Decision Tree based predictive Analytics for detecting Pregnancy Induced Hyper Tension

Ranjith Ramesh
Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA
rramesh3@jhu.edu

ABSTRACT- The major disadvantage of the present medical system is that it seriously lags in continuous evaluation of a patient‘s record. There is a big gap between the way medication is done and what the body really wants. This has a flabbergasting effect in every human body. The major disadvantage in the present system is the high difficulty in deciphering the patient‘s ailments. Hence it results in delayed diagnosis of the actual problem. This difficulty can be avoided and help can be provided to fill this communication gap with the help of CNC (Computing, Networking and Communications). Wireless Sensor networks can be used to monitor the vital sciences Medication and determination of a patient‘s ailment revolves around major vital science (pulse, respiration, blood pressure and temperature). Watching the vital signs return to normal is often an important variable that signifies the patient is improving. Wearable or implanted sensors can be used to continually collect a person‘s details. These sensors can be attached permanently in our body. For example temperature sensors can be placed in our ornaments we use. The information collected by the sensor networks can be routed to the cloud platform in which we create applications to monitor any abnormal rate in the vital sciences. If any discrepancy is detected the person can be intimated. The data stored can also be used during diagnosis. In this system the persons vital sciences are continually monitored and hence even a small change can be easily detected and any ailments can be cured at the initial stage

Keywords – **Decision tree, predictive Analytics , Wearable Sensors , Pregnancy Induced Hypertension.**

I. INTRODUCTION

There is a very big gap between the way the doctors diagnose and what really the body needs. In order to conform their diagnosis doctors use many tools like scan report, x-ray report etc. Even after using so many confirming tools the accuracy of their predictions lag far behind their need.This is due to lack of monitoring the patient‘s body at regular interval. The doctor may not know the pulse rate of a patient a few hours before he or she came to the hospital.Such a kind of continuous monitoring systems can pave way to solve many crucial diseases like cancer and alzeamir.

The major back lock in current medication system is that we are not able to continuously monitor a person‘s health. Added to this many are not good in communicating the problem they have to the doctors. Even the doctors make assessment based on the current state of the vital sciences which leads to improper medication. The main objective of this research is to bridge the gap between doctor-patient communications by continually monitoring a person‘s vital sciences. Wireless sensors are fitted in every person‘s body even though if the person is not suffering from any ailment because the main aim is to continually monitor the persons basic vital sciences .These sensors collect the person‘s data like pulse, respiration, blood pressure and temperature. The collected data are regularly stored in cloud data base. If the person gets affected with any ailment there will be changes in the vital sciences. The application created in cloud monitors these changes and intimates the same to the person.

As a result any problem can be found out at the early stages. More over the doctors too can have a clear detail about the history of the patient‘s vital sciences and with the help of the collected data medication can be made accurately.

II. EXISTING WORK

There are lots of works going on body sensor networks especially for monitoring patient‘s activities and health condition during hospitalization. But none of the system proposes a continuous monitoring scheme to regulate and monitor a normal human‘s health. In integration of wireless sensors network with cloud computing for secured architecture [1] the information collected through the sensors are collected into a mobile device and they are compared with the threshold values stored in mobile and any abnormality detected is sent to the cloud database. The greatest drawback of this system is that most of the process will be performed in PDA and hence it needs to be active all time. Added to this the overhead and radiations emitted by PDAs are high when they are admitted to too much of processing. In

Integrated Cloud for Automated Telemedicine Based e-Health Care Applications WSN [2] the data‘s collected from patients are added to the common database in cloud which can be accessed anywhere by the doctors. Here the cloud platforms provide Iaas and does not act as a monitoring tool for scrutinizing patients health. In Phone centered Body sensor network platform [3] Bluetooth technology is used as a medium of communication between the sensor nodes and cache watch which is used as an interface. This system though it sounds very much similar to the proposed system, the algorithm used for processing the collected data in cloud or in mobile has very low precession. The idea to have a cache watch to have a user interface suffers practical back lock as it is highly difficult to have a watch tied up all the time within. More over the objective of continuous evaluation is not satisfied as the monitoring is triggered at regular interval and the sensors are put at sleep mode in between. Though it saves battery power of the sensors the basic idea of continuous monitoring is not satisfied. In ubiquitous monitoring Environment for Wearable and Implantable Sensors (UbiMon) [4] an intermediate LPU (Local processing unit) is need. This emphasizes the need for keeping the LPU active all time more over the efficiency of the LPU gets reduced if it is a multipurpose LPU.

