DESIGN AND EXPERIMENT AN EMBEDDED CIRCUIT-BASED HEALTHCARE BIGDATA GENERATION AND DATA ANALYTIC MODEL

[1]M. Arun Kumar, [2]Dr. R. Vimala, [3]Dr. K.R. Aravind Britto

[1]Assistant Professor, Department IT, PSNA College of Engineering and Technology, Tamilnadu, India.
[2]Assistant Professor, Department of EEE, PSNA College of Engineering and Technology, Tamilnadu, India.
[3]Assistant Professor, Department of ECE, PSNA College of Engineering and Technology, Tamilnadu, India.
Email-ID: [1]kmarun_vicky@yahoo.co.in , [2]vimala79@gmail.com, [3]krbritto1975@gmail.com

Abstract: *Analysing healthcare bigdata is one of the emerging applications in recent days. It is practically identified that the number of patients entering to hospital industry is increasing day by day. So, the data generated for each patient is also increasing. The ratio of doctor and the patient in a day is 3: 50, where a doctor cannot spend more time with a patient for analysing the symptoms and diseases. So, it is very hard to find out the abnormal condition of a patient like high/low heartbeat, blood pressure and etc. Hence patients are not happy with the doctors and their treatment. This problem is taken into account and this paper motivated to design and implement an IOT based patient monitoring system incorporated with Cognitive technology-based Data Monitoring and Transmission (CDTM) as a data analytical model for analysing the healthcare data to predict the abnormal condition. The circuitry system attached in the patient body generates the healthcare data automatically and transmit to the analytical model for predicting the abnormality.*

Key words: *Healthcare Monitoring System, Bigdata Analytics, Healthcare Bigdata, Data Prediction, IOT based Bigdata.*

1.Introduction

Healthcare industry is one of the tremendously growing industry where predicting the condition of a patient is very hard. For example, finding the abnormality of a heartbeat is difficult. Because for a normal human being age below 55 the heartbeat rate is 140 to 170 beats per minute, but for an old person the heartbeat rate is 110 to 145 beats per minute. Patients are not satisfied with today treatment and searching for accurate health condition prediction. Also, the data generated on the patient, health diagnosis, treatment applied and medicine suggested are increasing speedily. In order to manage the large volume of patient and their data, each patient is monitored by interconnecting human body into a separate electrical-electronic embedded circuits or devices. Some of the sample circuit or device interconnected/attached into the patient body is shown in Figure-1.

Figure-1. Embedded Electrical and Electronic Circuit for Patient Monitoring System

Figure-1 shows the healthcare data generated using an embedded system which records various data

describe the health condition of a patient. ECG, heartbeat, EMG and temperature are some of the data recorded and transmitted into the computer.

In this paper, it is aimed to provide a cognitive technology for efficient transmission of medical data in a healthcare monitoring system. The entire contribution of the paper is:

- Hospital monitoring system
- Data analytics model and
- Prediction Model

Each patient in the hospital is interlinked with a set of medical devices like ecg, eeg, and body wearable sensors. Each device records the health condition in a predefined format and it transmit to the internal server (directly connected in a short distance, like within a room). Then the data is analysed by training process and create a prior knowledge information for testing purpose. Then the data is distributed with high security in cloud, where any authorized medical experts can fetch and use the data. The main objective of this paper is to train and test the data for predicting the health condition or the treatment.

Wireless Sensor Networks (WSN's) and mobile networks permitted in hospitals and the outside patients are supervised through Internet of Things (IOT) [1] where patients are operated with various smart devices like In-Plant Pacemaker, Electro Cardiogram (ECG), Electromyography (EMG), Electro Encephalography (EEG) and Motion Sensors. These wearable devices gather health correlated data such as blood pressure, heart rate and body temperature that would be useful in physical condition tracking and medical healing. Huge data in healthcare [2] is a systematic environment to grid the enormous quantity of organized and unorganized patient data. According to the psychoanalyst, the amount of data of USA healthcare system has attained to 150 Exabyte's in 2011[3] and has improved to zettabytes scale [4] in the recent point of time. Also, the California based health network Kaiser Permanente has 9 million members and the data [3] composed from Electronic Health Records [EHR's] together with medical records, pathological images doctor remarks vary from 26.5 to 44 peta bytes.

The health data are recognized has a huge data that is distinguished by 5V's in forms of Volume, Velocity, Value, Veracity and Variety. The patient data are composed of peta or zeta bytes that illustrate the volume. The velocity is stated in forms of data influx rate from the patients. Variety clarifies the expanded data sets by means of value to the organized, semi-organized and un-organized sets like medical records, EHRs and radiological images and veracity explicate the reliability of the data sets in accordance to data accessibility and validity. The composed data are distorted into significant insights that explicate the importance of 5V's. Thus, an appropriate raw data must be composed in a well-organized manner in medical surroundings. In sophisticated healthcare systems, the patient's data are composed [5] through wearable devices operated with special kinds of sensors. In recent times, the development in mobile devices [6] like multi-censored smart phones are also utilized as the data composed devices. Thus, massive quantity of patient data is produced inside a hospital network, that requires to be accumulated and analysed proficiently. So, a cloud computing [7] permits dispersed storage space and developed environment is necessary to collect and process the healthcare data that could be accessed everywhere and every time easily.

Currently different data intensive requests have come forward that requires certain well-organized logical models. Several stochastic approaches [8] are deemed by diverse authors in the current times for health care structure analysis. Also, the similarity [9] among health structure of a patient is regarded by the doctor for improved results. Huge data logic is used in healthcare [10] to recognize the group of patient's illness and future forecast with the assist of several machine learning devices [11]. The data are examined and utilized as insights, constantly for patient concern in the learning healthcare system [12]. One of the correct set of technologies for analysing the medical data is obtained through big data analytics. It helps in evolving data analysis model creation and increase the infrastructure architecture. Big data analysis can also address the problems faced in data visualization and manipulation. The entire architecture of a healthcare monitoring system considered in this paper is given in Figure-2(a) and in Figure-2(b).

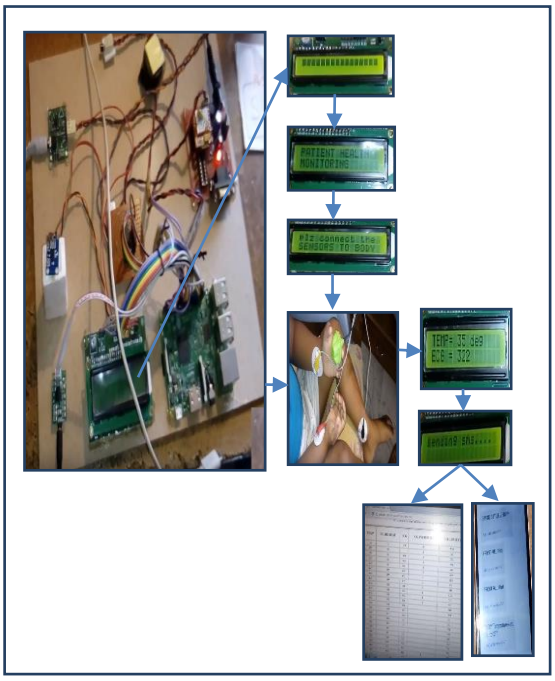

Figure-2(a). Embedded Circuit for Patient Monitoring System

Figure-2(a) shows the circuit system interconnected

into human body (IOT system) to monitor and record human health condition. The recorded parameters are transmitted to computer and mobile devices immediately. The recorded data is maintained in terms of date and time.

