THINK INDIA (Quarterly Journal)

ISSN: 0971-1260
Vol-22-Issue-2-April-June-2019

Predicting Heart Disease using an Intelligent Naive Bayesian Classifier in IoT Architecture

V.Narmadha,M.Tech (*Assistant Professor*)
Department of Information Technology
Sri Sairam Engineering College
Chennai, India
narmadha.it@sairam.edu.in

M.Priyadharshini *(Student)*
Department of Information Technology
Sri Sairam Engineering College
Chennai, India
dharshinimkumaran@gmail.com

Abstract—
There are enormous amount of IoT devices were developed in the medical field. IoT devices plays a requisite role in monitoring the health conditions of a patient. For medical practitioners, it is very important to monitor the physiological conditions of patient. The early prediction of unusual heart conditions is very essential to discover heart problems. This paper proposes a wearable sensor device for earlier prediction of heart diseases. This scheme involves measurement and evaluation of likelihood of heart disease in a patient. We use NAÏVE BAYESIAN CLASSIFICATION technique to detect the heart diseases before occurring. Naïve Bayesian classifier is a easy-to-do and influential algorithm for the classification task, based on bayes' theorem. The proposing IoT related health tracking scheme consists of three phases, namely phase 1 data acquisition, phase 2 data storage, and phase 3 data analytics. Once, the patient's physiological data is collected through wearable sensor device, the data will be transferred to Amazon S3 (storage device) using s3cmd method, then the stored information will be transferred to Apache HBase (Hadoop distributed file system) by using Apache Pig(platform for analysing data). The collected data will be classified based on NAÏVE BAYESIAN CLASSIFICATION technique. Apache Mahout is used for implementing NAÏVE BAYESIAN prediction method. After applying this technique, if the likelihood of getting heart diseases is higher,then the information will be immediately transferred to the doctor and the doctor will provide emergency services to the patient. So that, we can prevent the patient from heart diseases.

Keywords— Internet of Things,wearable sensor devices,heart diseases,emergency services,naïve bayesian classification.

I. INTRODUCTION

Heart is an important organ of the human body. If the blood circulation to the body is insufficient, the organs of the body (i.e.,brain and heart) stop working and death occurs in few minutes. Now a days, heart diseases are increased due to some factors such as food habit, stress, genetic problems. There are various parameters which are used by the doctors for the prediction of heart sickness.

*A.*INTERNET OF THINGS:

The Internet of Things(IoT) is a network in which many objects are interconnected and these objects exchange information through this network. The various use cases of Internet of Things includes smart cities, smart environment, smart homes, etc. A vast ranging IoT environment is proceeding to support the procedure of connecting real world objects. In medical industry, there are many IoT initiatives has revolved around the enhancement of care. The IoT has the capacity to make large impact in workflow optimization and other forms of healthcare.

*B.*IOT BASED HEALTH TRACKING SYSTEM:

If the sensor detects any abnormal changes in patient's body, the automatically alerts the medical practitioners about the patient's condition over IoT. The IoT based health monitoring system effectively makes use of internet to monitor patient's heart conditions and save lives from heart sickness. Detecting the heart diseases earlier can save many lives. For detecting the heart diseases earlier, the wearable sensor device is fixed to the human body which collects the information about the patient for each second and those information will be stored in the cloud database, if the heart condition exceeds the particular threshold value ,the information will be transferred to the concerned medical practitioner then the medical practitioner will provide emergency services to prevent the patient from heart related diseases.

II. RELATED WORK

*A.*Using logical regression model, priyan malarvizhi kumar and usha devi Gandhi(2017) have proposed a novel three-tier architecture for early detection of heart diseases. They used the concept of continuous health monitoring system. In this system, the information about a set of patients is already been stored in the cloud database, doctors can retrieve data from these these database and they check the possibility of getting heart disease is within the particular value or not. If it is higher than the threshold value then immediate action will be taken by the doctors[1].

