# Smartphone Based Ischemic Heart Disease (Heart Attack) Risk Prediction using Clinical Data and Data Mining Approaches, a Prototype Design

M. Raihan[1], Saikat Mondal[2], Arun More[3], Md. Omar Faruqe Sagor[4],
Gopal Sikder[5], Mahbub Arab Majumder[5], Mohammad Abdullah Al Manjur[5] and Kushal Ghosh[6]
Computer Science & Engineeing, Khulna University, Khulna, Bangladesh[1246]
Email : rianku11@gmail.com[1] , saikatcsebd@gmail.com[2] , shagor110202cse@gmail.com[4]
Ter Institute Of Rural Health, Murud, Latur, India[3] , Email : arunmoregsa@gmail.com[3]
AFC Fortis Escorts Heart Institute, Bangladesh[5]

***Abstract*—We developed a simple approach to predict risk of developing Ischemic Heart Disease (IHD) (Heart Attack) using smartphone. An Android based prototype software has been developed by integrating clinical data obtained from patients admitted with IHD. The clinical data from 787 patients has been analyzed and correlated with the risk factors like Hypertension, Diabetes, Dyslipidemia (Abnormal cholesterol), Smoking, Family History, Obesity, Stress and existing clinical symptom which may suggest underlying non detected IHD. The data was mined with data mining technology and a score is generated. Risks are classified into low, medium and high for IHD. On comparing and categorizing the patients whose data is obtained for generating the score; we found there is a significant correlation of having a cardiac event when low & high and medium & high category are compared ; p=0.0001 and 0.0001 respectively. Our research is to make simple approach to detect the IHD risk and aware the population to get themselves evaluated by a cardiologist to avoid sudden deaths. Currently available tools has some limitations which makes them underutilized by population. Our research product may reduce this limitation and promote risk evaluation on time.**

***Keywords*—Heart Disease, Chi-square, ACS, IHD, Prediction, Data Mining, Android, Smartphone.**

## I. Introduction

Ischemic heart disease is one of the major cause of deaths worldwide. As per WHO's report the death toll is 6.96 % in Bangladesh due to ischemic heart diseases (IHD), which ranks it first as the cause of death [17]. There should be extensive efforts at various levels to reduce the mortality and morbidity out of IHDs [1] [2]. In literature there is enough evidence to show that the risk of IHD increases with the risk factors a person has [12] [13].

Preventive medicine is gaining importance and popularity globally. For some diseases, prevention is better than cure. IHD episodes can be primordially, primarily and secondarily prevented. A strategy to prevent development of risk factors or control of risk factors will modify the risk of IHD.

Having episode of IHD has long lasting effects on the individual from having repeated episodes of chest pain to having limited quality of life due to heart failure and even death. It is the need of the hour to make maximum number of people aware for health and make them to walk up to a doctor for preventive health checks.Calculating and categorizing a person into a risk level may help the person to motivate for a preventive health check so that mortality can be reduced as much as possible.

Smartphone mobile devices are one of the most widely utilized technology worldwide. Through it, various applications are being developed and being used by general population for various means. In fact, there is good mental impact of information when mobile application tells about something to an individual. Development of a smartphone based tool would benefit masses of people. Currently there are certain tools available to predict risk of IHD which bears excellent academic excellence but are underutilized by general people due to few limitations like mandatory inputs of Lipid profiles and Blood Pressure (BP) values moreover, these risk scores do not individualizes the risk. Development of a tool to predict the heart attack risk based on risk factors may push much of population to check their own risk.

In hospitals everyday huge number of data is collected of the patients but it is not mined. As there is enough clinical data available, mining the data can lead to developing an android application which will calculate the risk for IHD by giving certain inputs and categorizing the individual into risk as low, medium or high.

We developed a simple approach to predict risk of developing Ischemic Heart Disease (IHD) (Heart Attack) using smartphone based application in android platform. The application generates a score based on the input and categorizes the person into low, medium or high risk. The application should guide the person for further work up based on the result.

The rest of the paper is organized as follows: in section 2 and 3 related works and methodology are discussed respectively. In section 4 experimental result and performance analysis is presented. Finally the paper is concluded with section 5.

