Early Predictive Analysis for Heart Attack Identification

MUHAMMAD ZARYAB KHAN
Information Technology
Altinbas University
Istanbul, Turkey

TIMUR INAN
Computer Engneering
Altinbas University
Istanbul, Turkey

***Abstract*—In many countries, especially in underdeveloped countries, proper health care service is a primary issue. Medical sectors are distant and on the other hand medical associated bodies are defective in comparison with the requirements of the people. For this reason, the daily health monitoring is difficult for the normal person. This lacks the monitoring of healthy people. To provide necessary primary care to the citizens, one of the ways could be the implementation of desirable systems get this issue solved. The "Internet of Things (IoT)" and Data Science are playing a vital role in variety of areas like safe and secure environment, Artificial intelligence-based commute systems, digital cities, smart operating factories, and health relate fields. The main purpose of this paper is to focus on the application of Data Science and "IOT" in the healthcare system and to bring forward a solution for human health through its blood vitals. The compositional proposed idea can be used to acknowledge blood pressure monitoring to secure heart attacks and other diseases and also it will aware for the possible diseases and also it will suggest for the medical cure which can be use as the first aids.**

Keywords*— *systolic, diastolic, BPM, pluses waves, Microcontroller

I. INTRODUCTION

Heart disease are one of the most common diseases in our society, about 17.9 million people are die each year in the account of cardiovascular diseases. Many of the patient does not aware about their disease before heart attack, in this paper we will discuss about to predicate heart attack and possible disease with the help of IOT gadget by the help of algorithms and also suggest for the medication related to the disease

II. EASE OF USE

A. Heart Attack

Heart attack is caused by the blocking in heart of blood stream which isn't giving an adequate sum of oxygen required for your heart. Which makes blood pressure in the heart, The threshold of hypertension condition is "**systolic pressure**" which should be less than 120 and other is "**diastolic pressure**" which should be less than 80 for the adults. Angina could also be caused as of receiving less amount of oxygen which is the signal of heart attack

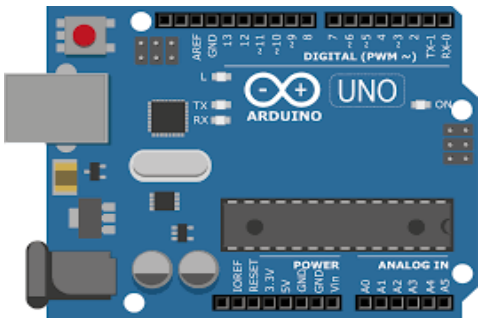
- The common signs and symptoms of heart attacks are pain, aches and stiffness, pressure, burdensomeness, tightness, squeezing, or pain in various upper body areas like in your chest or arm or it could be below your breastbone.
- The number of BPM (beats per minute) might be also affected on heart, a pulse rate of adults is between 60 and 100 beats per minute which is related to supply of oxygen in our heart

B. BPM sensor

A BPM sensor is commonly used for calculating the pulse wave in the volume of blood vessel that occurs when the heart pumps blood. It detects the pluses waves and also changing in pluses waves

C. Arduino

"It is a process of modeling the probability of a discrete outcome given and input variable." It is utilized to predicate the category or class of audience based on one or different indicator factors



D. Logistic Regression (Binary Classification)

"It is a process of modeling the probability of a discrete outcome given and input variable." Which is predicate the

class or category of multiple predictor variables or individuals based on one.

III. CONFIGURATION OF HARDWARES

With the help of BPM sensor, we will get the pulse wave in the volume of blood vessel which will be embedded in Arduino that will send to our mobile application which will through that data in our database.