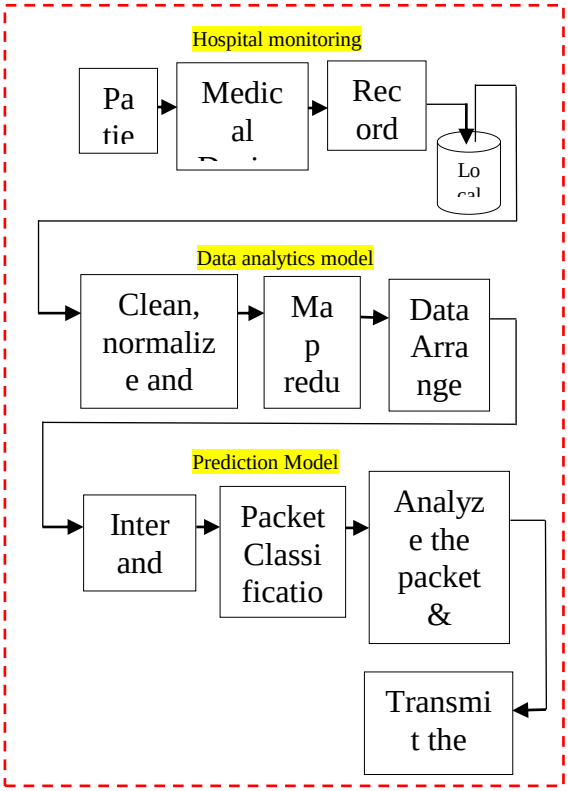

Figure-2(b): General Architecture of Medical Data Analytics

Figure-2(b) exemplifies, when a doctor orders a genetic test, it is tried to recognize a specific disease problem in the genome. Deciphering a wide-ranging genome would take up more space for each patient. Other than the test, data desires to be taking out from various other bases like medical sensors. Each sensor can provide 10GB of data at every second. All the results are compared to find out the genome patterns to take a decision. Due to the massive amount of data it is necessary to analyse the data large volume of data over various points of views. In order to increase the efficiency of the data analytical issues related to time, operational cost, and workflows it is necessary design and develop a novel approach whereas it can be obtained by using the proposed approach.

Yet, the existing design could not sustain both analysis and procedure for huge quantity of multi-organized healthcare data. In recent times, the larger presentation of cloud platform ensures a scalable and distributable similar procedural framework (i.e) MapReduce [13] for healthcare data dispensation. Map Reduce has the capacity to process the huge quantity of data in similar manner on a cloud. Thus, a hybrid form of stochastic similar processing structure is designed in medical surroundings to process and examine the large quantity of healthcare data. In this paper MapReduce similar processing structure [14] is used as a backbone for healthcare huge data analysis. Additionally, the CDTM work is expanded to prognosis a design for future health condition prophesy of the patients.

2.Related Works

The chronological and spatial connection of intellectual Pilate exercise data are examined in [15]. Still, partial works are carried out on correlation analysis of healthcare parameters among various patients. In hospital transience of critical unit, patients are forecasted using a confined huge data-driven arbitrary forecast representation [16]. A concise review is executed based on benefits and difficulties of application and technical necessities for hospital and BAN patient's observation in [17]. Hu et al [18] has projected the data acquirement process by using sensors, log files and web crawler in different applications. Moreover, the occurrences of the patient visit are not deemed in data compilation. A methodology, with unique huge data architecture for healthcare is proposed in [3] and zhang et al. [19], suggest a task-level adaptive and scalable MapReduce structure, which could calculate the future influx rate of work load on the map and reduced period. Likely MapReduce structure is designed to decrease the re-calculation for increased iterative calculations in [20]. An online society-based health services is proposed in [21], in which the health data are gathered and extracted through certain questionnaires and their individual answers. A scalable and distributable process is proposed in [22], to identify the similarity between patients by adjusting the MapReduce structure. Moreover, the appointment regularity health parameters and unknown symptoms of patients are most essential although they are not taken into deliberation for evaluating and processing the data in this work.

Future prediction of diseases is very decisive and significant for the patients with chronic illness. In [23] various kinds of Artificial Neural Network (ANN) methods are conversed for disease prophesy. Yet ANN approach takes more time for preparing the model due to expanded weights correlated with every layer. Still, a few small variations in the input data collection influence the model that provides unbalanced output. In [24], the upcoming solidity is experimented by using the corrosion-based attribute graph for the medical prognosis in high-dimensional electronic medical documents. An analytical structure is designed in [25] to incorporate the HER data to efficiently predict the osteoporosis and bone fractures. Henrique's et.al [26] envisages the de-compensation of heart malfunction by taking into consideration the physiological data of the patients. Conversely the hidden symptoms of the sickness are not taken into account in the present

prophesy models. The bio-sensors like ECG, EMG and EEG are used to accumulate and transmit the health constraints to backend servers for progression. Although various researchers have projected the operation and sensing strategies of body sensors for set data, none of them have enlarged the data set models of inside and outside patients based on the occurrence of appointment to the hospital. Moreover, the parallel analysis is integrated with sickness prophesy between the patients in the hospital.

2.1 Contribution of the Paper

Thus, the major perspective of this paper is to discover the eminent characteristics of the diseases by establishing the correlation analysis of healthcare parameters that are summarized below,

1) Projects a cognitive data transmission method based on the occurrence of out-patients appointment and number of data generated from the patients with BAN.
2) Correlation analysis is applied for the patient's data of intra and inner sections of the hospitals.
3) An algorithm for foreseeing upcoming health condition of patients based on their present health condition is planned.

3. Cognitive Data Transmission Model

Generally, in medical domain, the patient's records are stored in a multi-dimensional data structure. It has patient number, patient name, date, disease details and the medical advises with the doctor comments is stored. But it should be noticed that arranging, maintaining and searching the patient data is very difficult. Searching a data using the index (patient number) takes more time and takes more comparison complexity and it reduces the efficiency of the entire data analytical process. The proposed CDTM model carry out several steps of data analyzing processes continuously. Initially it does data cleaning, redundancy removal and error removal in the data and applies dimensionality reduction. After dimensionality reduction the entire data is mapped using a defined index value. The dimensionality reduction and mapping functions integrate the entire data even if the size of the data grows as large. The integration helps to combine any hospital data comes from anywhere in the world in cloud.

3.1 Data Collection Model

In conventional healthcare systems, the patient data are gathered, piled up and examined in a customary method that cannot maintain the diagnosis of multifaceted health conditions. One of the aspects which influence the efficiency is data collection and arranging. CDTM used a windowing system for data collection and arrangement. The set of all data flows in a time interval is divided into number of windows. Each window represents a set of patient data at a time $\mathbf{t_1, t_2, \ldots. t_n}$. The entire dataset is divided into a periodic manner like days; the data in a day is divided into hours. Each window comprises of three hours data and manipulated. Moreover, in our CDTM data set scheme, patients, doctors and BANs are deemed in general as the basis of data generated depending upon the rate of visits(r) of a patient rather assuming only number of patients as examined in conventional schemes. A window based [28] temporary data set and monitoring models are utilized to augment the efficacy of patients monitoring.