III. PROBLEM IN EXISTING SYSTEM

The sensors in the existing system just sense certain information and analyses it in the abstract way and reports it back. There is lots of drawback in this system. The accuracy in which the sensors detect and diagnose the vital science is a big question. For example when the pulse rate increases suddenly the sensors will indicate the person that his heart is weak. But there are numerous reasons for increase in pulse rate. For example the pulse rate would have increased because of increase in thyroxin which is totally a different issue. When the sensitivity of the sensors and the threshold values of the sensors to indicate abnormality are low, then there is a great problem of miscalculation in detecting abnormalities. For example when a person runs his heart beat increases rapidly and that is indicated as abnormality in some existing system. Hence the existing system suffers rapidly with respect to computation and accuracy of data compution.

IV. PROPOSED SYSTEM

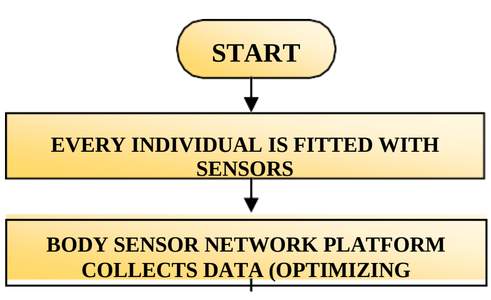

Figure 1 Flow chart of the proposed system

The proposed system follows the following sequence

- Wireless sensors are placed on the surface of the body depending upon the persons need.
- The sensor senses different vital sciences of the body.
- An optimizing algorithm is used to have an optimal use of sensors.
- The data collected from the sensors are compared with the base threshold value
- When the difference is same then the values are updated in the cloud database.
- Rather than updating the data as soon as it is collected the data is stored in a local cache and later all the data along with the abnormality are updated in the database in cloud. This kind of approach increases the efficiency of the sensors.

Process in cloud database:

- All the data collected from the sensors are stored in the database in different tales as shown.
- The data are analyzed not only based on the fixed comparison value but they are analyzed based on high optimization algorithm. This

algorithm provides high efficiency when compared with all the other existing system.

- By using the results of this algorithm many complex diseases can be found at a very early stage and diagnosis cum medication can be given to the patient.
- Monthly medical reports are given to the doctors for analysis.
- Reports can also be sent through mobile.

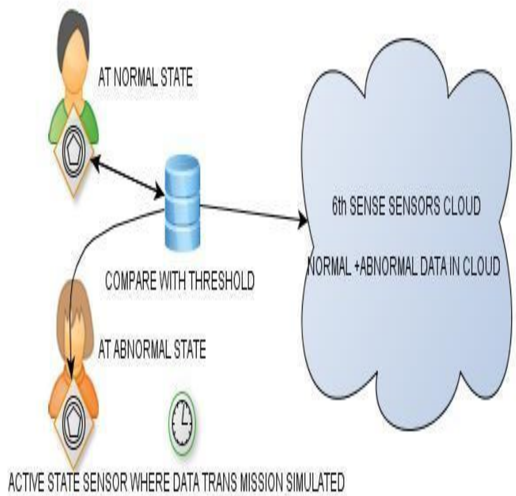

Figure 2 Optimizing Algorithm at the sensor side

Proposed-Optimizing Algorithm at the sensor side.

- The sensors are made to operate in two modes. They are passive mode and active mode.
- In passive mode the sensors just senses the vital sciences and compares it with the threshold and then stores it in the cache. This does not require more amount of energy.
- Active mode is activated when the collected data crosses above the threshold(which is kept low for accuracy) all the data in the cache are updated to the database
- The sensors are also capable of detecting the person‘s state like whether he is walking, running or doing any aerobic activity. At that time the vital sciences will rise up and those details are also not updated in the database

Proposed filter algorithm

This algorithm gives an intelligent view of the operation of the software in the cloud platform.

