)	0

It may get rid of problems connected with human fatigue and habituation raise awareness of abnormalities identifications and alter fast prediction in real time. Development of a mobile application to predict the possibilities heart attack risk would profit lots of individuals. Having a leading framework to suspect the pain as alarming an attack or not an attack might facilitate several such those who tend to neglect the pain and later finally end up within the catastrophe of heart attacks. In this circumstance, we hope that our mobile application PredictAttack will serve the people for useful purposes to taking care of heart through its flexible interactive approach.

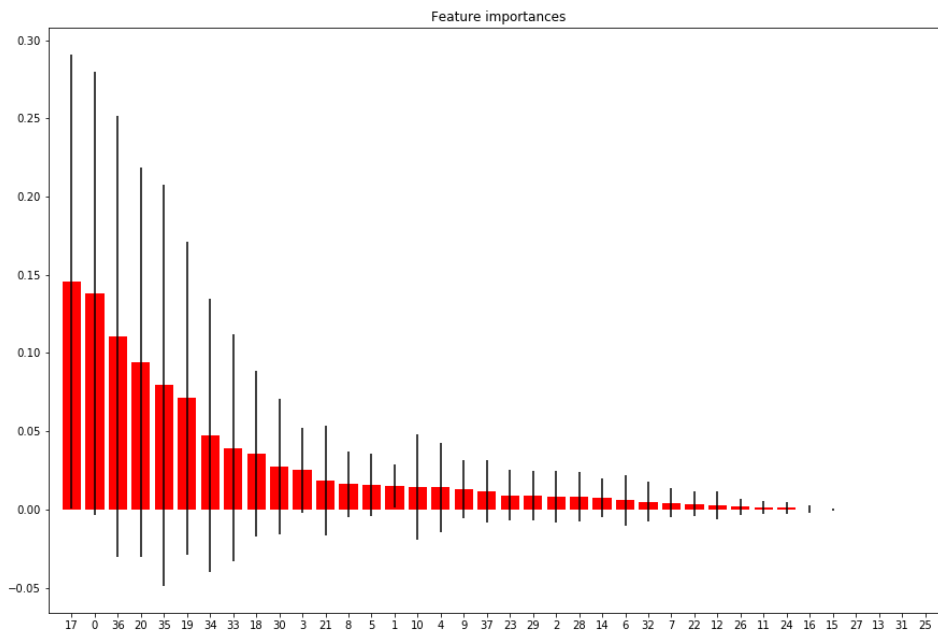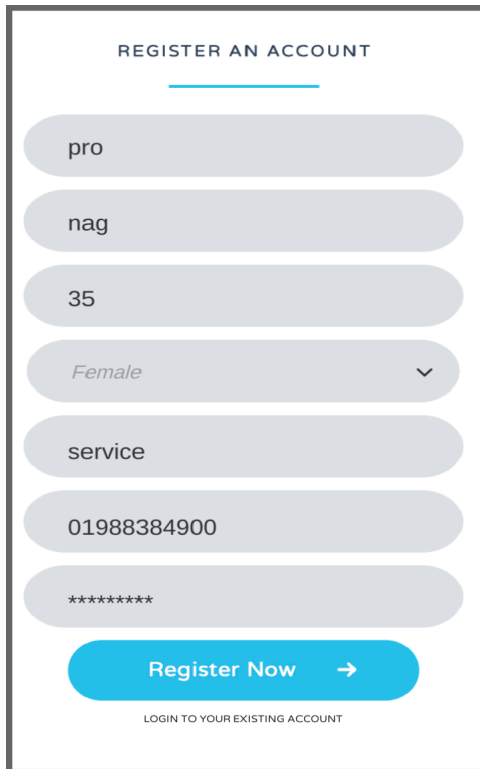**Conflict of Interest**   The authors declare no conflict of interest.

Figure 7: Feature importance

References

[1] P. Nag, S. Mondal, F. Ahmed, A. More, and M. Raihan, "A simple acute myocardial infarction (heart attack) prediction system using clinical data and data mining techniques," in Computer and Information Technology (ICCIT), 2017 20th International Conference of. IEEE, 2017, pp. 1–6. https://doi.org/10.1109/iccitechn.2017.8281809

[2] Islam, AKM Monwarul, and A. A. S. Majumder. "Coronary artery disease in Bangladesh: a review." Indian heart journal65, no. 4 (2013): 424–435. https://doi.org/10.1016/j.ihj.2013.06.004

[3] White, Harvey D., and Derek P. Chew. "Acute myocardial infarction." The Lancet 372, no. 9638 (2008): 570–584. https://doi.org/10.1007/springerreference_44106

[4] Country statistics and global health estimates by WHO and UN partners website: World Health Organization, "Bangladesh: country profiles", 2015. Available: http://www.who.int/gho/countries/bgd/country _profiles/en/, retrieved on 26 July, 2018.

[5] http://www.worldlifeexpectancy.com/bangladesh–coronary–heart–disease, retrieved on 26 July, 2018.

[6] L. E. Burke, J. Ma, K. M. Azar, G. G. Bennett, E. D. Peterson, Y. Zheng, W. Riley, J. Stephens, S. H. Shah, B. Suffoletto et al., "Current science on consumer use of mobile health for cardiovascular disease prevention," Circulation, vol. 132, no. 12, pp. 1157–1213, 2015. https://doi.org/10.1161/cir.0000000000000232

[7] C. Alexander and L. Wang," Big data analytics in heart attack prediction," J Nurs Care, vol. 6, no. 393, pp. 2167–1168, 2017. https://doi.org/10.4172/2167-1168.1000393

[8] J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques,Elsevier, 2011.