3.2 Data Structuring

Let us consider a cloud-based health care environment, with 'h' number of hospitals in a set $\boldsymbol{H = \{H_1, H_2, \ldots, H_h\},\ \forall h \in H}$ as stated in figure1. Let different departments be connected by one hospital and for simplicity, it is implicated that similar and identical number of departments are present in each hospital. Let $\boldsymbol{DP = \{DP_1, DP_2, \ldots, DP_d\}, \forall dp \in DP}$ be the collection of $\boldsymbol{dp}$ number of departments connected with every hospital. Every department nearby is combined with various number of doctors out-patients and BAN patients, that are the basis for generating the huge data. It is to be distinguished that out-patients are the patients who attend a hospital for treatment without residing there whole night. BAN patients are the chronic patients fixed with smart body sensors to examine their health conditions regularly.

Let, $\boldsymbol{do}$ represents number of doctors available in a set $\boldsymbol{Do_{ij}^{x}}$, where $\boldsymbol{j = \{1, 2, .., do\}}$ in the $\boldsymbol{i^{th}}$ department of $\boldsymbol{x^{th}}$ hospital $\boldsymbol{\forall i \in DP}$ and $\boldsymbol{\forall x \in H}$. Hence,

$\boldsymbol{Do_{ij}^{x} = \{Do_{1d}^{x} \cup Do_{2d}^{x} \cup \ldots \cup Do_{dpdo}^{x}\}\ \forall i \in DP}$ and $\boldsymbol{\forall x \in H}$. For example, $\boldsymbol{Do_{12}^{5}}$ says the doctor-2 from department 1 in hospital 5.

Let $\boldsymbol{P_{ij}^{k}}$ be the set of patients in which $\boldsymbol{j = \{1, 2, \ldots p\}}$ in the $\boldsymbol{i^{th}}$ department of $\boldsymbol{x^{th}}$ hospital, $\boldsymbol{\forall i \in DP}$ and $\boldsymbol{\forall x \in H}$. Thus P, represents number of patients present in the i[th] department of $\boldsymbol{x^{th}}$ hospital. Thus, $\boldsymbol{P_{ij}^{x} = \{P_{1p}^{x} \cup P_{2p}^{x} \cup \ldots P_{mp}^{x}\}, \forall i \in DP, \forall x \in H}$ .For example, $\boldsymbol{P_{34}^{2}}$ denotes the patient-4, that is associated to the department-3 in hospital -2. It is considered that patients with BANs are also hospitalized, that could be neither a patient nor a BAN at a time. Likely, $\boldsymbol{b}$ represents the number of BANs appeared in a set $\boldsymbol{B_{ij}^{x}}$, where $\boldsymbol{B_{ij}^{x} = \{B_{1b}^{x}, B_{2b}^{x}, \ldots, B_{8b}^{x}\},\ \forall i \in DP}$ and $\boldsymbol{\forall x \in H}$ and various number of BAN are accessible in several departments within the hospital. For example, $\boldsymbol{B_{13}^{2}}$ denotes BAN-3 is associated to the department-1 in

hospital-2.

In the proposed PDC model, a window based chronological data set and monitoring model is used to improve the quality of patient monitoring. Let $T = \{0, 1, 2, \ldots t\}$ be a constant time frame that is partitioned into W number of windows, in which every window contains Z units of time interval. Every time interval could be deemed as a minute, an hour, a week, a month or a year based on the application going to be handled. Consequently $Do_{ij}^{x}(w)$, $P_{ij}^{x}(w)$, and $B_{ij}^{x}(w)$ denotes the amount of data produced from BAN, doctors and patients correspondingly in every window w. The gathered data inside window w are piled up in various cloud data centers as shown in Figure-3. Let $\{DC_1, DC_2, \ldots, DC_n\}$ be the N numbers of geo-dispersed data center positioned in the cloud, where $n \in N$. These data centers are joined through M numbers of gateways $G = \{GW_1, GW_2, \ldots, GW_m\}$ where $m \in M$. In our structure, H represents number of those hospitals that are joined by those N numbers of geo-dispersed data center via M number of gateways.

The overall data is collected from the data center which is available in the cloud. Initially the data gathered from various medical devices interconnected with the healthcare network. The data is initially processed and corrected. After correction the data is converted into data packets. Each data packet is created and stored in a predefined format where it has a filed called as criticality bit field. Based on the criticality bit, the data packet is transmitted with highest priority assignment. Since, emergency patient's information is transmitted with highest priority to get appropriate treatment suggestion from the medical experts interconnected in the same network. Hence it improves the efficiency of medical data analysis and treatment prediction in the cloud.

3.3 Inter-Intra Clustering Process

Speed and accuracy of the mining process is increased by clustering method. Two types of clustering methods are used in the paper, are inter-cluster and intra-cluster. Intra-clustering method cluster the data within a hospital H_i where it classifies the patient data based on the patient ID, body area network ID, and classify the data based on the nature of the disease. The entire patient data is clustered and arranged in such a manner where predicting any data is very speedy and accurate. In order to globalise the data using cloud data centres, the overall data persisted in the cloud data centres is clustered and it is called as inter-cluster process. The inter cluster process cluster the entire data comes from various hospitals interconnected into cloud. Inter clustering is applied on the results of the intra-cluster. Hence the efficiency of the inter clustering is improved in terms of reduced time complexity, increased speed and improved accuracy.

3.4 Prediction Model

The conditions of the patient are predicted by validating the medical data (E.g. ECG) and compared with the well-defined threshold values. The patient health condition is well-defined as normal, mild, moderate and severe by medical experts or laboratory experts and available in the medical industry. Each condition has a threshold value whereas it can be used for comparing with the recorded values. The patient data is collected in periodical manner and compared. In order to compare effectively, the recorded data is divided using widow system.