- When an abnormality is detected, the condition is checked with the previous conditions and a mark is made for the data.
- When there is frequent abnormalities an indication is sent to the user and the data is compared with the standard database containing the symptoms for various problems
- Depending upon the comparison result the possible problem for the users is guessed and it is ranked

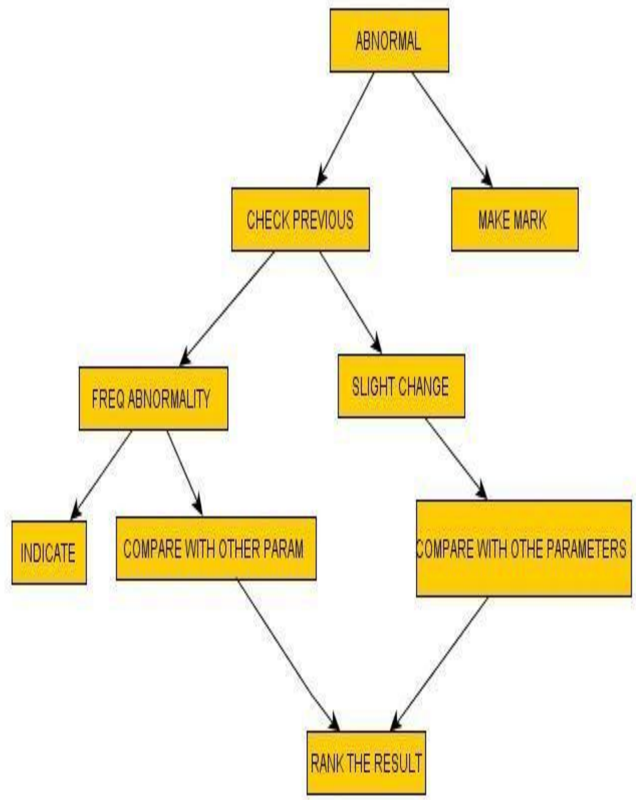

Figure 3 Dynamic Filtering Algorithm

Technical details of Implementation:

The body sensors developed should satisfy certain criteria. They should be micro-powered miniaturized and cost should be low. Earlier Body sensors were prepared by modifying the current wireless sensors, but recently numerous Body Sensor Networks platforms are developed.
These platforms can be divided in three categories as ARM based platform, Microcontroller based platform and RF integrated platform.

The ARM bases platforms require more amount of processing power. But ARM Processors has SDIO (Secured Digital Input Output) and hence enables to use Bluetooth and wireless technology.

The microcontroller based platforms are considered better in terms of power consumption as they consume less power when compared to ARM Based platform. These processors are fitted with radio transceivers based on IEEE 802.15.4. Chipcon CC2420 is one of the transceivers used.

The size of the sensors can further be reduced by using RF-microcontroller chips. For example Nrf24E1 is an RF Transceiver integrated with INTEL 8051 microprocessor core. The size is further reduced to 1.2x1 inch.

By using RF-Integrated microcontroller it is stated that they are slow due to utilization of maximum energy by wireless transmission. But this can be overcome by utilizing the proposed optimization algorithm where processing is more than the wireless transmission**.**

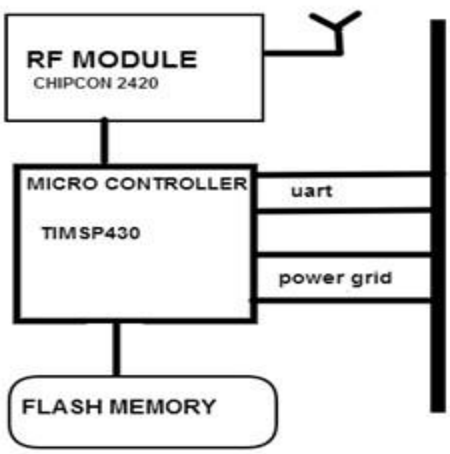

Figure 4 Body sensor Node

Figure 4 illustrates the construction of body sensor nodes. It has three blocks and they are RF module, Microcontroller and Flash Memory.