ISSN: 0971-1260
Vol-22-Issue-2-April-June-2019

*B.*SomayehNazari,MohammadFallah,Hamed Kazemipoor,Amir Salehipour (2018) proposed a fuzzy inference and fuzzy analytical hierarchy process based clinical decision support system system for diagnosing heart diseases. They have developed the clinical decision support expert system for patient who have highest possibility of heart problems. It uses expert system and fuzzy logic,which is useful for diagnosing occurrence of heart diseases. They retrieves information from the datasets stored in the database[2].

C. The Clinical Decision Support System for predicting the risk level of heart problems using weighted fuzzy rules, developed by P.K.Anooj(2011). The system automatically obtains information from the patient's medical data. They are using mining techniques,attribute selection and weightage techniques to obtain fuzzy rules. The process is carried out using datasets from UCI repository[4].

D. Mehrbakhsh Nilashi, Othman bin Ibrahim, Hossein Ahmadi, Leila Shahmoradi(2017) proposed an analytical method for heart sickness prediction using Classification And Regression Technique. Initially, EM algorithm is used for data clustering and for reducing multicollinearity in datasets they used PCA algorithm and obtained fuzzy rules by using CART techniques. This paper proposes an health monitoring system using Classification And Regression Technique[3].

E. Using IoT based m-healthcare tracking system,Prabal Verma,Sandeep K.Sood, has proposed the Cloud-centric IoT based disease diagnosis health framework. This system updates the health details. Based on the details collected, the system process health measurement. The results are generated based on accuracy, sensitivity, F-measure and specificity[6].

III. PROPOSED METHODOLOGY

The proposed IoT based health tracking system consists of three phases to collect and classify wearable sensor data. In phase 1, the physiological data is collected through wearable sensor devices and those information will be transferred to cloud database. This transformation of information is done in phase 2, where Amazon S3 and Apache HBase are used. In phase 3, Naïve Bayesian Classification technique is used to classify and determine the probability of getting heart diseases.

A. Phase 1:

Data Acquisition:

Using the wearable sensors , we collect the data and then the collected data must be converted into the format essential for the prediction of heart disease. The collected information will be classified based on factors such as respiratory_rate,heart rate, blood_pressure,systolic_range, diastolic_range,body temperature, blood_sugar. The gathered information needs to be analyzed and transformed for further steps.

B. Phase 2:

Data Storage:

The IoT devices have the efficiency of projecting clinical data uninterruptedly,to store and access such data we use the classification technique called Naïve Bayesian technique. The cloud communication technologies are efficient in terms of scalability and elasticity. Apache HBase is an important component for storing data in a distributed pattern. The data collected from the patient is carried to AmazonS3 which is the storage service.the transformation of information from AmazonS3 to Apache HBase is done by ApachePig. For extraction,loading and transformation of huge structured,unstructured,semi-structured data,the ApachePig is used.

C. Phase 3:

Data analytics:

Data analytics part is used for the enhancement of prediction model using naïve bayesian classification technique. Data analytics process is used to test data with varying data properties to provide meaningful results. There are various types of data analytics in IoT such as STREAMING ANALYTICS(real time data streams are examined to predict urgent situations and immediate actions), SPATIAL ANALYTICS(analyses geographic patterns to obtain the spatial relationship between the physical objects), TIME SERIES ANALYTICS(time based analysis for trends and patterns) and PRESCRIPTIVE ANALYTICS(combination of descriptive and predictive analysis)[4]. This system uses STREAMING ANALYTICS since we are transferring real time data streams.The proposing scheme uses machine learning libraries for developing the naïve Bayesian classification, methods for the earlier prediction of heart disease.