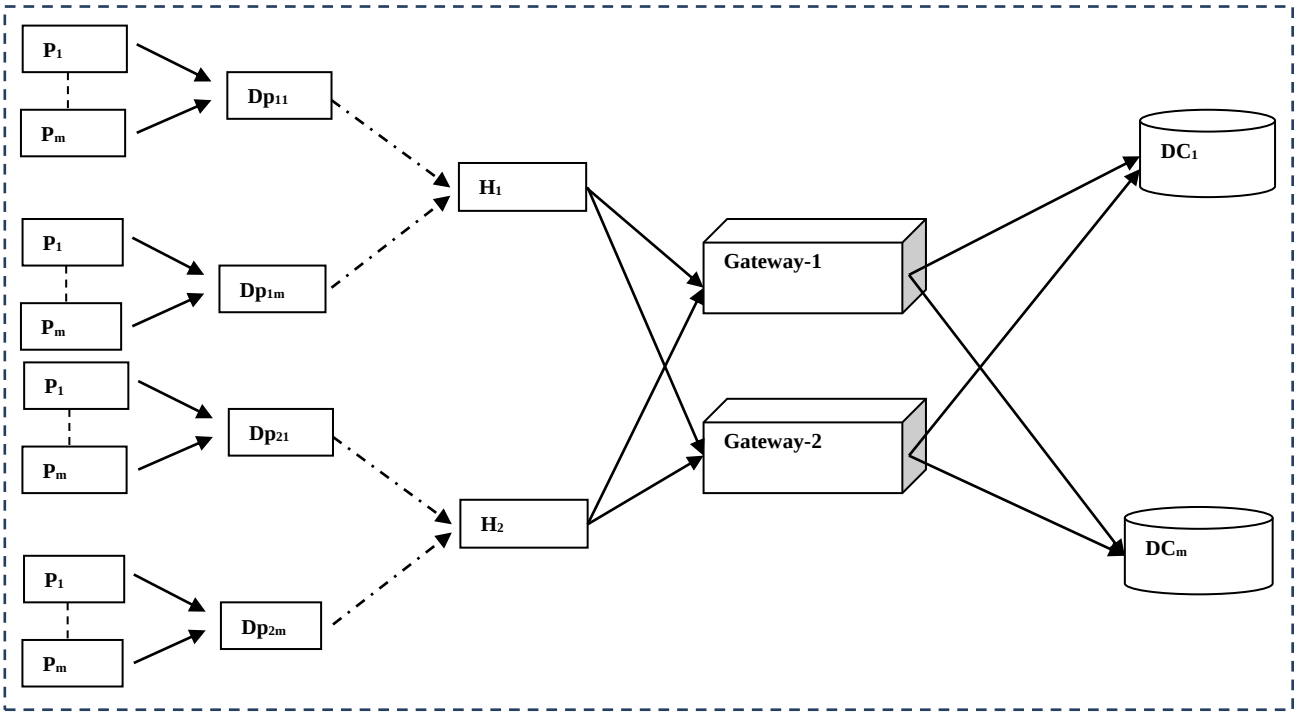

Figure-3: Proposed CDT System Model

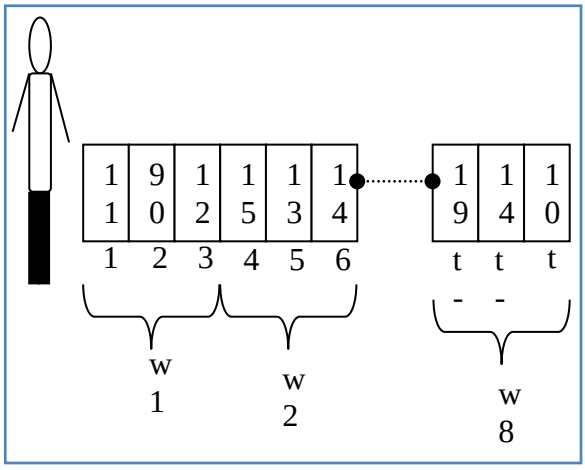

Figure-4: Data Collection in Windows Based Model

The window volume could be customized based on health conditions. Consider this example, let A be the blood pressure data of a patient, that is verified on the time framework $\boldsymbol{T = \{1, 2, 3, \ldots . t\}}$, with window frames $\boldsymbol{G = \{W_1, W_2, \ldots, W_m\}}$, as shown in figure-4, where the interval of every window $\boldsymbol{W_i}$ holds 3units. In second window $\boldsymbol{(W_2)}$, the normal blood pressure data (**Avg.** $\boldsymbol{W_2}$) of three time slots (4, 5, 6) is verified as 140, that is higher than the vital condition **(if (Avg.** $\boldsymbol{W_w}$**> 125)).** Those verified data by time interval are gathered by using our proposed data acquirement scheme and is conveyed to the data center in the cloud for storage and analysis.

In huge data healthcare surroundings, the EHR, 3D imaging, physiological data, medical, radiology images, genomic sequencing and billing data are the basis of huge data which illustrate the volume. Synchronized and urgent situation monitoring is the basis of streaming data that illustrates the velocity of data. Although the majority of the papers believe the patients physiological data as the huge data [3], we incorporate the visiting rate of the patients to the hospitals in our huge data processing models.

3.5 Data Transmission Model

Generally, the sensor mote is a charged device with unique radio broadcasting features. For example, Imote-2, Mica-2, Mica-z are the sensors used for human health care monitoring. Each patient in ICU is bounded by some human body sensors where they can sense and record the appropriate human body information like heartbeat, pulse rate, and so on. The information from each sensor is formed as a packet which contains BED-NO, PAT-ID, SEN-ID, EXP-DAT, MON-DAT, CRI and it can be written in the form of data packet is shown in Figure-5.

BED-NO	PAT-ID	SEN-ID	EXP-DAT	MON-DAT	CRI

Figure-5. Packet Format

BED-NO : denotes unique bed number with no replica
PAT-ID: denotes unique patient ID
SEN-ID: denotes sensor ID
EXP-DAT : denotes the nominal range of sensor data
MON-DAT : denotes the obtained data from the sensors
CRI : denotes a Boolean type indicates the critical/no-critical

3.6 Cognitive Technology Based Packet Transmission

In order to provide cognitive technology for data analysis and prediction, this paper uses Simulated Annealing method (SA). SA comprises two functions for changing the structure of the metal such as heating and cooling. It changes the physical properties of the metal. The SA kept the temperature as variable. In the beginning stage, the temperature is high and assign the state is "cool". If the temperature is high, the frequency is high and it is the worse solution. Here the algorithm is applied to change the priority based on the critical bit value. If the critical bit value is "0", then the data packet transmission is normal, that is as such in the queue the data transmitted. If it finds the critical bit is "1", then that particular data packet is transmitted with high priority. After transmitting the data with highest priority, other data packets will be transmitted.

The simulated annealing algorithm used for verifying the critical data and transmitting the data packet is given in the form of pseudo code.

```
Simulated_Annealing Algorithm ( )
Start
Initialize i = 0; N = number of data packets at the time interval t;
For i=1 to N
  If ( Datapacket.CRI (i)) == "1" ) then
     Priority=1;
  Else
  Gateway (i) = Datapacket (i)
End if
Repeat the above steps until the time interval over
End
```

Say sensor S1 from Table-2, gives the information about Blood Pressure (BP) and pulse rate. Whereas the nominal range of the sensor is given as 120/72. But the sensor observed data in time **t** is 135/90. From the observed data according to BP range it is abnormal, by this CRI is updated as "**Y**" and packet is placed immediately in the bus. Similarly,

sensor S2, gives the information about Blood Gluco Level (BGL). The nominal range of the sensor is given as 135, but the sensor observed data in time **t** is **133**. From the observed data according to BGL range it is normal, by this CRI is updated as "**N**" and packet is placed in the bus in a scheduled time. Similarly, all the human body sensors are recording and placing the data on the bus immediately or in the scheduled time by referring the CRI value. In this paper two different patient at two bed with sensors used and the CRI value computed in given in Table-2 and in Table-3. The sensors used in this paper and its purpose are given Table-1.