IV. CONFIGURATION OF HARDWARES

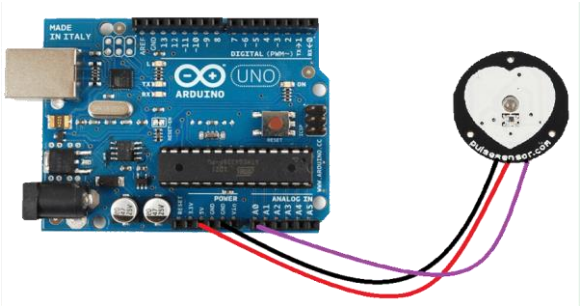
With the help of BPM sensor, we will get the pulse wave in the volume of blood vessel which will be embedded in Arduino that will send to our mobile application which will through that data in our database. Each person has their own BPM values separately, also we have a dataset which will contain all the history of patient which was affected by heart disease. When the threshold of bpm will be cross the data will be check on our dataset. That which data is related that our current data and what was the outcomes of related data history. To providing the desire result logistic regression algorithm will help us.

1. Configuration of bpm Sensor with Arduino

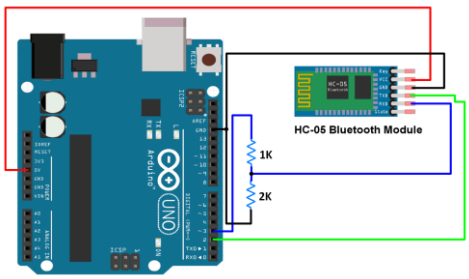
   There are three wires in BPM sensor

   - Ground
   - VCC
   - Signal

The ground pin put in ground slot, VCC require 5v, and signal put in A0 slot



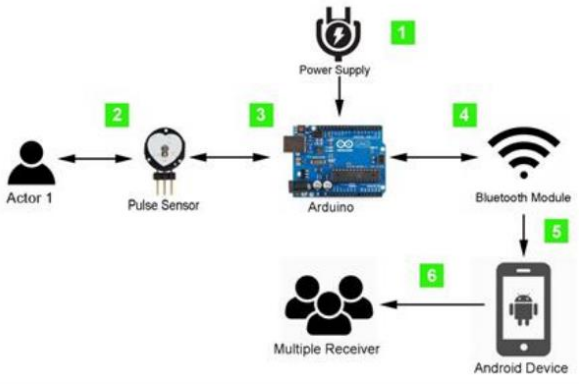
2. Configuration of Bluetooth module



The TX (Transmitter) put in RX slot and Rx (Receiver) put in TX slot, The VCC pin connect with +5V slot and Ground connect with ground slot

V. METHDOLOGY

Framework is isolated into hardware as well as software parts. The equipment module is embedded on Microcontroller. The purpose of this microcontroller is to serve as the main control for all the operations to be carried out by the equipment components. Bluetooth module performs as a middleware of sensor and the smartphone device, through a bi-directional communication, and networking the hardware and program elements of the framework with the Android.



The BPM sensor provide the data to micro controller which send to android device that will send data in our database, it matches the data with dataset and provide the appropriate result to the end users.

1. ALGORITHM WORKING

	Patient_Number	Blood_Pressure_Abnormality	Level_of_Hemoglobin	Genetic_Pedigree_Coefficient	Age	BMI	Sex
0	1	1	11.28	0.90	34	23	1
1	2	0	9.75	0.23	54	33	1
2	3	1	10.79	0.91	70	49	0
3	4	0	11.00	0.43	71	50	0
4	5	1	14.17	0.83	52	19	0

This dataset has provided Pavan Bodanki on kaggle site

- Data Exploration

  By the following code we are creating the plot of our dataset, which takes the coordinates of ages and blood pressure

```
plt.scatter(heart_data.Age,heart_data.Blood_Pressure_Abnormality,marker='+',color='red')
```

- Splitting Dataset into training and Test dataset

  Here we are splitting 10% dataset in test dataset and rest of the 90% into train dataset also we are separating the sizes of our train and test dataset. Which help to get the lengths of our datasets

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(heart_data[['Age']],heart_data.Blood_Pressure_Abnormality,test_size=0.1)
```

After the splitting we have data in our train set 1800

