IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PREDICTION AND DIAGNOSIS OF CARDIOVASCULAR DISEASE USING CLOUD AND MACHINE LEARNING DESIGN

Dr.K.VENKATASALAM, Mr.S.T.LENIN

Associate Professor, Department of Computer Science and Engineering, Mahendra Engineering College(Autonomous), Mallasamudram, Namakkal-637503.

Assistant Professor, Department of Information Technology, Mahendra Engineering College(Autonomous), Mallasamudram, Namakkal-637503.

ABSTRACT

In medicine, predicting and accurately diagnosing heart disease is a huge problem, and cardiovascular disease predetermine in health services is considered an important problem. In these growing health-care organizations, more expensive surgeries are offered to patients. Recently, heart disease has become a common disease, that is, even though medicine is growing on one side, cardiovascular disease are growing exponentially on other side. The main cause of these diseases is poor lifestyle, alcohol consumption, lack of physical activity and tobacco consumption. Thus, there is a need for a cloud-based framework (CBF) for monitoring health information and predicting it efficiently. Recently, machine learning methods have been used to solve these types of problems. But in this proposed system, to improve the process of predicting patients' health information and cloud-based four steps are also used to improve monitoring. So here are two types of methods used in machine learning to detect and classify heart disease. Then the accuracy of those methods is examined. Evaluation criteria are used to examine their effectiveness.

Keywords: Heart Disease, Prediction, Classification, CBF, Machine Learning Algorithms (MLA).

A. INTRODUCTION

Nowadays people's lifestyle is increasing day by day due to inheritance. This creates not only time but also a lot of data, thus wasting data on health surveys. But nowadays data analysis has come into play. The project was developed to create a large amount of data from hospitals to be effective. Heart disease is a deadly enemy during this period. The disease can be cured as much as possible otherwise it affects that particular

person. And the right time is to diagnose patients, and it is a very difficult job. Hospital tests can sometimes be misdiagnosed, leading to bad names. The disease is very difficult to treat and most patients are unable to treat it.

In the paper [1], two algorithms in machine learning have been used to predict heart disease. In paper [2], machine learning is used to predict the presence of heart disease. In paper [3], a hybrid machine learning method has been used to diagnose heart disease. In paper [6], data mining technique has been used to diagnose and monitor heart disease. In paper [5], the MLA system has been used to diagnose and predict atherosclerosis.

B. PROPOSED METHODOLOGY

a) Data-set

The final attributes are selected first for pre-processing. This data set contains the information of 300 people. And of these there are fifteen columns. It is explained below.

1. ***Age***:

   The age of each person is indicated here.

2. ***Sex***:

   Of these, gender is defined using the forms given below.

   1 = female

   0 = male

3. ***Chest-pain type (CPT)***:

   It describes the chest pain that people experience using the following guidelines.

   1 = asymptotic

   2 = non-angina pain

   3 = atypical angina

   4 = typical angina

4. ***Resting Blood Pressure (RBP)***:

   The unit of mm-Hg is shows the RBP value of an individual.

5. ***Serum Cholesterol (SC)***:

   The unit mg/dl shows SC.

6. ***Fasting Blood Sugar (FBS)***: The FBS value of a human is compared to that of 120mg/dl.

if

FBS > 120mg/dl

then:1(true)

else

0 (false)

7. ***Resting ECG (RECG)***: This shows RECG results,

0 means normal

1 means it having ST-T wave abnormality

2 means left ventricular hypertrophy

8. ***Max heart rate (MHR):*** It shows the MHR reached by a human.

9. ***Exercise induced angina (EIA):***

1 means yes

0 means no

10. ***ST depression (STD)***: It shows the value that is a float or an integer.

Table 1: Data-set

Index	Age	Sex	CPT	RBP	SC	FBS	RECG	MHR	EIA	*STD*
1	64	0	2	140	234	0	0	151	1	2.4
2	68	0	3	165	287	1	0	109	0	1.6
3	68	0	3	125	230	1	0	130	0	2.7
4	38	0	4	120	251	1	2	188	1	3.
5	42	1	1	120	205	1	0	173	1	1.5
6	57	0	1	110	237	1	2	179	1	0.9
7	63	1	3	150	269	1	0	161	1	3.7
8	58	1	3	130	355	1	2	164	0	0.7
9	64	0	3	140	255	1	0	148	1	1.5
10	64	0	3	130	204	0	0	156	0	3.2

b) Architecture

In this proposed method, machine learning systems, including cloud, are designed here. Here are four steps designed to monitor and predict the incidence of heart disease. Devices and monitoring sensors are used to monitor people's health. So in step 1, that surveillance collects data from sensors and devices and combines them.In the second step, as the data is monitored and collected daily, the data is getting bigger. A huge amount

of device is needed to collect all this. It uses the cloud server to store large an amount of real-time information. Step 3 is developing a real-time based design for early detection of heart disease. It uses the MLA necessary for prediction and classification to create them. Step 4 sends the end result to the users. Here is using an application to do so.The structure of these four steps is given in Fig. 1.