Table-1: List of Human Body Sensors and its Purpose

Sensor ID	Purpose
S1	Pulse Oximeter Sensor
S2	Glucometer Sensor
S3	HG sensor
S4	Clinical Thermometer sensor
S5	Patient Position Sensor
S6	Sleep Monitoring Sensor

Table-2: Sensor Data Recorded at ICU-1

BED-NO	PAT-ID	SEN-ID	EXP-DAT	MON-DAT	CRI
107	AP-1001	S1	120/72	135/90	Y
107	AP-1001	S2	135	133	N
107	AP-1001	S3	13.0-17.5	11.25	Y

Table-3: Sensor Data Recorded at ICU-2

BED-NO	PAT-ID	SEN-ID	EXP-DAT	MON-DAT	CRI
307	AP-1023	S4	95°C - 97°C	103°C	Y
307	AP-1023	S5	SS	SS	N
307	AP-1023	S6	25 (rpm)	24 (rpm)	N

After recording the patients' information, the recorded data is passed to SPA (Sensor Packet Administrator). Means sensor device sends the data in the specified format (see Figure-5) to the common bus. For every second, SPA looks for fresh data in the bus. SPA maintains two different arrays as LP-array (low prioritized array), HP-array (high-prioritized array) with two different pointers in order to represent each array called store-pointer and transfer-pointer. Based on the CRI, SPA stores the data packets in the specified array, actually array is maintained in such a way that if CRI is "Y" then data packet store in a high prioritized array and store pointer increments by **1**. Now the transfer-pointer sends the data packet from HP-array to aggregator called master collector device which in turn sends the data to physician's knowledge to take an appropriate action. Store and transfer pointer work mutually exclusive, store-pointer will keep on storing the sensor data and move to the next location and wait for the new data. Whereas transfer-pointer will wait for a positive acknowledgement from master collector device and then delete sensor data from the current location of the array and increments to the next location. Here the arrays are considered as a circular queue. Once service is done, the transfer-pointer clears the data. After clearing the last location of the array pointer will be initialized to 0^{th} location. Similar mechanism is applied for LP-array, where the only difference is once HP-array become serviced and cleared then only SPA will serve for LP-array.

Sensor units are integrated sensor devices, which collects data from human body sensor devices which are connected in-bed patients. Here total six human body sensors are considered to monitor the patient in equal time intervals. The sensor unit associated with each bed is acquiring data from sensor devices and based on the criticality it will place the data packet in the bus with high priority. The data packet steam lined in the bus is available for SPA. The allocator unit is the scheduler which controls and coordinates the sensor data from different sensor units. Scheduler is all time process unit which has redundancy. In case one scheduler is busy or fails to serve the process from sensor unit bus, other scheduler will be considered to continue the process. If criticality parameter is high then SPA assigns the data packet into HP-Array, otherwise data packet containing sensor information is stored into LP-Array.

Generally, these arrays are circular arrays having two pointers; the one is to store the data packet and increments the pointer to the next location. Once packet is stored into this array, the aggregator which is in loop will take packet and transferred to physician's bay. The same data packet will be sent to base station to keep track the information and store it to the data sheet. If SPA allotted data packet in Physician's bay is not getting attended and that data packet will be controlled by other Physician which in turn given by base station. If HP-Array is fully loaded and there is no room to allocate new data packets from the sensor unit, then LP-Array is utilized for this purpose, this is done automatically by SPA. LP-Array holds the data packet which is nominal and non-critical cases only, hence non-availability of HP-Array case, the algorithm can use the LP-Array as a storage platform and very finite time the same data packet is transmitted to the base station and physician's bay. By using higher priority array and lower priority array the congestion is avoided and also data traffic is controlled in an efficient manner. This PPI method not only for health care industry, it also can be applied in various WSN applications.

The entire process of the CDTM is given in the form of algorithm where it can be implemented in any

programming languages and the results can be verified.

Algorithm_ CDTM Method ( )

{

- Initialize number of patients $\boldsymbol{P}$, hospitals $\boldsymbol{H}$, departments $\boldsymbol{D_p}$ and collect all the data
- Separate the data in terms of hospitals $\boldsymbol{H}$, departments $\boldsymbol{D_p}$ and doctor $\boldsymbol{Do}$
- Map the data $\boldsymbol{P_{ij}^x}$ based on the hospitals $\boldsymbol{H}$, departments $\boldsymbol{D_p}$ and doctor $\boldsymbol{Do}$
- Group the data in terms of time into Windows $\mathbf{W_i}$
- Cluster the data in $\mathbf{W_i}$
- Classify the data in terms of disease
- Predict the data for user query $\mathbf{Q_i}$

}

3.7 Data Acquisition Scheme

The number of appointments of a patient to seek advice from doctor in various hospitals, required to be analysed since they could produce data during every appointment. Let us consider that a patient ($\boldsymbol{P_{ij}^x}$) makes an appointment $\boldsymbol{f}$ times to a department (DP_i) within $\boldsymbol{W}$ time intervals. Every patient at a time of appointment let $\boldsymbol{pf, pDP, pD}$ be the probabilities of patient's appointment rate of hospitals, departments and doctors. To be considered that $\boldsymbol{pV}$ and $\boldsymbol{pBA}$ are the smallest value of the probability of a patient who makes an appointment in the hospital and BAN correspondingly.

It is experimented that the possibility $\boldsymbol{pBA}$ rises if the BANs connected with several doctors and department increases. As stated in [29], the medical test data and radio-logical images are deemed as organized and unorganized collections of data correspondingly. In accordance to the author's one tera byte of medical text data and 19 tera bytes of image data are produced by 25000 patients per year. In our data acquisition scheme, text and image data of the patients are also deemed that are produced during the appointment of the patients. Consider $\boldsymbol{U}$ and $\boldsymbol{V}$ megabytes, be the size of every text and image data correspondingly and $\boldsymbol{\varphi^p}$ be the sum of data produced by a patient during a sole appointment. Hence $\boldsymbol{\varphi^p}$ is the total sum of text ($\boldsymbol{\varphi_{TD}^p}$) and image ($\boldsymbol{\varphi_{ID}^p}$) data of a patient $\boldsymbol{p}$ which can be calculated in an efficient manner in accordance to the number of department and number of hospitals. In order to increase the efficiency of data retrieval and prediction on the patient's data an inter-cluster and intra-cluster methods are used. These clustering methods give an arranged data in a uniform manner which increases the speed of the searching. The prediction process looks into the patient's information under a department in a hospital according to association among the Identification key assigned to each element. This process is implemented and experimented in MATLAB software and the results are given here. In order to verify the efficiency and performance of the proposed approach the data is collected from cardiac disease patient's information with mild, moderate and sever attack information records taken from All the data are collected from Cleveland and Hungarian clinic [40] available for public use.

3.8 Data Sharing

Once the data processing like clustering classification is completed, the patient data is persisted in the cloud data centers. The medical experts and doctors need to be registered in the healthcare cloud, for accessing and sharing their patient data in the cloud. Unless the doctors are not registered in the cloud, cannot access the patient data. Same time, the data having highest priority is shared among the doctor's group immediately for fast treatment.

4. Experimental Results and Discussion

In accordance to the architecture (given in Figure-2.), the entire proposed approach of the paper is experimented in MATLAB software and the stage wise results are given here. Initially the data is generated for a set of patients linked in the healthcare network. The generated data has more irrelevant data as well as redundant with errors. Hence the data needs to be pre-processed and after preprocessing the data size is decreased. Preprocessing of data increased the accuracy of mining and prediction. Figure-6 shows the comparison of data generated and pre-processed in terms of size. The pre-processed data is considered as the valid data and it is used for global access and further data analysis. The volume of data generated is increased based on the number of patients involved in the network.

Data cleaning, error correction and redundancy removal are the most important process which improves the quality of the data. Hence the data size is compared after preprocessing with the original data. It is experimented and the result is obtained and it is shown in Figure-7. From the comparison, the pre-processed data size is comparatively lesser than the original data. The data size is measured in terms of Giga Bytes (GB). The less and accurate data can improve the efficacy of the prediction and mining in healthcare data inter linked with cloud. The size of the original data and the pre-processed data size is increasing based on the time interval and patients.