```
len(X_train)

1800
```

After the splitting the dataset is providing random data

```
X_test
```

	Age
443	37
904	64
537	68
498	72
1805	64
...	...
548	68
1071	29
1352	26
1278	43
945	30

- Logistic Regression Prediction

Now data can be predicted by logistic method

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.fit(X_train,y_train)
LogisticRegression()
```

```
array([0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1,
       1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0,
       1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1,
       1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1,
       1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0,
       0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0,
       1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0,
       1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0,
       0, 1])
```

Here the logistic function is showing how many patients has been affect by heart disease by "1" and how many does not have heart disease by "0".

A. *Equations*

All sigmoid functions have the property that they outline the complete number line into a little run such as between and 1, or -1 and 1, so one utilizes of a sigmoid work is to change over genuine esteem into one that can be deciphered as a likelihood:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

Here e = Euler's number ~2.71828 which converts input into range 0 to 1

In logistic regression, we determine the probability of edges. We categorize that component in a single group or the other way round if the probability of specific component is higher than the probability of edges.

Firstly, we establish the best possibly fitted line by applying linear Regression.

$$Y = mx + c \tag{2}$$

To determine the best-fitted, assign random values to m and c. By this method, we can conclude the values of Y for a given value of x.

The value of Y is a dependent value because "logistic Regression is a supervised Machine Learning algorithm". So, it is easy to verify the prediction value is accurate or not.

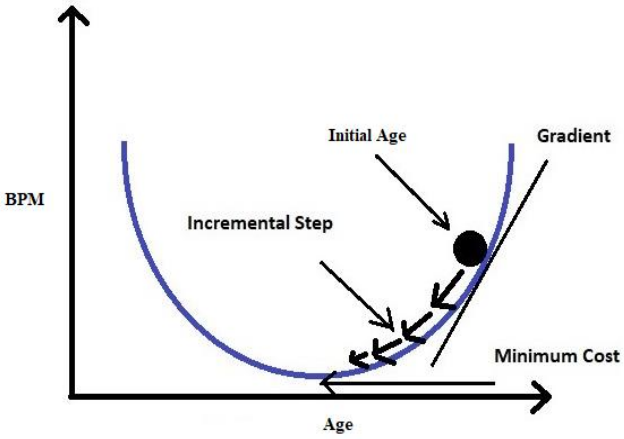
Y_1 (bpm)	Y_2 (bpm)	Error rate $\lvert Y_1 - Y_2 \rvert / \lvert Y_2 \rvert$	Standard deviation (σ)
98	100	0.02	
91	96	0.052	4.1533
92	94	0.0212	
101	103	0.0194	

In case of Error values, we can use MSE method that is also known as the loss function.

$$L = 1/n \sum((y - \hat{y})^2) \tag{3}$$

Here n is the set of taken observations.

In order to get the best-fitted line, we will need to reduce the count of loss function by the technique of gradient descent.



On the similar note, we will get the first-order derivative for the loss function of age. Afterwards, we can abstract the derivate through the initial age to multiplying with a learning rate (α). Loop this step until gets the minimum value.

After the loss function is minimized, we conclude the final equation of best possibly fitted line.

After that apply Logistic Regression by using equation 1

The sigmoid function's output is now converted into 0 and 1 based on its threshold value. Normally we establish the threshold as 0.5

Here, we have an extra domain BPM and we have to classify whether an individual is corpulent without depending on their given Age as well as BPM

This is an obvious categorization issue where we got to isolate the dataset into two major categories, one is Obese and other is Not-Obese

- Comparison between "Linear and Logistic Regression"

Linear Regression: Utilized when the required yield requires nonstop esteem based on anything input/dataset given to the calculation. Assume you need to create a program that would anticipate the average temperature of saying tomorrow, based on specific highlights, like average temperature, least temperature, greatest temperature, etc. of the past week. Since this issue needs yield as esteem of ceaseless nature, it is classified as a linear regression problem.