[9] A. Liaw, M. Wiener et al., "Classification and regression by randomforest," R news, vol. 2, no. 3, pp. 18–22, 2002.

[10] Khalilia, Mohammed, Sounak Chakraborty, and Mihail Popescu. "Predicting disease risks from highly imbalanced data using random forest." BMC medical informatics and decision making 11, no. 1 (2011): 51. https://doi.org/10.1186/1472–6947–11–51

[11] Livingston, Frederick. "Implementation of Breimans random forest machine learning algorithm." ECE591Q Machine Learning Journal Paper, 2005.

[12] D. Selvathi, V. V. Sankar, and H. Venkatasubramani,"Embedded based automatic heart attack detector and intimator," in Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017 International Conference on. IEEE, 2017, pp. 1–6. https://doi.org/10.1109/iciiecs.2017.8275839

[13] Saxena, Kanak, and Richa Sharma. "Efficient heart disease prediction system using decision tree." In Computing, Communication & Automation (ICCCA), 2015 International Conference on, pp. 72–77. IEEE, 2015. https://doi.org/10.1109/ccaa.2015.7148346

[14] S. A. Pattekari and M. A. Yadav, "Heart attack prediction system using data mining techniques." International Journal of Ethics in Engineering & Management Education, vol. 1, no. 1, pp. 34–37, Jan. 2014.

[15] P. Leijdekkers and V. Gay, "A self-test to detect a heart attack using a mobile phone and wearable sensors," in Computer–Based Medical Systems, 2008. CBMS08. 21st IEEE International Symposium on. IEEE, 2008, pp. 93–98. https://doi.org/10.1109/cbms.2008.59

[16] K. Srinivas, G. R. Rao, and A. Govardhan, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques," in Computer Science and Education (ICCSE), 2010 5th International Conference on. IEEE, 2010, pp. 1344–1349. https://doi.org/10.1109/iccse.2010.5593711

[17] Shafique, Umair, Fiaz Majeed, Haseeb Qaiser, and Irfan Ul Mustafa. "Data mining in healthcare for heart diseases." International Journal of Innovation and Applied Studies 10, no. 4 (2015): 1312.

[18] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[19] Cutler, Adele, D. Richard Cutler, and John R. Stevens. "Random forests." In Ensemble machine learning, pp. 157-175. Springer US, 2012.

[20] Sazonau, Viachaslau. "Implementation and Evaluation of a Random Forest Machine Learning Algorithm." University of Manchester, UK, 2012.

Appendices

Smart-phone application process screen-shot:
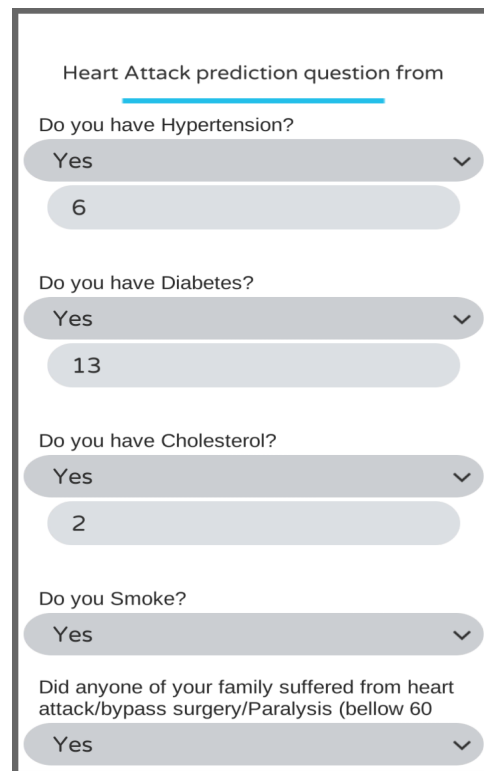
Figure 9: Registration page.

Figure 10: Application screenshot.
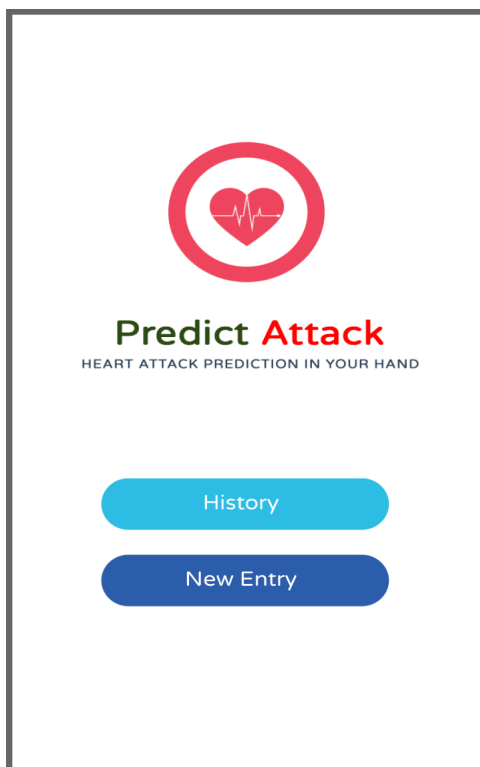
Figure 11: Heart attack prediction question set.

Figure 12: Required questions(continue)

Figure 13: Required questions(continue)

Figure 15: Result screen.

Figure 14: Submition screen.

Figure 16: Checked previous result history