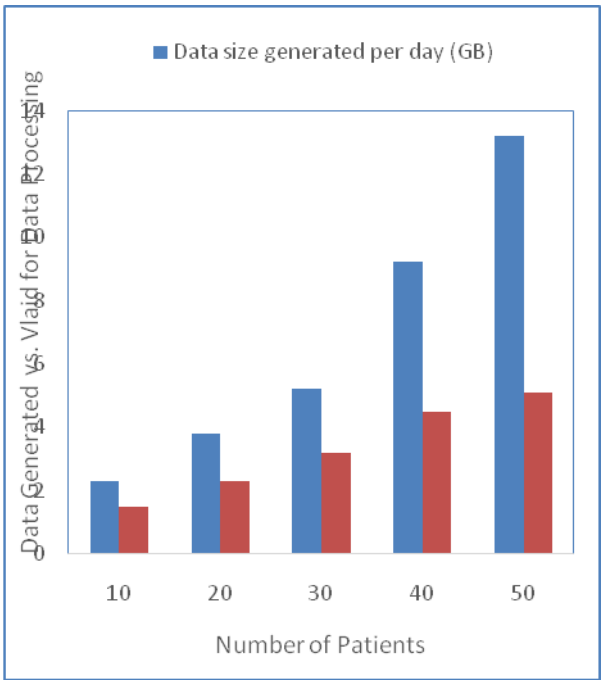

Figure-6. Data generated versus Data Valid for Analysis

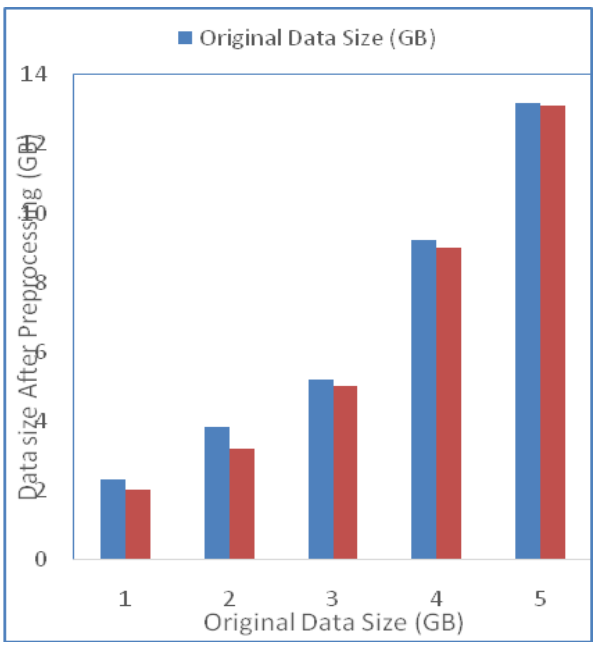

Figure-7. Data Size Comparison Before and After Preprocessing

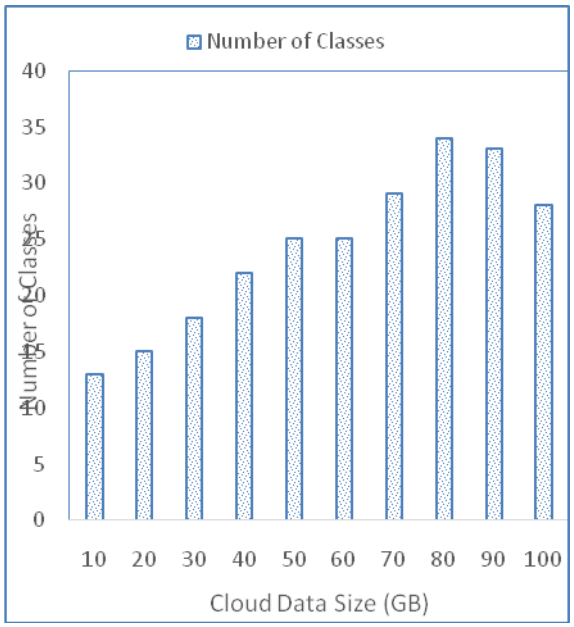

Figure-8. Number of Classes Obtained after Clustering

After preprocessing the healthcare data is clustered using inter and intra clustering. The clustering process is carried out based on disease-ID and patient-ID. Initially the healthcare data is clustered in hospital based (local DB) then the data is clustered at cloud data centers (global DB). After clustering, all the data is clustered under certain classes based on the diseases. Each class represents each disease of the patient. The number of classes obtained from clustering process is shown in Figure-8.

Figure-8 shows the results after inter and intra clustering method. But the clustering result after inter and intra cluster is calculated separately. Initially the correlation among the data is obtained by calculating the similarity values. It is calculated among inter and intra clusters. The correlation can be calculated at any point of view like data under same hospitals, same disease, and severity of the disease, locality and other aspects. It is very much useful in classification and mining process also it increases the efficiency in terms of accuracy.

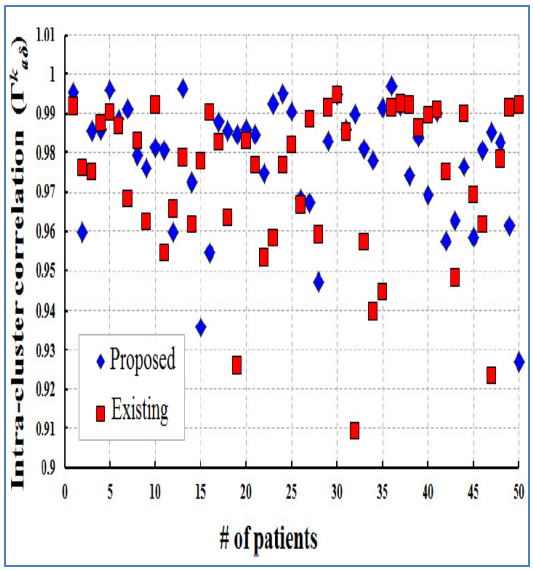

Figure-9: Correlation Obtained on Intra Cluster Method

In the initial stage, the preprocessing method is applied to normalize the raw data using some reserved functions. In accordance to the number of patients the intra cluster method clusters the data in terms of correlation. The results obtained from both existing and proposed are shown in Figure-9. Instead of all the patients, 50 numbers of patients are considered in each set to observe the data more clearly. The plot shows that almost all the patients correlation values are greater than 0.95, which leads to disease-based correlation exist between two sets in one cluster. Therefore, the healthcare attributes of set 1 are highly correlated with the attributes of set 2. From the figure, it seems that our proposed algorithm is efficient for correlation analysis of the heart disease patients.

In Figure-9, the inter cluster method is applied to cluster the patient's data. There are two different data sets are used in the experiment. One is Cleveland and Hungarian with severe disease and the other one is mild disease. From Figure-10, it is observed that some correlation values are less than 0.6 and most of them are greater than 0.95. The values less than 0.6 are treated as the less correlated values, where greater than 0.9 values are treated as highly correlated patients. This correlation analysis could be extended to a classification analysis based on the correlation values.