The microprocessor is a 16-bit ultra low power RISC processor with 60 KB flash memory, 2 KB RAM and 12 bit ADC. The node requires 3mW at active mode and 15 micro watts at sleep mode. For wireless communication, the Chipcon CC2420 is used in the BSN node. The CC2420 has a maximum throughput of 250kpbs with a range of over 50m. In addition, it has a built-in the AES-128 (Advanced Encryption Standard) hardware encryption/decryption and the IEEE 802.15.4 MAC (Medium Access Control) functions. With the built-in buffers and MAC, the CC2420 can act as a coprocessor to handle all the packet communications, and which significantly reduce the computational demands on the microcontroller. [5]

To facilitate the sensor data collection and enable dynamic reprogramming of the BSN node, 512KB of serial flash memory is incorporated in the BSN node. With the 512KB of memory, almost 1.5hour of ECG (100Hz) data can be stored without any compression. By applying the DPCM and LZW in series, ECG data can be compressed down to 11% of its original size, which means that the memory may be able to hold up to 13 hours of ECG data.

In terms of software, the BSN node is designed to run TinyOS by U.C. Berkeley, which is a small, open source and energy efficient sensor board operating system. It provides a set of modular software building blocks, of which designers can choose the components they require. The size of these files is typically as small as 200 bytes and thus the overall size is kept to a minimum. The operating system manages both the hardware and the wireless network—taking sensor measurements, making routing decisions, and controlling power dissipation.

By using the ultra low power TI microcontroller, the BSN node requires only 0.01mA in active mode and 1.3microA when performing computational intensive calculation like a FFT. With a size of 26mm, the BSN node is ideal for developing wireless biosensors. In addition, the stackable design of the BSN node and the available interface channels ease the integration of different sensors with the BSN node. Together with TinyOS, the BSN node can significantly cut the development cycle for wireless biosensor development.

The data collected from the sensors are sent to the cloud database using 802.11 protocols, HTTP or XML.
In the cloud software is created using many cloud tools like salesforce, windows azure etc and the above mentioned algorithms are implemented and effective results are obtained.[6]

Proximate Data Collection

Data collection is a fundamental but challenging task for Wireless Sensor Networks due to the limitations in communication bandwidth and energy budget. Since sensor nodes are often battery powered and deployed in harsh environments, data collection strategy must be carefully designed to reduce energy consumption on sensor nodes, so as to extend the network lifetime as much as possible. Proximate data collection is a wise choice for long term data collection in WSNs with limited bandwidth. There are several factors to be considered in the design of an approach for proximate data collection.

First, the data collection approach should be scalable. Second, in proximate data collection, the spatial temporal correlation model used for data suppression should be light-weight and efficient so as to meet the constraints on sensor node's memory and computation capacity. Most of the existing models are too expensive, i.e. demands large amount of computing capacity or storage Capacity, to be run on the existing sensor nodes. Third, the data collection scheme should be self-adaptive to environmental changes.

Proximate Data Collection (ADC) achieves low communication cost by using the fact that physical environments generally exhibit predictable stable state and strong temporal and spatial correlation. The sink can estimate the sensor readings according to the model parameters updated by the clutch heads.
ADC consists of following two parts: The local estimation and the data approximation. The local estimation builds a local temporal correlation model for each sensor node to estimate its local readings. The **local estimation achieves self-adaption by periodically checking the differences** between its estimation and the actual sensor readings. If the actual sensor readings consistently differ from the estimated model, the local estimation will regulate its parameters automatically.

Figure 5 describes the integration of wireless sensor network with web services. At first sensor nodes are plotted and sink node is created. Clutches are formed by using k-means clutching algorithm and the head is chosen by the node which has high energy than all other nodes. The sensor will automatically learn their value and an estimation model will be generated inside the clutches to reduce the communication cost. Clutch head will collect data from all sensor nodes and it will send to the sink node.