## II. Related works

Peter W.F. Wilson et. al. [3] proposed some guideline and identified risk factor coronary heart disease (CHD). This paper was to examine blood pressure and cholesterol categories with coronary heart disease risk and to incorporate them

into coronary prediction algorithms. They set a community based group. The sample size was 5345 ; with 12 years of follow up and demonstrated the association between IHD and hypertension, diabetes and dyslipidemia.

Ralph B. D'Agostino et. al. [4] proposed a system to accurately predict CHD risk, of the Framingham functions compared with the performance of risk functions developed specifically from the individual cohorts data. Their comparisons included evaluation of the equality of relative risks for standard CHD risk factors, discrimination, and calibration. They used cox re-gression coefficients and chi square test as statistical measures. The Framingham functions performed well for both black and white men and women. This events was taken within 5 years of follow up. In their score system High density Lipoprotein ( HDL), Low Density Lipoprotein (LDL), diabetes, blood pressure etc. inputs are mandatory.

Lora E. Burke et. al. [5] and a group of scientist reviewed how mobile health can play important role in cardiovascular disease prevention. They showed many statistics and gave some idea how mobile health (mHealth) can prevent cardio vascular disease. They recommend some future research such as mobile application for treating obesity encouraging regular physical activity, smoking cessation, control of hypertension, and dyslipidemia; and treating diabetes mellitus.

## III. Methodology

Data of 506 patients admitted with symptoms in a cardiac hospital and 281 individuals from general population was recorded and analyzed. Total 70 attributes were collected. The variables analyzed for the purpose of this study were age, gender , presence or absence of hypertension, diabetes, Dyslipidemia, smoking status, a significant family history, habit of exercise, presence or absence of symptoms suggestive of IHD were considered which are shown in Fig.3 and Fig.4. These risk factors are well proven for their association with developing IHD in various studies. For developing score we used chi square correlation and Fisher's exact test. The score was integrated for developing an android application.

The system architecture:

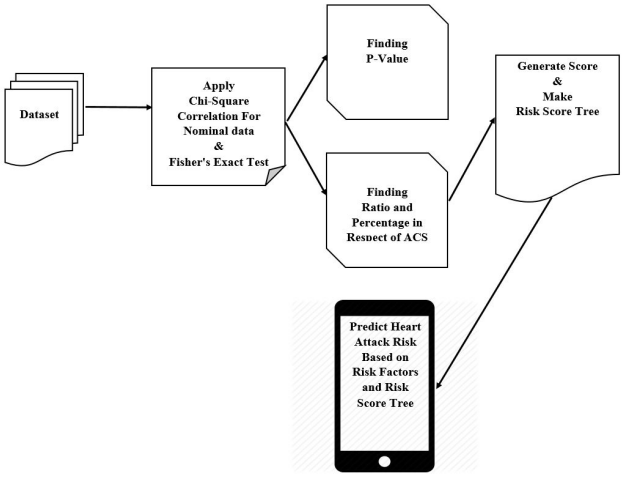


Fig. 1: System Architecture

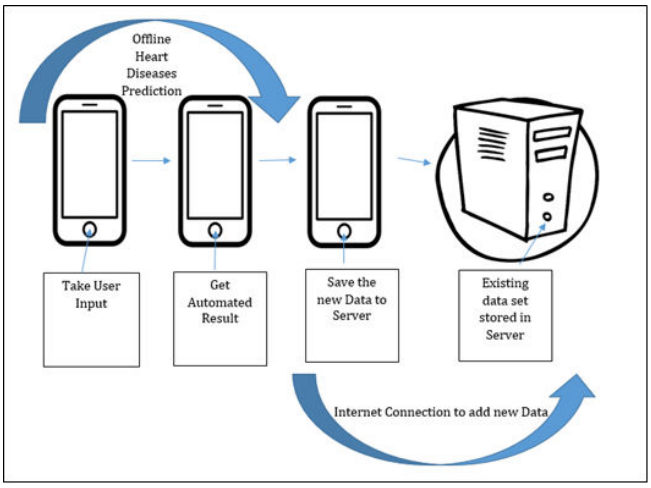


Fig. 2: Server communication with mobile application

*1) Risk Score Tree:* To construct this tree we used chi square correlation and finding the p-value. p-value was used to find out the level of significance [10] [11]. A score was given which was analyzed from ratio and percentage in respect of Acute Coronary Syndrome (ACS) which is IHD and a histogram was generated.