Presently assume your issue was not to yield the temperature, but the sort of climate that tomorrow might have (e.g., sunny, cloudy, stormy, stormy, etc.). This issue will donate a yield having a place to a specific set of values predefined, consequently it is fundamentally classifying your yield into categories. Classifying issues can be double (yes/no, o/1) or multiclass (just like the issue depicted over). Calculated logistic regression is utilized in classifying issues of machine learning

VI. SUGESSTION OF MEDICATION

In our dataset we will a have field which will preserve the history of patient, where it is determine that which medication doctor give him at that certain condition, by the help of cluster we can also suggest the medication to the user in the case of emergency.

Suppose Beta-blockers is one of the most generic medication doctors suggest while increasing the heart rate.

Patient	Favorability of Current Regimen of Medications	Additional Treatment Needed For:	Candidate Medications Being Considered:	Utility Value	Favorability Updated to Include Candidate Medication	Candidate Medication Chosen
A (male)	0.500	Arthritic pain	Acetaminophen	0.520	0.504	NSAID
			NSAID	0.561	**0.511**	
B (male)	0.545	Hypertension	Beta-blocker	0.452	0.525	Diuretic
			Diuretic	0.505	**0.537**	
C (female)	0.629	Symptoms of schizophrenia	Olanzapine	0.698	0.642	Risperidone
			Risperidone	0.705	**0.644**	
D (female)	0.657	Depression	SSRI	0.615	**0.649**	SSRI
			TCA	0.593	0.644	
E (male)	0.544	Parkinson disease	Levodopa	0.683	0.569	Pramipexole
			Pramipexole	0.710	**0.574**	

Medication suggestion = $(0.499 \times 0.452 \times 0.521 \times 0.530 \times 0.520)^{0.20} = 0.504$.

By the help of probability, we can suggest the medication

Medication Suggestion = [Beta-blockers x Digoxin x Beta-blockers x Beta-blockers x Antianxiety medications]1/5

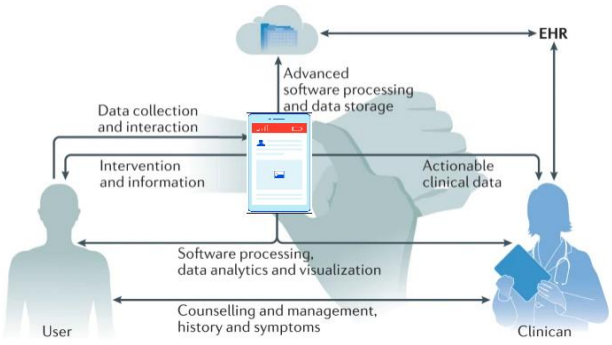
In the paper of author "Peter wanger & Lene Martin from department of Clinical Neuroscience, Karolinska institute St Erik's Eye hospital", Sweden suggests the pseudocode to suggest the medicine in a different method.

```
function reko(){
.....
.....
if (systolic_dysfunction == true){
document.forms[0].digoxin_text.value = "Useful";
document.forms[0].antiarrythmics_text.value = "Stop treatment"
document.forms[0].aceinhibitors_text.value = "Useful, decreases
    afterload";
document.forms[0].angiotensin2_inhibitors_text.value = "Useful,
    decreases afterload";
document.forms[0].other_info_text.value = "Inotropic drugs are
    useful. Stop treatment with NSAID and non-heartselective
    Ca++-blockers";
        }
.....
.....
.....
}
```

Here it is clearly defined if the patient BPM threshold cross the systolic condition the treatment will be stop also it is suggesting the drugs according to data facts

VII. DIAGNOSES THE POSSIBLE DISEASE

By the classification of dataset, we can do same thing to diagnoses the possible disease as the suggestion of medication. In the below image there is a Clinician person who is identify the patient disease by the asking his current situation, As the same thing in our system that will ask the patient condition and will tell the possible disease from the data storage.