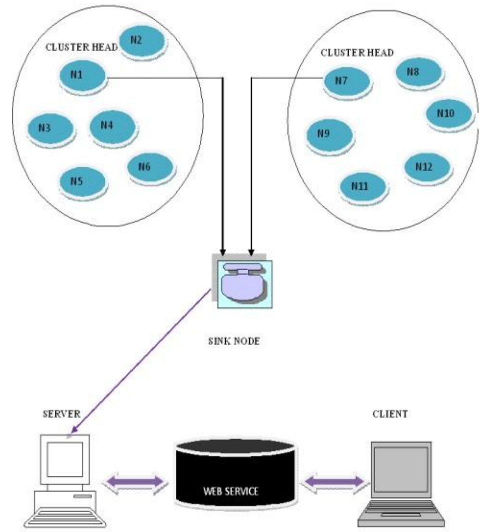

Figure. 5. Integrating WSN and Web Service

The sink node acts as a source and control the data collected by all other sensor nodes. The sink node is responsible for receiving messages from other corresponding nodes, store information & perform periodically sending all the information to the server where it is treated and inserted into the database. Web service is integrated to view the details of temperature which has been collected by the sensors.

Sensor Nodes Creation and Clutch formation

Sample sensor nodes and a Sink node (S) with different energy levels are plotted. Clutch the sensor nodes. For clutching, K indicates the usage of algorithm. Then select clutch head for all the clutchs. The node with highest energy level is chosen as clutch head. Figure 7 depicts the flow diagram of deployment of sensor nodes and clutch head selection.

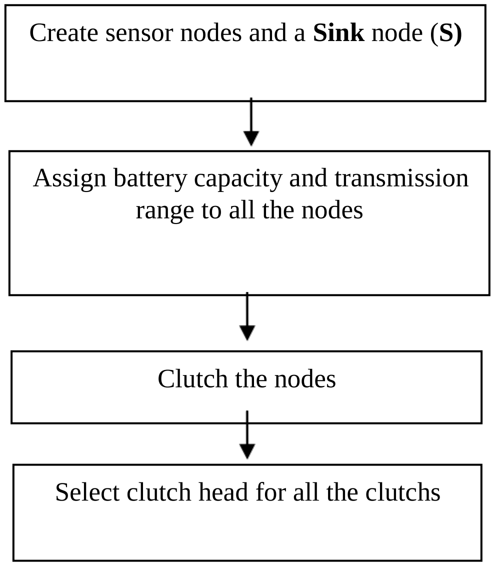

Figure. 6. Clutch head selection

K Means Clutching Algorithm

In this algorithm we determine the number of K clutch and assume the centroid or center of these clutchs. We can take any random node as the initial centroids or the first K objects in sequence can also serve as the initial centroids.

K means Clutching algorithm [7]

1. Initialize the total number of nodes required.
2. Place all the nodes at specific location.
3. Determine the centre of grouping nodes

$$De=\sum_{j=1}^{k} \sum_{i=1}^{n} \| m_i^{(j)}-c_{(i)}^{h} \|^2$$

Where

$\| m_i^{(j)}-c_{(i)}^{h} \|^2$ is the Euclidean distance

m_i data point.

$c^h_{(i)}$-clutch centre.

4. If centre is finalized (centre data point), the distance between centre point and all other node must be calculated.
5. Distance = a metric of Euclidian distance.
6. Euclidian distance should be a minimum for formation of clutch
7. Grouping nodes as clutch.
8. Determine the energy level for every node.
9. Clutch head = Higher energy level node
10. Recalculate the Euclidian distance from clutch head and other node in clutch

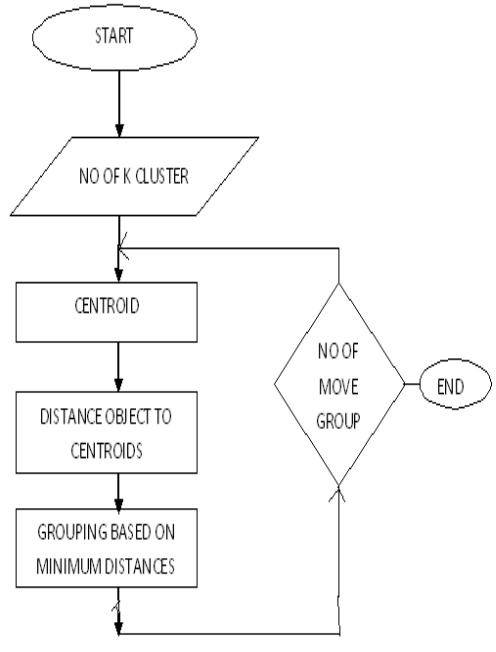

Figure. 7. Flow diagram of k means algorithm

Figure 7 shows the flow diagram of k means algorithm. The position of nodes and clutch head should be in a way that distance between node and clutch should be a minimum for attaining efficiency. This algorithm can be executed multiple times in order to mitigate the effect. The clutch formation is shown in Figure. 6.