*2) Chi-Square Correlation:* Chi-square test discovers the nominal data correlation between two attributes [7][8][9]. If a has c distinct values, namely a1, a2 .... ac and b has r distinct values namely b1, b2 ... br. The data tuples can be shown as table where with the c values of a making up the columns and the r values of b making up the rows [6][7]. The chi-square value (also known as Pearson chi-square statistic) is computed as [6][7][8][9]:

$$\tilde{\chi}^2 = \sum_{i=1}^{c}\sum_{j=1}^{r}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

*3) p-value:* p stands for probability. The p value, or calculated probability, is the probability of finding the observed, or more extreme results when the null hypothesis (H0) of a study in question is true [14][16]. The definition of extreme depends on how the hypothesis is being tested. p is also described in terms of rejecting H0 when it is actually true, however, it is not a direct probability of this state [14][15][16].p-value may take a value between 0 and 1[14].
0.1 : 10% significance level ; weak evidence against H0
0.5 : 5% significance level ; strong evidence against H0
0.1 : 1% significance level ; very strong evidence against H0

*4) Fisher's Exact Test:* It is for a single contingency table [15]. It is a statistical significance test used in the analysis of contingency tables [16]. Although in practice it is employed when sample sizes are small, it is valid for all sample sizes [16]. For hand calculations, the test is only feasible in the case of a 2*2 contingency table. However the principle of the test can be extended to the general case of a m*n table [15][16] (Table I).
To check significance among a,b,c,d The p is calculated as

p= ((a+b)! (c+d)! (a+c)! (b+d)!)/(a! b! c! d! n!)

TABLE I: Calculating P Value in Fisher's Exact Test

| | A | B | Total |
|---|---|---|---|
| X | a | b | a+b |
| Y | c | d | c+d |
| Total | a+c | b+d | a+b+c+d |

TABLE II: Risk score

| Attribute | Sub Category | Score |
|---|---|---|
| Age | 15-30 | **1** |
| | 30-40 | 2 |
| | 40-55 | 3 |
| | Above>=55 | **4** |
| Sex | female | 2 |
| | male | 4 |
| Smoking | yes | 4 |
| | ex | 3 |
| | no | 1 |
| HTN | no | 2 |
| | yes | 4 |
| DLP | no | 2 |
| | yes | 4 |
| DM | no | 2 |
| | yes | 4 |
| Physical Exercise | no | 4 |
| | yes | 1 |
| Family History | no | 2 |
| | yes | 4 |
| Drug History | no | 2 |
| | yes | 4 |
| Psychological Stress | no | 2 |
| | yes | 4 |

*5) The Score Calculation:* Age was sub-categorized into 4 and given scores based on level of significance and association with having IHD. Age between 15-30 years is sub categorized as score 1 , between 30-40 years as score 2 and between 40-55 years as score 3 and a highest score of 4 for users above 55 years of age (Table II).

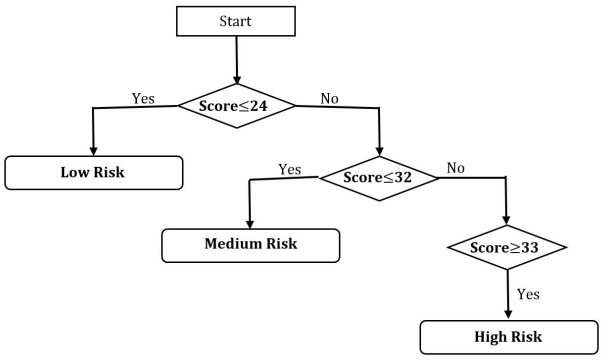


Fig. 3: Risk score tree

With similar calculations score was generated for other variables (Table II).

Total Highest score 40 and the Lowest was 17.