NAÏVE BAYESIAN CLASSIFICATION :

When we apply the principle Bayesian inference to a classification problem, preceding evidence and new evidence are summed up to calculate the likelihood of an occurrence of heart disease belonging to a certain class(posterior likelihood). This can be given by,

$$\text{Posterior} = \frac{preceding \times likelihood}{evidence} \quad \frac{preceding \times likelihood}{evidence}$$

The evidence is independent of the class. Therefore,we can express the proximate for the posterior P of an occurrence belonging to a certain class C by the equation,

$$P_j = p(C_j|X_1, \dots, X_n) \propto p(C_j)p(X_1, \dots, X_n|C_j) = p(C_j, X_1, \dots, X_n)$$

ISSN: 0971-1260
Vol-22-Issue-2-April-June-2019

$$= p(C_j)\prod_{i=1}^{n} p(X_i|C_j)$$

We express the posterior of each occurrence belonging to a certain class C. The ultimate classification of an occurrence can be carried out by a maximization function of the posteriors P to select the type with the largest posterior

Likelihoods can be very useful to assess the probability of occurrence of heart diseases. It can be normalized for each cluster j.

$$Likelihood_j = p(X_1, \dots X_n|C_j) = p(C_j)\prod_{i=1}^{n} p(X_i|C_j)$$

The normalized likelihood for a cluster acknowledges the existence of other clusters while neglecting the relative abundance of that of other clusters(p(C) and prior in eqn 1& 2)

$$Normalized\ likelihood_j = \frac{\prod_{i=1}^{n} p(X_i|C_j)}{\sum_{j=1}^{m}\prod_{i=1}^{n} p(X_i|C_j)}$$

Where,

p(C_jC_j)=probability of occurrence of heart disease belonging to class C_jC_j

p(X_1X_1 ,… X_nX_n)=probability of occurrence of heart disease provided characteristics X_1X_1 ,… X_nX_n

i,j=no. of clinical parameters

The most significant parameters such as Respiratory rate, Blood pressure-Systolic range, Blood pressure-Diastolic range are identified with the help of ROC analysis, whereas Body temperature, Blood sugar-Fasting, Blood sugar-Postmeal parameters are non significant parameters for predicting heart illness.

There are certain metrics for monitoring patient's health condition,

R1=respiratory rate
R2=heart rate
R3=blood pressure-systolic range
R4=blood pressure-diastolic range
R5=body temperature
R6= blood sugar-fasting
R7=blood sugar- post meal

if(Age == New born)
{
R1 == 30-50 &&R2 == 100-160&& R3 == 75-100&& R4 == 50-70&&R5 == 36.6 -37&&R6 == 70-100&&R7 == 70-140 Send all parameter values to Amazon S3storage. }

elseif(Age == 0-5 months)
{
R1 == 25-40 &&R2 == 90-150&&R3 == 75-100&&R4== 50-70&&R5 == 36.6 -37&&R6 == 70-100&&R7== 70-140 Send all parameter values to Amazon S3storage.
}
elseif(Age == 6-12 months)
{
R1 == 20-30 &&R2 == 80-140 &&R3 == 75-100&&R4== 50-70&&R5 == 36.6 -37&&R6 == 70-100&&R7 == 70-140 Send all parameter values to Amazon S3storage. }

elseif(Age == 1-3 years)
{
R1 == 20-30 &&R2 == 80-130&&R3 == 80-110&&R4== 50-80&&R5 == 36.6 -37&&R6 == 70-100&&R7 == 70-140 Send all parameter values to Amazon S3storage. }

elseif(Age == 3-5 years)
{
R1 == 20-30 &&R2 == 80-120&&R3 == 80-110&&R4== 50-80&&R5 == 36.6 -37&&R6 == 70-100&& == 70-140 Send all parameter values to Amazon S3storage. }

elseif(Age == 6-10 years)
{
R1 == 15-30 &&R2 == 70-110&&R3 == 85-120&&R4== 55-80&&R5 == 36.6 -37&&R6 == 70-100&&R7 == 70-140 Send all parameter values to Amazon S3storage. }