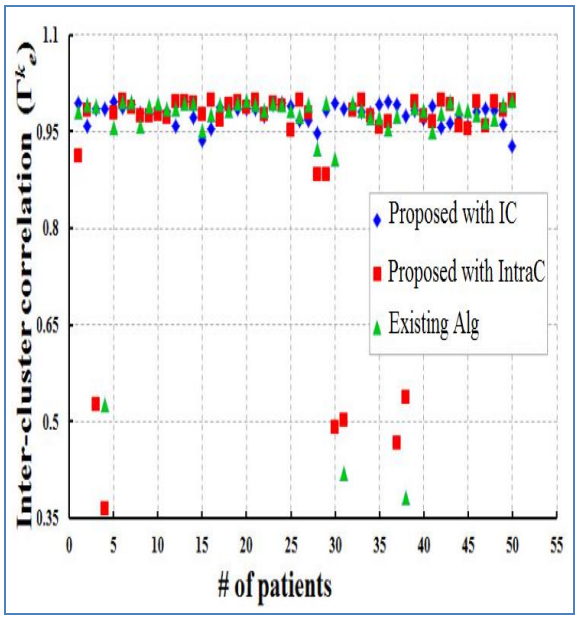

Figure-10: Correlation Obtained on Inter Cluster Method

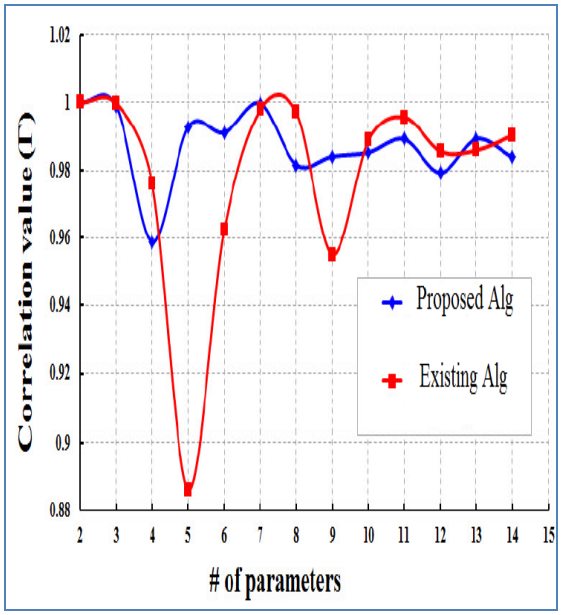

Figure-11: Number of Parameters versus Variations in Correlation

The number of attributes shown in Figure-11 plays a vital role in the correlation analysis. From the obtained simulation result in the beginning stage of clustering process, the correlation analysis is varied due to the less numbers of parameters. Even though this variation is minimized during the number of parameters is increased. For both datasets the correlation value is more drifted than the correlated values using both clusters. From the result it is observed that the correlation values are more stable according to large number of parameters.

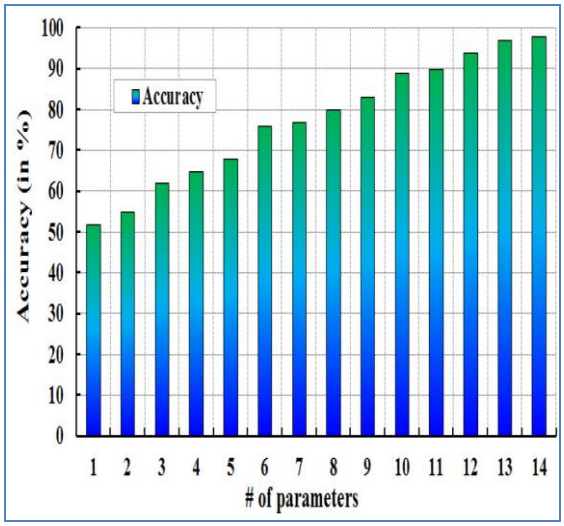

Figure-12: Accuracy in terms of Number of Parameters

It is without a doubt observed in Figure-12 that the accuracy of prediction is also elevated with increase in the quantity of attributes. When best one characteristic is considered, the accuracy is touched around 51%. However, it's far expanded up to 58% with attributes. The accuracy is about 80%, whilst the numbers of attributes are expanded to 8 and eventually, we got the maximum accuracy of 98%, when the numbers of attributes are accelerated to 14.

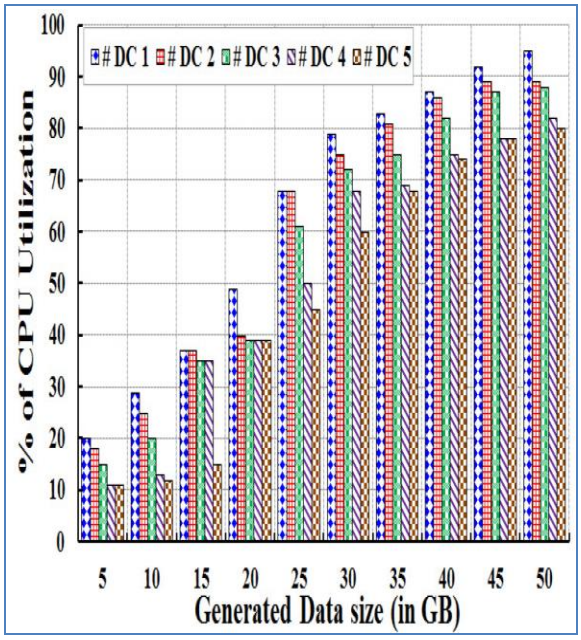

Figure-13: CPU Time in terms of Increased Data Size

Processing time is a critical element to take a look at the performance of the algorithms, which is proven in Figure-13. Here, the processing time is described as the summation of challenge scheduling time with information transfer time from distinct data centers to attain the records locality and execution time. It is found that the processing time is decreased with the aid of using a couple of virtual machines in distinctive records facilities. The processing time of proposed set of rules is longer than the prevailing algorithm as the responsibilities inside one statistics center await the execution until the jogging responsibilities are finished. The proposed set of rules with 3 facts centers is faster than the prevailing algorithm with 2 records centers due to parallel processing of the map obligations. Here, the main advantage of using cloud platform is to lessen the processing time.

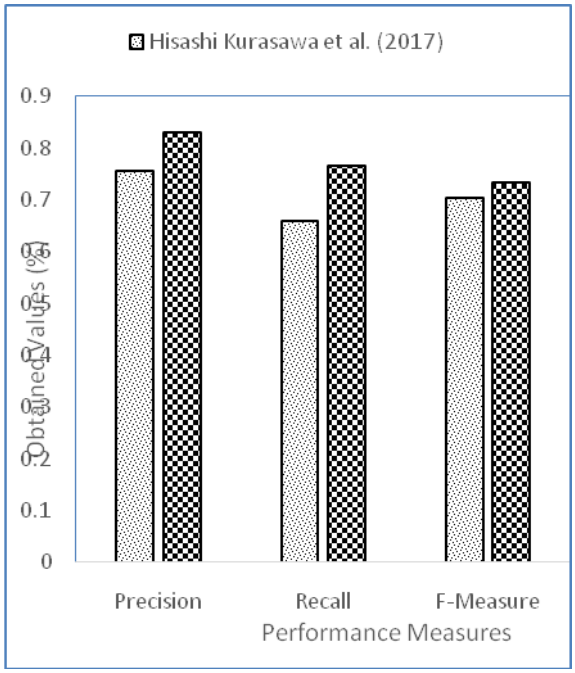

Figure-14. Performance Comparison with Existing Approach

To evaluate the performance of the proposed approach, some of the performance measures such as precision, recall and F-measures are calculated and compared with the existing approach presented in [28]. The compared result is shown in Figure-14. From the figure, it is clear and noticed that the proposed approach obtained higher performance than the existing approach. From the above discussion it is identified that the data is initially collected using a data collection model, arranged, mapped and clustered. The clustering and classification process are done by computing the correlation among the data under various aspects and it is calculated over inter and intra clusters. This process can increase the accuracy of classification. Finally, the prediction accuracy is calculated. The performance is also evaluated by predicting the data over various sizes of data. From all the experimental verifications, it is clear and identified that the proposed CDTM is more suitable than the other approaches.