Requests from the monitoring system are accepted and stored by the cloud server. If the new data is equal to pre trained data, that data is stored in cloud server. If the new data is or higher than or lower than or not equal to the pre trained data, that data go to as notification to user.In this paper, for heart disease, to come up with a better solution, provides MLA, cloud servers, NoSQL database and real-time predictive service.

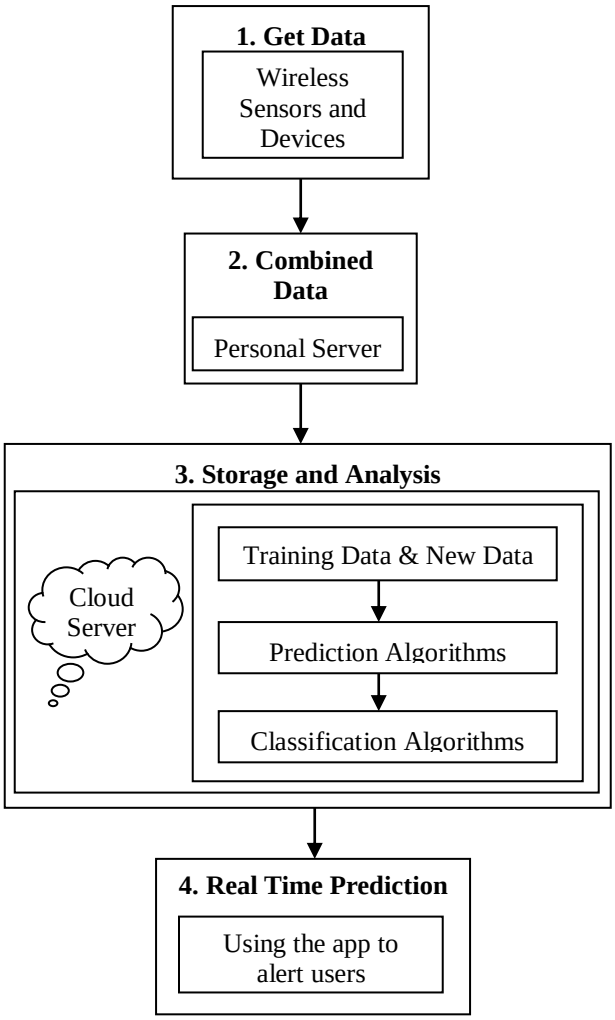

Fig.1: Proposed Architecture

The main contributions of this paper are mentioned in the following points:

- Using machine learning, cloud-based systems are recommended for monitoring and predicting heart disease. This system helps to cardiologists make good, effective decisions.
- From various health monitoring services, to handle big data like real-time data and health services data, this paper deals with many excellent systems.If this paper does not handle this large amount of information, some of the important features of that data will be missed.

- Most analyzes do not use real-time settings. But this paper just uses it. And some studies are based on accuracy only. But this paper is all about accuracy and real-time system.