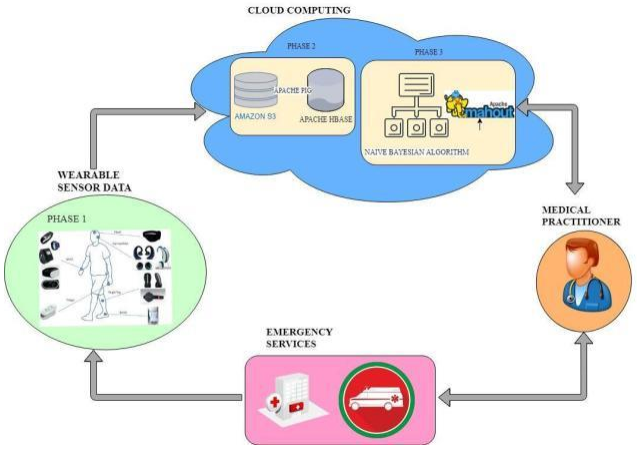
elseif(Age == 11-34 years)
{
R1 == 12-20 &&R2 == 60-105&&R3 == 95-140&&R4== 60-90&&R5 == 36.6 -37&&R6 == 70-100&&R7 == 70-140 Send all parameter values to Amazon S3storage. }

elseif(Age == 35-100 years)
{
R1 == 12-30&&R2 == 60-100&&R3 == 95-140&&R4== 60-90&&R5 == 36.6 -37&&R6 == 70-100&&R7 == 70-140 send all the parameter values to Amazon S3 storage. }

else
{
If clinical parameters is abnormal, send these parameters to doctors as well as Amazon S3 storage }

ARCHITECTURAL DIAGRAM

ISSN: 0971-1260
Vol-22-Issue-2-April-June-2019



IV. CONCLUSION

This paper proposes the Naïve Bayesian Classification technique based earlier prediction of heart diseases with the objective of evaluating the possibility of getting heart sickness. The main aim is to identify the key factors and features from the medical data of the patient to predict the heart problems, before it causes to serious issues. This paper comprises the evolution of framework based on NAÏVE BAYESIAN CLASSIFICATION technique for early detection of heart problems. The medical practitioners can help the patients by predicting the heart diseases before occurring. The future enhancement is to increase the accuracy of naïve Bayesian classification model.

REFERENCES

[1] Priyan Malarvizhi Kumar,Usha Devi Gandhi, "A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases",Computers & Electrical Engineering,2017,vol.65,pp.222-235.

[2] Somayeh Nazari , Mohammad Fallah , Hamed Kazemipoor , Amir Salehipour,"A Fuzzy Inference- Fuzzy Analytic Hierarchy Process-Based Clinical Decision Support System for Diagnosis of Heart Diseases",Expert Systems With Application(2018),vol.95,pp.261-271.

[3] Mehrbakhsh Nilashi,Othman bin Ibrahim,Hossein Ahmadi,Leila Shahmoradi,"An analytical method for diseases prediction using machine learning techniques",Computers & Chemical Engineering(2017),vol.106,pp.212-223.

[4] P.K.Anooj,"Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules",Journal of King Saud University-Computer and information sciences(2011),vol.24,pp.27-40.

[5] Wim Paas,Jeroen C J Groot,"Creating adaptive farm typologies using Naïve Bayesian classification",Information Processing in Agriculture(2017),vol.4,pp.220-227.

[6] Prabal Verma,Sandeep K.Sood,"Cloud-centric IoT based disease diagnosis healthcare framework",Journal of Parallel and Distributed Computing(2017).

[7] T.Vivekanandan,N Ch Sriman Narayana Iyengar,"Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease",Computers in Biology and Medicine(2017),vol.90,pp.125-136.

[8] Anum Saeed,Vijay Nambi,"Short-term global cardiovascular disease risk prediction in older adults",Journal of the American College of Cardiology(2018).

[9] Purushottam,Kanak Saxena,Richa sharma,"Efficient heart disease prediction system",Procedia Computer Science(2016),vol.85,pp.962-969

[10] Andrea driscoll,Elizabeth H Bames,"Predictors of incident heart failura in patients after an acute coronary syndrome: The LIPID heart failure risk prediction model",International journal of cardiology(2017),vol.248,pp.361-368.