Data model generation and parameter learning

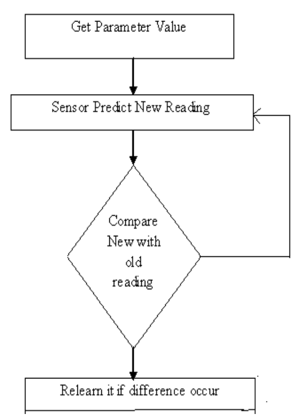

Figure. 8. Flow diagram of learning parameter

The Data Model and the parameter learning takes place in the following way.
Step1: Initially the sensor nodes are plotted in the simulation environment.
Step2: The parameters are obtained from these sensor nodes. Then the sensor nodes predict the new readings using estimation model.
Step3: Then the comparison of predicted readings with the old readings is calculated. If the difference is greater than the threshold then the nodes relearn the parameter values. The flow diagram is shown in Figure. 8.

Selection of minimum dominating sets

The clutch heads calculate the estimation distance between the nodes. Then they find ϴ similar nodes based on estimation distance. Now they form correlation graph by connecting ϴ similar nodes. The representative node for each ϴ similar set is found out. Clutch head forms their own dominating set .Then by applying greedy heuristic approach minimum dominating sets are chosen. The clutch heads send the dominating set & corresponding predictor set to the sink node.

To Determine Dominating Sets:

Step 1: Intialize all sensor nodes
Step 2: Find the slope among sensor nodes using linear prediction
$\hat{Y}$=ax+by where
a=$R_{\sigma x}^{\sigma y}$, b= $\mu_{y-} R_{\sigma x}^{\sigma y} \mu_{x}$
Step3: By using the slope, we are going to find regression for prediction value using R=
$\frac{1}{n-1} \sum_{i=1}^{n} \frac{xi-\mu x}{\sigma x} \frac{yi-\mu y}{\sigma y}$
Step 4: Then we are going to draw correlation graph by
$|p_i (t)-p_j (t)|$ where
$P_i (t)$ = Actual value
$P_j (t)$ =Expected value
Step 5: Find the mean difference between $p_i (t)$ & $p_j (t)$. Then it will form similar sets.
Step 6: Else other nodes are joined in similar sets.

Proximate data collection

Whenever the clutch heads receive updated parameters from their member nodes they will check whether that node is still ϴ similar to the set. Likewise the nodes who are no longer ϴ similar then taken out from the graph and they can be added into another ϴ similar set or new ϴ similar set is created. The clutch heads send *set update or create new set* message to the sink node. Then the sink updates the dominating and predictor sets for whole network.

Step 1: For proximate data collection the following steps are necessary
Step 2: Set the threshold value
Step 3: If the message lies within the threshold value update the message
Step 4: If the message lies outside the threshold value we can create newest
Step 5: Move the sensor to new clutchs and updation and creation of messages occurs in a cyclic manner by checking the above conditions.

Integration with web service

The data from sink is collected by a server. Then the server provides the collected data as a web service to the clients. The clients obtain the service from the server.

Expected Outcome:

- The outstanding impact of the proposed system is found in the increase in efficiency of diagnosis.
- For certain diseases like cancer, alzheimer, blood thickening the impact is not felt by the person at the initial stages. For such kind of ailments continuous monitoring system helps in identifying the problem before the symptoms are felt and hence early diagnosis can be made
- The doctor can make diagnosis by not only examining the patient at that time but the doctor can take a look at the status of vital sciences for at any time before.
- Even if the patient is not able to explain about his/her symptom the doctors can diagnose using the data collected from the sensors
- It helps in early diagnosis. Early diagnosis is useful in implementing the treatment plan and to follow up regularly how the patient react
- When the diagnosis is done earlier the patient can choose the treatment plan and cope up well with the treatment.