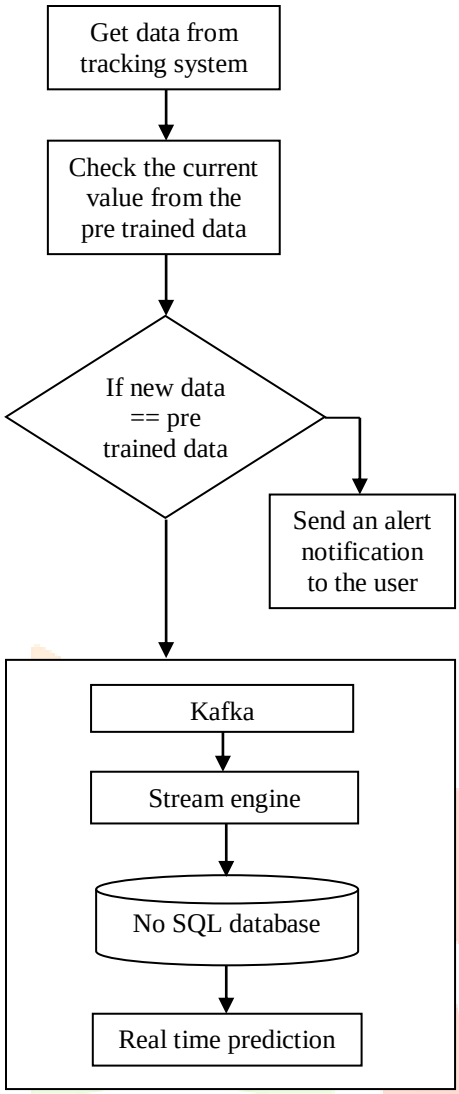

Fig.1.2: Flow diagram

There are four steps in this proposed method, they are, respectively,

1. Data collection
2. Data storage
3. Analysis module
4. Application presentation

The data collection step is used to extract data from a specific patient using devices and sensors. To continuously collect the health data of a particular person, health monitoring devices are integrated with the human body. Moreover, monitoring systems are sending data seamlessly. To this vast amount of data, by using traditional database techniques and tools, making their analysis and storage difficult.This proposed system uses

NoSQL data base and cloud computing technologies to continuously collect health data. And in the app step, users can view their usage reports using the mobile app.

C. Prediction Method using K-Nearest Neighbors (K-NN) Algorithm

The KNN system is a supervised machine learning system. It is used to solve classification and regression problems. The KNN mechanism is based on actual observations in nearest.

Let D_S = {(a_1, b1), (a_2, b_2)... (a_N, b_N)} A training data set (observations set) of q-dimensional patterns with A = $[a]_{j=1}^{N} \subset D^q$, B =$[b]_{j=1}^{N} \subset D^q$ as a corresponding labels set, and N is the no. of training entities. In the data space, Minkowski metric is used to define a homogeneity function.

$$\|a' - a_i\|^p = \left\{\sum_{j=1}^{q} \left|(a_j)' - (a_j)_i\right|^p\right\}^{\frac{1}{p}} \quad (1)$$

Pretty much the distance is said to be the Euclidean distance.This is used to determine the distance between the query and the data set points.In multi-class classification mode, to unknown Model a0, in the data space, the KNN system predicts the majority class label near K.

$$F_{KNN}(x') = \arg \max_{b \in N_k(a')} \sum_{i \in N_k(a')} L(b_j = b) \quad (2)$$

Where, the L () is indicator function. L (b_j = b) = If argument is true, the value is 1 and otherwise 0.The choice of the K value will always be odd numbers.

D. Classification Method using K-means clustering algorithm

Hard classification called KMC consists on a partitioning N training dataset or observations dataset A of n attribute vectors into C classes.The purpose of the classification algorithm is to find class centroids that reduce objective function. The objective function is given by the equation given below:

$$I = \sum_{j=1}^{C} \left\{\sum_{k, x_k \in v_j} \|a_k - v_j\|^2\right\} \quad (3)$$

Where, v_j is the centroid of the j^{th} class, d (a_k, v_j) is the distance between j^{th} centroid C_j and the k^{th} data of A. Generally, theEuclidean distance is utilized to define the objective process.

Here the data must belong to only one denominator. Therefore, the membership team M has two characteristics. Their equation is given below.

$$\sum_{j=1}^{c} a_{jk} \quad (4)$$

Where, ∀k = 1, . . . , N

$$\sum_{j=1}^{C} \sum_{K-1}^{N} a_{jk} = N \quad (5)$$

The value v_j of every and each class centroid is calculated by the mean of all its attribute vectors:

$$v_j = \frac{1}{|C_j|} \sum_{k, a_k \in v_j} a_k \quad (6)$$

Where, $|C_j|$ = the cardinal or the size of C_j.

C. RESULTS AND DISCUSSION

Several methods are used for the assessment of atherosclerosis. All these methods are sensitivity (S_E). It quantifies the amount of patients who are most accurately identified as having the disease.

The specificity (S_P) is a computation of the number of patients correctly identified as having no disease. Accuracy (A) depends on the operation of the algorithms used. MCC (Matthews's correlation coefficient) is a measure of the strength used in binary classifications machine learning.

In machine learning, the CM is also called the error matrix. Refer to Table 2 to see the effectiveness of the algorithms. This matrix has 2 information types. That is the efficiency of the predicted algorithm and the actual performance of the algorithms.

