Cloud based Data Mining framework for monitoring Healthcare in rural India

Vijesh kumar Patel[#1]

[#]Computer Science & Engg., SRMCEM, Lucknow , India
[1] Vijeshkumarpatel4@gmail.com

Abstract—Advance Information and Communication Technology (ICT) helps in making healthcare at affordable cost and in improving the effectiveness of healthcare services in remote areas. This paper proposed a framework for monitoring the healthcare of rural people. The proposed framework is designed to utilize a private cloud and data mining techniques for monitoring healthcare especially for rural areas. The main objective of this framework is to improve the quality of healthcare and provide healthcare facility at least for primary care in rural areas.

Keywords-Data Mining, Cloud Computing, Rural Healthcare, Information and Communication Technology.

I. INTRODUCTION

Health is an important issue for everyone. No nation can prosper with unhealthy people. In India, Healthcare facilities are not very good in rural areas. Most of the population of India lives in rural areas. The number of Primary Healthcare Centre (PHC) in India is around twenty three thousand, which is not sufficient as compared to the population of this country.
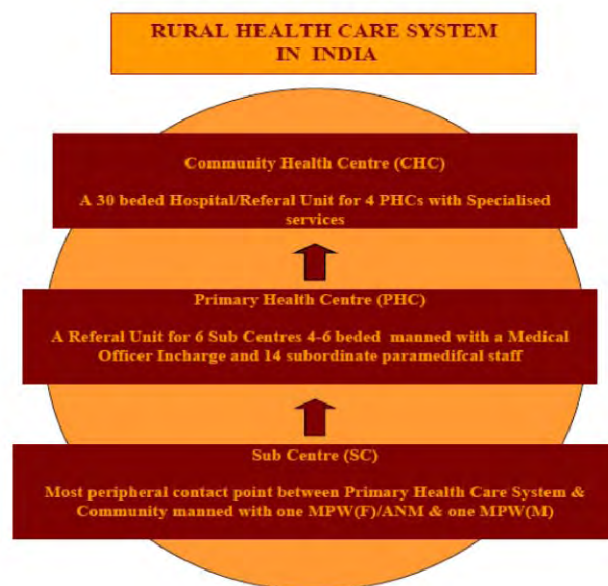

Figure 1 Rural Healthcare System in India [1]

Figure 1 shows the rural healthcare system in India. The healthcare system in rural areas is constructed using 3-tier system. The first tier is of Sub-Centre (SC), second tier is of PHC and third tier is of Community Health Centre (CHC). PHC acts as contact point between CHC and SC and it is established by state government. SC is operated by ANM (Auxiliary Nurse Midwife) and Male Health Workers. But the ratio of these health centers per population is not adequate [1]. Even in villages the healthcare facility is not available for primary care, which provides the healthcare facility for common and minor diseases and for those diseases where prevention is more effective [2]. Around 700 million people are living in rural parts of the country. Approximately, there are 6 lakh villages in India. The income of most of the villagers is very small. Even they don't have access to the basic healthcare facilities due to their poor income. Another major problem in rural areas is that the people don't have access to the safe drinking water which is the major cause of various diseases such as Diarrhoea, Leukemia, Fluorosis etc. In rural areas there is a low- doctor to patient ratio which in turns increases the cost of healthcare facilities and also it becomes inefficient. ICT could help in making these facilities available at lower cost and also enhance the efficiency of the basic healthcare [3]. Due to unavailability of doctors and poor number of PHCs, people don't have access to the basic and quality healthcare in villages. Supply of Healthcare services in India is also varying from urban to rural part [2]. Figure 2 shows the demand for healthcare services

and supply in India. The maximum demand is for Primary care, which provides healthcare facility for common and minor diseases and for those diseases where prevention is more effective, and then there is a demand for secondary care where patients require constant medical attention and short period of hospitalization, and at last tertiary care where patients require care from specialized clinics or expert. Healthcare supply in India is also varying from urban to rural part as shown in figure 2.
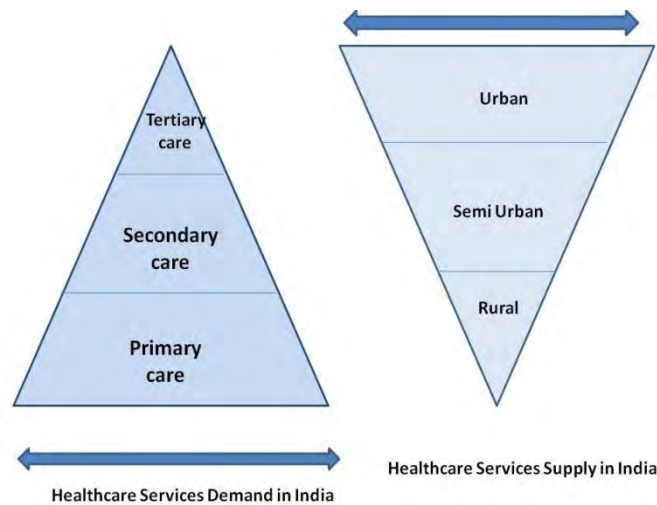