Simulation Results for the proposed Optimizing Algorithm:

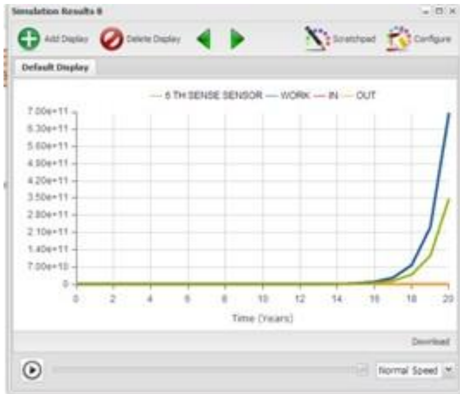

Figure 9 Effeciency(blue line) of sensor when during abnormal state

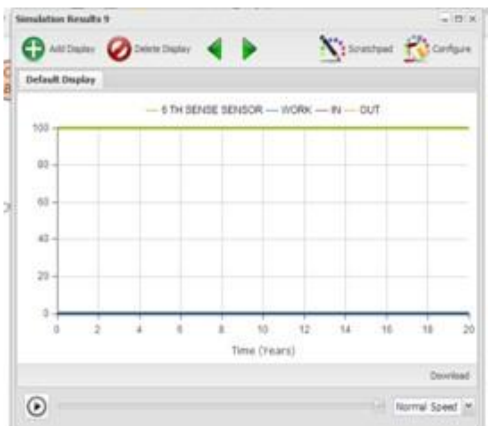

Figure 10 Effeciency(blue line) of sensor when during normal state

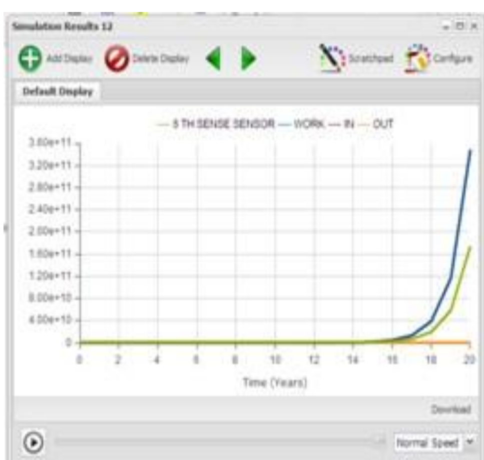

Figure 11Sensor always active even during unwanted state(when algorithm not used)

V. CONCLUSION

Thus the proposed system paves way for complete analysis of body and helps in detection of different diseases at different levels. It also optimizes the energy demand by adopting two newly proposed algorithms.

References

[1] Integration of wireless sensors network with cloud computing for secured architecture-Naveen Raj Dhanpal

[2] WSN Integrated Cloud for automated telemedicine based e-health Perumal.B 1+, Pallikonda Rajasekaran.M 2 and Ramalingam.H.M 1

[3] Phone centered Body sensor network platform-Lin Zhong Mike Sinclair Ray Bittner

[4]Ubiquitous Monitering Environment for Wearable and Implantable Sensors(UbiMon)-Jason W.P.Ng,Bennyy P.L.Lo,Oliver Wells,Morris Sloman,Nick Peters,Ara Darzi,ChrisToumazaou and Gauang-Zhong Yang

[5] Architecture of Body Sensor networks -Benny Lo and Guang-Zhong Yang

[6]A cloud computing Solution for patients data collection in Health Care Installation Carlos Oberdan Rolim, Fernando Luiz Koch, Carlos Becker Westphall, Jorge Werner, Armando Fracalossi, Giovanni Schmitt Salvador Network and Management Laboratory – LRG Federal University of Santa Catarina
Florianopolis - SC – Brazil

[7] Michael W. Trosseta, –Representing Clutchs: K-Means Clutching, Self-Organizing Maps, and Multidimensional Scaling‖, Technical Report 08-03, Department of Statistics Indiana University Bloomington, INaD, February 20, 2008.

Biographies and Photographs

Ranjith Ramesh- I am doing my Masters in Computer Science and Information System in Johns Hopkins University. I am also working as a Data Science Research Assistant in Institute of Genetic Medicine, Miller Research Lab, Johns Hopkins Medical Institute. I worked as an Application Developer at iNautix Technologies. I did my B.Tech Information Technology in Sri Sairam Engineering College.