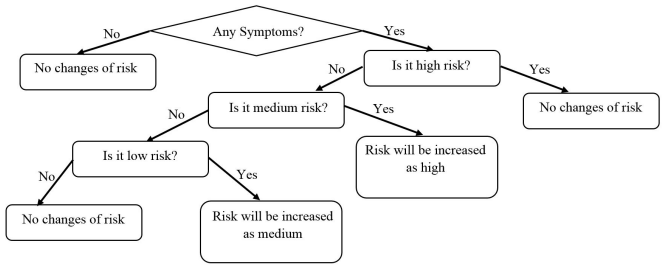


Fig. 4: Symptom tree

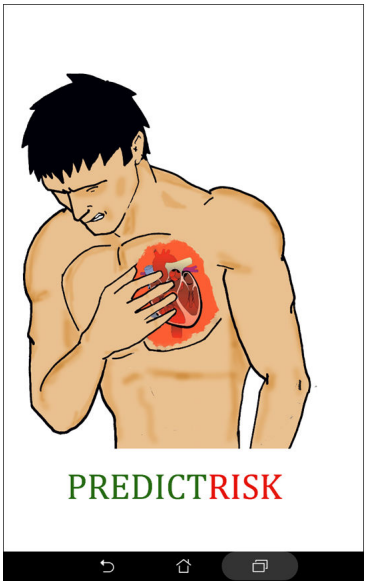


Fig. 5: Application screen shot 1

Difference = 40-17 = 23
Interval = 23/3 =7.66 = 8

So, score of 17-24 was categorized into Low Risk, 25-32 as Medium and 33-40 as High Risk category . It was combined into risk score tree (Fig.3) with symptoms (Fig.4) and was integrated it to our mobile application.

*6) Android Application Features:* The data and score tree generated was integrated to the android software named **PredictRisk**. After reading and accepting the terms and conditions the user is registered with information like identity, address, age, gender , profession and contact details. A input format consisting of questions related to history of Hypertension, Diabetes, Dyslipidemia (Cholesterol Abnormality) , Family history, Drug history, Stress, exercise habit and symptoms was developed. The user gives a specific input about the existing known risk factors and symptoms upon which the application calculates the score and categorizes the person into a risk.

The risk is categorized as per score generated for variables of risk factors but if the user gives an input of having one or more symptoms, the risk level ascends up by one; e.g. By risk factor input if the generated score categorizes the user to