Figure 2 Rural and Urban Health Divides

In urban areas the healthcare facilities are available for corporate health and but in rural there is no facility for primary healthcare. Rural areas have inequity to access the hospital resources such as doctors to population and hospitals beds to population ratio is lower in rural areas as compare to urban area. If villagers fall sick they travel so far for getting the treatment in urban hospitals. Healthcare becomes expensive for them, so for the same illness they spend more money as compared to urban people. Due to their poverty they are not able to afford the healthcare facilities which are available in urban hospitals. There is a huge difference between healthcare facilities in urban and rural areas. In India Maternal Mortality ratio (MMR) is also high in rural part. MMR measure the number of deaths of women due to maternal causes per 1 lakhs live birth. Institutional delivery in rural areas is lower as compared to urban areas and also there is no system for monitoring the pregnant women health condition in rural. Due to lack of awareness and inaccessibility to the healthcare facilities MMR is high in these areas [4]. The trends in MMR in India in different years are shown in figure 3.
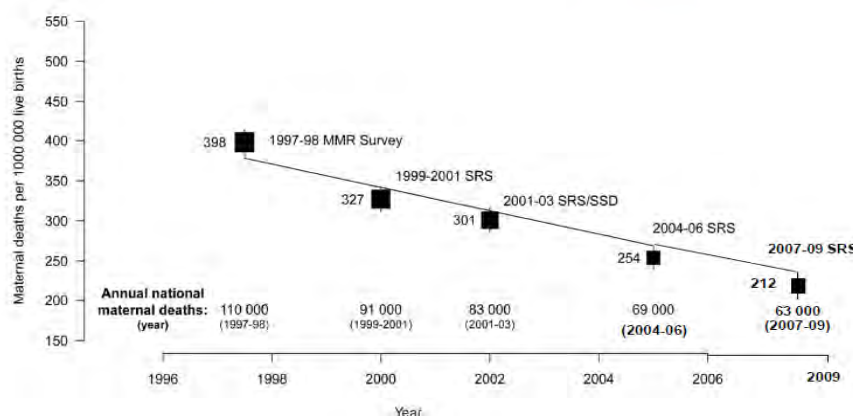

Figure 3 Maternal Mortality Ratio in India [4]

As per MDG (Millennium Development Goal), India has to improve the MMR by 109 by 2015 which requires lots of efforts. Although, government has started various schemes such as JANANI SURAKSHA YOJANA etc. for improving the MMR yet there is a need to effectively run these schemes. .
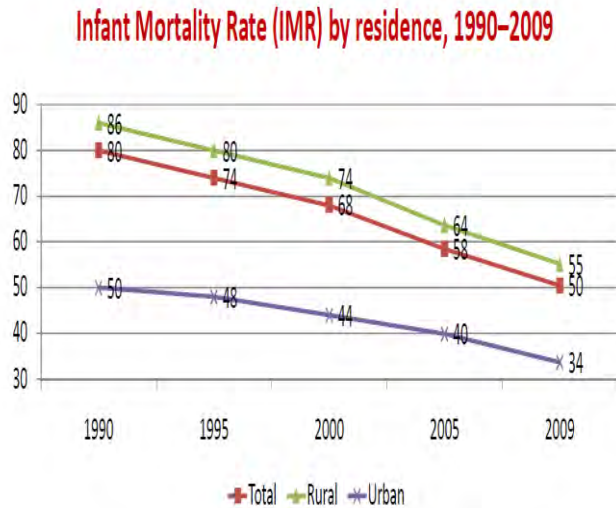
Figure 4 Infant Mortality Ratio trends in rural and urban part of India [4]

Infant Mortality rate, which measure the number of infant (age<1 year) deaths per 1000 live births is also high in rural India. As shown in figure 4, IMR in India is 50. But there is a huge difference between IMR in rural areas and IMR in Urban areas. Figure 4 shows the U5MR which is under 5 mortality rate, in India. U5MR is also high in rural areas.U5MR measure number of children died before reaching their fifth birthday [4]. So, all this data represents the huge difference of healthcare facilities in rural and urban areas.
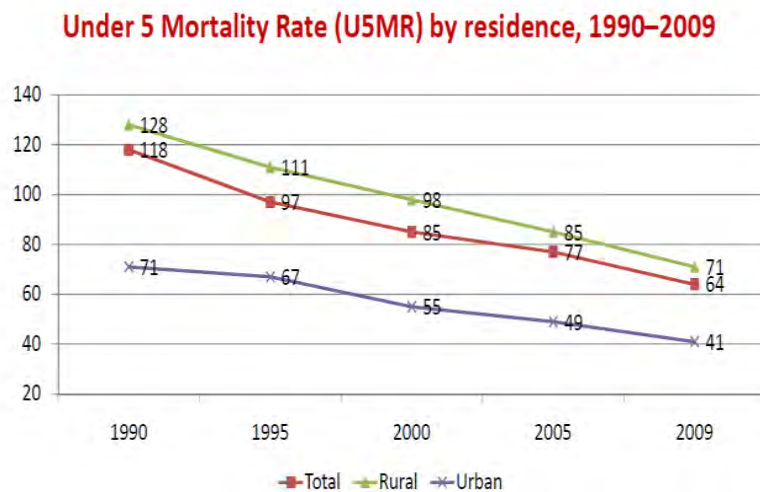

Figure 5 U5MR trends in rural and urban parts of India [4]

Table 1 indicates differences in mortality rate, access to healthcare facilities in rural and urban part [5].