Our system will take the information from the patient and it will check in the data storage by the advanced software processing and algorithms.

VIII. COMPARISON

Be Alert with blood disease is slightly similar to other research. Like to predict the heart attack, maintaining blood pressure like smart pills. Alarming messages are sent on smartphones, predicting different diseases which could be possible by the blood. However we will work on the suggestions of medicine to maintain the heart vitals..

IX. CONCLUSION

In conclusion, we can describe IoT and data science to be the most advanced approach that will continue to advance and develop requiring latest and unconventional forms of Software Engineering, Systems Engineering, Project Management, as well as innumerable multiple disciplines to grow it further and maintain in the coming times.

Its troublesome to physically decide the chances of hypertension/heart disease based on chance components. However, machine learning techniques with the help of IoT is quite helpful to predict the outcomes from existing dataset. Through Data Science and IoT, it is possible to have remote patient care and a heart diagnostic system.

This has gotten to be indeed more apparent, as distinctive governments around the world have appeared intrigued by the idea of IoT by providing more financial assistance within the field that's implied to encourage assist inquire about and advance within the field.

REFERENCES

[1] Task Force of the European Society of Cardiology. (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use. Eur Heart.

[2] Shih-Chang Huang, Hong-Yi Chang, Yu-Chen Jhu, Guan-You Chen. The intelligent pillbox – design and implementation ICCE-Taiwan (May 2014).

[3] Home health hub internet of things (H 3 IoT): an architectural framework for monitoring the health of elderly people Sci Eng Manag Res (2014)

[4] R.J. Rosati Evaluation of remote monitoring in-home health care

[5] 2009 international conference on eHealth, telemedicine, and social medicine (2009)

[6] Agostini JV. Measuring drug burden: A step forward. *Archives of Internal Medicine.* 2007;167:753–754. [PubMed] [Google Scholar]

[7] Cranney A, Wells G, Willan A, Griffith L, Zytaruk N, Robinson V, et al. Meta-analysis of alendronate for the treatment of postmenopausal women. *Endocrine Reviews.* 2002;23:508–516

[8] https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-4-10

[9] https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-4-10

[10] Palatini P, Benetos A, Julius S. Impact of increased heart rate onclinical outcomes in hypertension: implications for antihypertensivedrug therapy. Drugs 2006;66:133– 44

[11] Nagatsu M, Spinale FG, Koide M, et al. Bradycardia and the role of-blockade in the amelioration of left ventricular dysfunction. Circu-lation 2000;101:653–9

[12] Andrews TC, Fenton T, Toyosaki N, et al. Subsets of ambulatorymyocardial ischemia based on heart rate activity. Circadian distributionand response to anti-ischemic medication. The Angina and SilentIschemia Study Group (ASIS). Circulation 1993;88:92–100.

[13] Traub O, Berk BC. Laminar shear stress: mechanisms by whichendothelial cells transduce an atheroprotective force. ArteriosclerThromb Vasc Biol 1998;18:677– 85.

[14] Black A, Murray LJ, Cardwell C, Davey Smith G, McCarron P.Secular trends in heart rate in young adults, 1949 to 2004: analyses ofcross-sectional studies. Heart 2006;92:468 –73.

[15] Heidland UE, Strauer BE. Left ventricular muscle mass and elevatedheart rate are associated with coronary plaque disruption. Circulation2001;104:1477– 82.

[16] Lechat P, Hulot J-S, Escolano S, et al. Heart rate and cardiac rhythmrelationships with bisoprolol benefit in chronic heart failure in CIBISII trial. Circulation 2001;103:1428 –33.

[17] Simpson E, Beck C, Richard H, Eisenberg MJ, Pilote L. Drugprescriptions after acute myocardial infarction: dosage, compliance,and persistence. Am Heart J 2003;145:438 – 44.