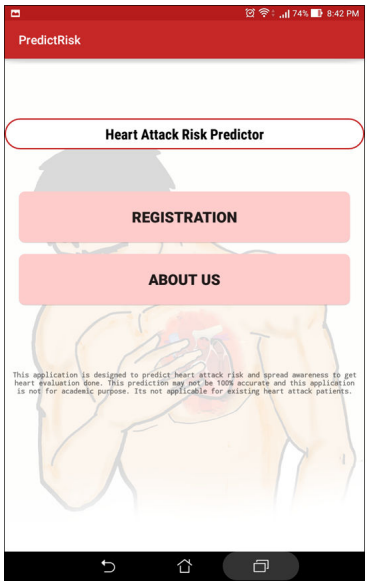


Fig. 6: Application screen shot 2

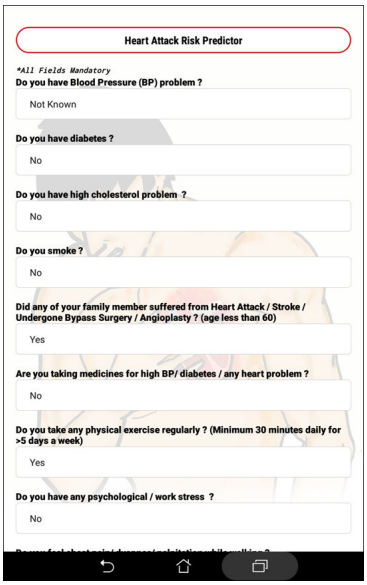


Fig. 8: Application screen shot 4

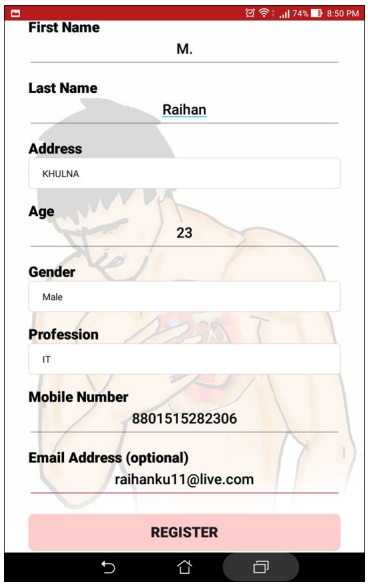


Fig. 7: Application screen shot 3

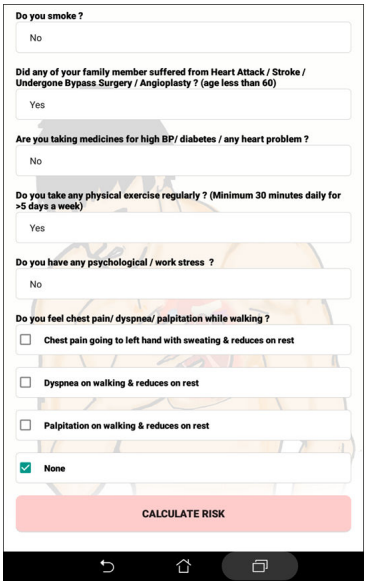


Fig. 9: Application screen shot 5

low category, additionally the user gives input to have one or more typical symptoms; the end result shown will be medium category. This has been designed to avoid underestimation of risk due to unawareness of the risk factor the user may have, as irrespective of awareness of having the hypertension and/or diabetes and/or dyslipidemia, the disease process is active for the person which may lead to having an episode of IHD.

The result is reflected on a page denoting the risk category and with an advice to get evaluated by a cardiologist. The page also reflects addresses of some of medical centers where the risk assessment can be performed on a Google map. The application is synchronized with a server and the data along with the score generated gets recorded in our server.

*7) Project Environment and Necessary Tools:*

- SPSS
- Java
- XML
- PHP
- Android Studio 2.1.1
- Server Storage and Domain

## IV. Experimental Results and Analysis

On comparing and categorizing the patients whose data was obtained for generating the score; we found there is a

significant correlation of having a cardiac event when low and high category was compared ; p=0.0001. Among patients categorized into high; 86.7% had IHD while in low category, only 12.5% had IHD. The difference between medium and high was also significant with p=0.0001; However there was no significant difference between medium and low risk categories; thus medium category remaining in the grey zone. Our application needs to be tested further for sensitivity and specificity. The mobile screen shots with active application **PredictRisk** Fig.5 to Fig.11

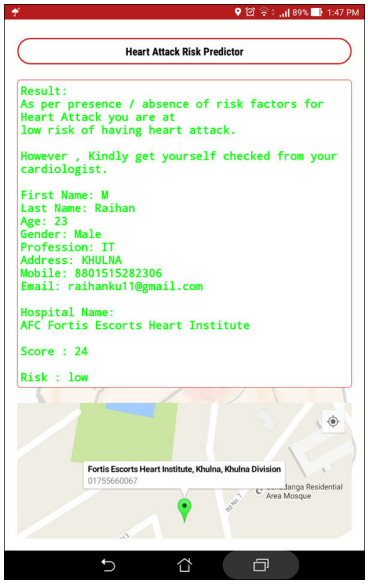


Fig. 10: Application screen shot 6

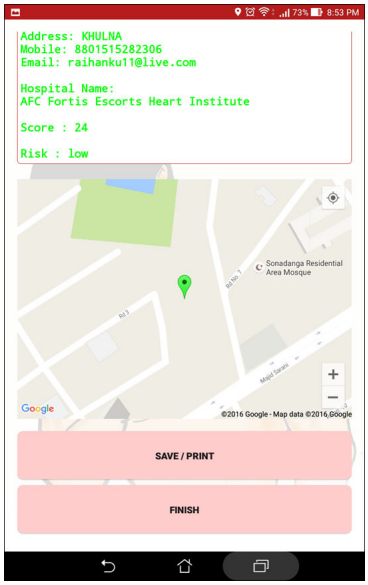


Fig. 11: Application screen shot 7

## V. Conclusion

Mobile applications are playing vital roles in day today life. They could help billions of people to access health information, health advices and guidelines. Our research is motivated to make simple approach to detect the IHD risk and aware the population to get themselves evaluated by a cardiologist to avoid sudden deaths and morbidities. Currently available tools has mandatory input of lipid values which makes them underutilized by population; though, those risk calculators bear excellent academic importance . Our research application **PredictRisk** reduce this limitation and promote a risk evaluation on time.