Table 2: CM for Binary Classification

Predicted Output (PO)	**Actual Outputs (AO)**	
	Patient has not disease	*Patient has disease*
Negative Test (NT)	TN	FN
Positive Test (PT)	FP	TP

The overall efficiencies were presented using equations. The TP (True Positive) is that the patient correctly diagnosed a condition called atherosclerosis. The FP (False positive) is a misdiagnosed prediction of disease. The TN (True Negative) is the prediction that the patient is healthy and fairly accurate. The FN (False negative) is a misdiagnosed prediction that the patient is healthy.

$$S_P = \frac{TN}{TN+FN} \quad (7)$$

$$S_E = \frac{TP}{FN+TP} \quad (8)$$

$$A = \frac{FN+TP}{FP+TP+TN+FN} \quad (9)$$

$$MCC = \frac{TN.TP-FN.FP}{\sqrt{(FP+TP).(FN+TP).(FP+TN).(FN+TN)}} \quad (10)$$

The Clinical Diagnostic Support System for atherosclerosis diagnosis (AD) is provided using the KNN and KMC Methods. There are two types of AD classes. One is healthy and the other is atherosclerosis. There are 280 samples in the NoSQL database. Seventy of percentage is used in training data. And fifteen percent of validation data were used. The remaining data have been used as test data. Two types of machine learning have been used for atherosclerosis. The results of the prediction and classification of the proposed system are built using the Confusion Matrix (CM) in Table 3. Each CM cell contains assorted source number of actual and predicted outputs.

Table 3: CM of two MLA

	FN	TN	FP	TP
KNN	6	43	5	29
KMC	9	26	22	26

The performance results and evaluation of the AD presented in table 4. The results have been detected utilizing the KNN, and KMC algorithms. The performance results gave that the proposed system had a sensitivity of 98%, an accuracy of 97%, and specificity of 97% as the good rates.

Table 4: Performance of MLA

	KNN	KMC
MCC (%)	0.73	0.27
A (%)	87	62
S_E (%)	84	74
S_P (%)	91	53

On the other hand, checking the MCC would be the proposed system performance indicator. In this process, the value of MCC is 0.94 utilizing equation (10). If the result is one, the proposed system has a good and correct prediction. The accuracy of the proposed system is arranged in the order of increasing the rate.

D. CONCLUSION

In this paper, two supervised learning-based machine learning techniques are described. Then the efficacy of the two classifiers used in the prediction of heart disease was compared. Then their performance is estimated using ten times the cross-sectional and confusion matrices. This paper used machine learning-based cloud architecture to predict and diagnose heart disease. Therefore, a real-time, four-step cloud system has been proposed to monitor and predict heart disease. This test used two machine learning techniques to predict heart disease based on several parameters. In addition, these cloud-based applications were able to detect heart disease by collecting data from the health center. This includes using the cloud system to detect cardiovascular diseases and send warning messages to patients. This system can therefore be used to predict and monitor heart disease.

REFERENCES

1. Santhana Krishnan J, Geetha S, “Prediction of Heart Disease Using Machine Learning Algorithms”, 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 20 June 2019.
2. Rahma Atallah, Amjed Al-Mousa, “Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method”, 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), 05 December 2019.
3. Senthilkumar Mohan, Chandrasegar Thirumalai, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques”, IEEE Access (Volume: 7), 19 June 2019.

4. Mamatha Alex P, Shaicy P. Shaji, "Predictionand Diagnosis of Heart Disease Patients using Data Mining Technique", 2019 International Conference on Communication and Signal Processing (ICCSP), 25 April 2019.
5. Oumaima Terrada, Bouchaib Cherradi, "Classification and Prediction of atherosclerosis diseases using machine learning algorithms", 2019 5th International Conference on Optimization and Applications (ICOA), 03 June 2019.
6. Divya Krishnani, Anjali Kumari, "Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms", TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), 12 December 2019.
7. Kumar G Dinesh, K Arumugaraj, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms", 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), 29 November 2018.
8. Kun Chen, Akrit Mudvari, "Heart Murmurs Clustering Using Machine Learning", 2018 14th IEEE International Conference on Signal Processing (ICSP), 28 February 2019.
9. Sanchayita Dhar, Krishna Roy, "A Hybrid Machine Learning Approach for Prediction of Heart Diseases", 2018 4th International Conference on Computing Communication and Automation (ICCCA), 29 July 2019.
10. Sneha Borkar ; M. N. Annadate, "Supervised Machine Learning Algorithm for Detection of Cardiac Disorders", 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 25 April 2019.