From the experiment a set of data is applied a

greater number of times and the accuracy is calculated. For various sizes of data, the experiment is carried out a greater number of times and the results are predicted.

5. Conclusion

In this paper a CDTM model is intended for the cloud-based healthcare system. The existing and the CDT algorithms are structured for the intra and inner cluster correlation analysis of the healthcare huge data. A CDT algorithm is planned to prophesy the upcoming health status of the patients, based on their present health condition with the precision of 98% cloud-based MapReduce model is utilized as the procedural framework for our huge data analysis. It is experimented that our procedure could be used for several applications based on healthcare and patient monitoring like heart sickness prediction, cancer severity categorization.

From the results and discussion, it is identified that this CDT approach is more suitable for data analytics over medical data. It is also identified that it is more robust, reliable, dynamic and more scalable. Hence it is proved that CDT is a better approach. The performance is evaluated by comparing the calculated performance measures with the existing approaches in terms of precision, recall and F-measure. The proposed approach in terms of precision is improved 0.074%, recall is improved 0.106% and F-measure is improving 0.03% than the existing approach discussed in [28].

Our future work is to execute the proposed data analytic structure in the actual healthcare field to examine the data in real-time data analytic proposal like SPARK.

References

[1] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, ``The Internet of Things for health care: A comprehensive survey," *IEEE Access*, vol. 3, pp. 678_708, 2015.

[2] J. Andreu-Perez, C. C. Y. Poon, R. D. Merri_eld, S. T. C. Wong, and G.-Z.Yang, ``Big data for health," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 4, pp. 1193_1208, Jul. 2015.

[3] W. Raghupathi and V. Raghupathi, ``Big data analytics in healthcare: promise and potential," *Health Inf. Sci. Syst.*, vol. 2, no. 1, pp. 1_10, 2014.

[4] (Apr. 2014). *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*. [Online]. Available: https://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf

[5] K. Lin, F. Xia, W. Wang, D. Tian, and J. Song, ``System design for big data application in emotion-aware healthcare," *IEEE Access*, vol. 4, pp. 6901_6909, 2016.

[6] L. A. Tawalbeh, R. Mehmood, E. Benkhlifa, and H. Song, ``Mobile cloud computing model and big data analysis for healthcare applications," *IEEE Access*, vol. 4, pp. 6171_6180, 2016.

[7] C. K. Dehury and P. K. Sahoo, ``Design and implementation of a novel service management framework for IOT devices in cloud," *J. Syst. Softw.*,vol. 119, pp. 149_161, Sep. 2016.

[8] Z. Yu *et al.*, ``Incremental semi-supervised clustering ensemble for high dimensional data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 701_714, Mar. 2016.

[9] M. Mukaka, ``A guide to appropriate use of correlation coef_cient in medical research," *Malawi Med. J.*, vol. 24, no. 3, pp. 69_71, 2012.

[10] S. Rallapalli, R. R. Gondkar, and U. P. K. Ketavarapu, ``Impact of processing and analyzing healthcare big data on cloud computing environment by implementing hadoop cluster," *Procedia Comput. Sci.*, vol. 85, pp. 16_22, May 2016.

[11] S. Wang, X. Chang, X. Li, G. Long, L. Yao, and Q. Z. Sheng, ``Diagnosis code assignment using sparsity-based disease correlation embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3191_3202, Dec. 2016.

[12] V. Tresp, J. M. Overhage, M. Bundschus, S. Rabizadeh, P. A. Fasching, and S. Yu, ``Going digital: A survey on digitalization and large-scale data analytics in healthcare," *Proc. IEEE*, vol. 104, no. 11, pp. 2180_2206, Nov. 2016.

[13] T. Huang, L. Lan, X. Fang, P. An, J. Min, and F. Wang, ``Promises and challenges of big data computing in health sciences, " *Big Data Res.*, vol. 2, no. 1, pp. 2_11, 2015.

[14] J. Dean and S. Ghemawat, ``MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107_113, Jan. 2008.

[15] C.-H. Wu and Y.-C. Tseng, ``Data compression by temporal and spatial correlations in a body-area sensor network: A case study in Pilates motion recognition," *IEEE Trans. Mobile Comput.*, vol. 10, no. 10, pp. 1459_1472, Oct. 2011.

[16] R. A. Taylor *et al.*, ``Prediction of in-hospital mortality in emergency department patients with sepsis: A local big data_driven, machine learning approach," *Acad. Emerg. Med.*, vol. 3, no. 23, pp. 269_278, Mar. 2016.

[17] M. Patel and J.Wang, ``Applications, challenges, and prospective in emerging body area networking technologies," *IEEEWireless Commun.*, vol. 17, no. 1, pp. 80_88, Feb. 2010.

[18] H. Hu, Y. Wen, T.-S. Chua, and X. Li, ``Toward scalable systems for big data analytics: A technology tutorial," *IEEE Access*, vol. 2, pp. 652_687, 2014.

[19] F. Zhang, J. Cao, S. U. Khan, K. Li, and K. Hwang, ``A task-level adaptive mapreduce framework for real-time streaming data in healthcare applications," *Future Generat. Comput. Syst.*, vols. 43_44, pp.

149_160, Feb. 2015.
[20] Y. Zhang, S. Chen, Q. Wang, and G. Yu, ``*i*2 MapReduce: Incremental MapReduce for mining evolving big data," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 7, pp. 1906_1919, Jul. 2015.
[21] L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, and T. S. Chua, ``Disease inference from health-related questions via sparse deep learning," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 8, pp. 2107_2119, Aug. 2015.
[22] M. Barkhordari and M. Niamanesh, ``ScaDiPaSi: An effective scalable and distributable mapreduce-based method to _nd patient similarity on huge healthcare networks," *Big Data Res.*, vol. 2, no. 1, pp. 19_27, 2015.
[23] C.-H. Weng, T. C.-K. Huang, and R.-P. Han, ``Disease prediction with different types of neural network classi_ers," *Telematics Inform.*, vol. 33, no. 2, pp. 277_292, 2016.
[24] S. Gopakumar, T. Tran, T. D. Nguyen, D. Phung, and S. Venkatesh, ``Stabilizing high-dimensional prediction models using feature graphs," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 3, pp. 1044_1052, May 2015.
[25] H. Li, X. Li, M. Ramanathan, and A. Zhang, ``Prediction and informative risk factor selection of bone diseases," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 12, no. 1, pp. 79_91, Jan./Feb. 2015.
[26] J. Henriques *et al.*, ``Prediction of heart failure decompensating events by trend analysis of telemonitoring data," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1757_1769, Sep. 2015.
[27]. Hisashi Kurasawa, Akinori Fujino, and Katsuyoshi Hayashi, (2017), "Predicting Patients' Treatment Behavior by Medical Data Analysis Using Machine Learning Technique", NTT Technical Review, Vol. 15 No. 8, pp. 1-6.