Our research is in experimental phase to check sensitivity and specificity and how the population takes it to walk upto a cardiologist to check their risks to have IHD. The **PredictRisk** could be path breaking research to make population aware about the complications of risk factors and prevention or early health check may save ones life.

## References

[1] A. Islam and A. Majumder,"Coronary artery disease in Bangladesh: A review", Indian Heart Journal, vol. 65, no. 4, pp. 424-435, 2013.

[2] M. Abu Sayeed, H. Mahtab, S. Sayeed, T. Begum, P. Khanam and A. Banu, "Prevalence and risk factors of coronary heart disease in a rural population of Bangladesh", Ibrahim Med. Coll. J., vol. 4, no. 2, 2010.

[3] P. Wilson, R. D'Agostino, D. Levy, A. Belanger, H. Silbershatz and W. Kannel,"Prediction of Coronary Heart Disease Using Risk Factor Categories", Circulation, vol. 97, no. 18, pp. 1837-1847, 1998.

[4] R. D'Agostino, Sr, S. Grundy, L. Sullivan, P. Wilson and for the CHD Risk Prediction Group,"Validation of the Framingham Coronary Heart Disease Prediction Scores", JAMA, vol. 286, no. 2, p. 180, 2001.

[5] L. Burke, J. Ma, K. Azar, G. Bennett, E. Peterson, Y. Zheng, W. Riley, J. Stephens, S. Shah, B. Suffoletto, T. Turan, B. Spring, J. Steinberger and C. Quinn, "Current Science on Consumer Use of Mobile Health for Cardiovascular Disease Prevention", Circulation, vol. 132, no. 12, pp. 1157-1213, 2015.

[6] I. Witten, E. Frank and M. Hall, Data mining. Burlington, MA: Morgan Kaufmann, 2011.

[7] J. Han, M. Kamber and J. Pei, Data mining. Amsterdam: Elsevier/Morgan Kaufmann, 2012.

[8] P. Tan, M. Steinbach and V. Kumar, Introduction to data mining. Boston: Pearson Addison Wesley, 2005.

[9] "Statistical Analysis 5: Chi - squared test for 2 - way tables", statstutor. [Online]. Available: http://www.statstutor.ac.uk/resources/uploaded/coventrychisquared.pdf.

[10] K. Ahmed, T. Jesmin and M. Zamilur Rahman, "Early Prevention and Detection of Skin Cancer Risk using Data Mining", International Journal of Computer Applications, vol. 62, no. 4, pp. 1-6, 2013.

[11] T. Jesmin, K. Ahmed, M. Zamilur Rahman and M. Badrul Alam Miah, "Brain Cancer Risk Prediction Tool using Data Mining", International Journal of Computer Applications, vol. 61, no. 12, pp. 22-27, 2013.

[12] Country statistics and global health estimates by WHO and UN partners website: World Health Organization, "Bangladesh: country profiles", 2015. [Online]. Available: http://www.who.int/gho/countries/bgd/country_profiles/en/.

[13] World Health Organization - Noncommunicable Diseases (NCD) Country Profiles, 2014 website: http://www.who.int/nmh/countries/bgd_en.pdf

[14] "P Values", Statsdirect.com, 2016. [Online]. Available: http://www.statsdirect.com/help/default.htm#basics/p_values.htm

[15] C. R. Mehta and N. R. Patel, "IBM SPSS Exact Tests", University of Sussex, 2012. [Online]. Available: http://www.sussex.ac.uk/its/pdfs/SPSS_Exact_Tests_21.pdf.

[16] "Fisher's exact test", Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Fisher%27s_exact_test#cite_note-1.

[17] "Coronary Heart Disease in Bangladesh", World Life Expectancy, 2014. [Online]. Available: http://www.worldlifeexpectancy.com/bangladesh-coronary-heart-disease.