TABLE I.   RURAL-URBAN HEALTH DIFFERENCES IN INDIA

	Total	Urban	Rural
IMR*	53	36	58
Birth Rate*	22.8	18.5	24.4
Death Rate*	7.4	5.9	8
Institutional Delivery**	40.8	69.4	31.1
Access to safe drinking water	89	96	86
Access to sanitation**	28	18	52
Adult Literacy rates**	66	80.3	59.4

* SRS-2009
** NFHS-3
*** World Health Statistics 2009

This paper is divided into six sections. First section includes the current scenario of healthcare facilities in rural areas. The healthcare challenges are discussed in second section. The third and fourth sections give an overview of various data mining techniques and cloud computing. The fifth section describes the proposed framework and conclusion is discussed in last section of this paper.

II. CHALLENGES

Many healthcare organizations also face some challenges while providing healthcare service to rural areas. Mainly these challenges are due to unavailability of trained and skilled manpower in rural areas and due to poor rural connectivity. In this paper we discuss the challenges from two perspectives one is from organizational perspective and one is from rural people perspective as shown in figure 6. The challenges faced by the rural people are that due to their low income sometimes they are not able to afford the healthcare services and due to the distance constraints they are not able to access these facilities. Another major problem arise due to lack of education and awareness they are not even know about various healthcare schemes which is available for them at PHCs at free of cost. Healthcare organization is also facing various challenges. The main challenge for them is to tackle with the social beliefs of rural people and lack of trained staff. Rural people believe in self-medication. To overcome these challenges there is a need to recruit the skilled staff and also create the awareness among rural people regarding their health, nutrition etc. and provide them healthcare at affordable cost.

Figure 6 Challenges in rural areas

This paper constructs a framework for early detection of diseases by using data mining and cloud computing technique so that early and effective treatment could be provided to the patient. There are some diseases such as malaria, tuberculosis, flu, diabetes etc. which are easy to diagnose and for which prevention is available at standardized form. These diseases are responsible for most of the deaths in rural areas. So, our main objective is to provide the healthcare facilities at least for those diseases for which there are some standard symptoms and the disease we can easily diagnose and provide treatments as per standard methods.

III. DATA MINING

Data Mining is a method of finding out valuable information from the enormous data for decision making. It came into existing in the middle of 1990's [6]. Data mining is a powerful tool that provides support for fetching previously unknown pattern and useful information from huge dataset. It helps the data holder to analyze and discover unsuspected relationship among their data which in turn helpful for making decision [7]. It is one of the important stages of knowledge discovery process. According to Fayyad et al. there are five stage in knowledge discovery process – Data gathering is the first stage, second stage is the pre-processing of gathered data, then conversion of these data into appropriate form is the third stage, data mining is performed then on the converted data and last stage is of results evaluation as shown in figure 7 [8-9].
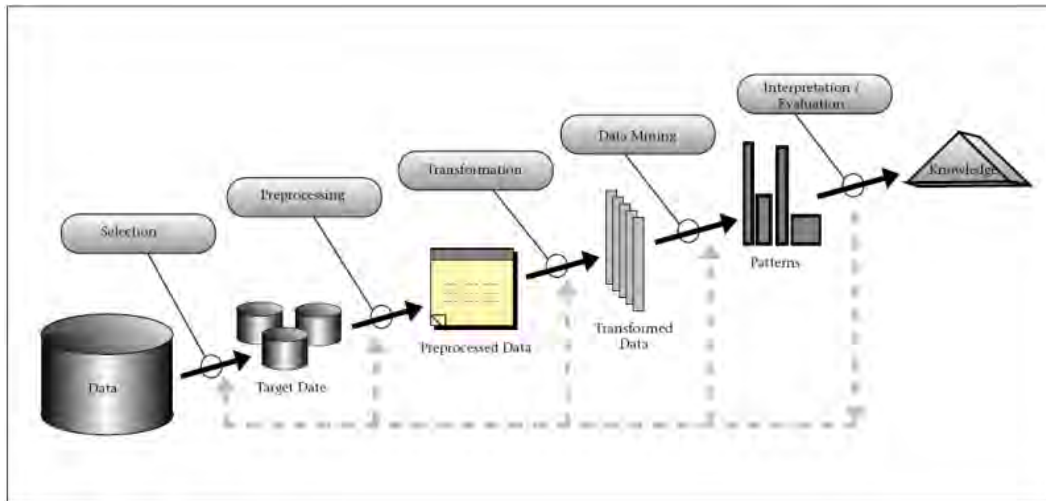
Figure 7 Knowledge Discovery Process

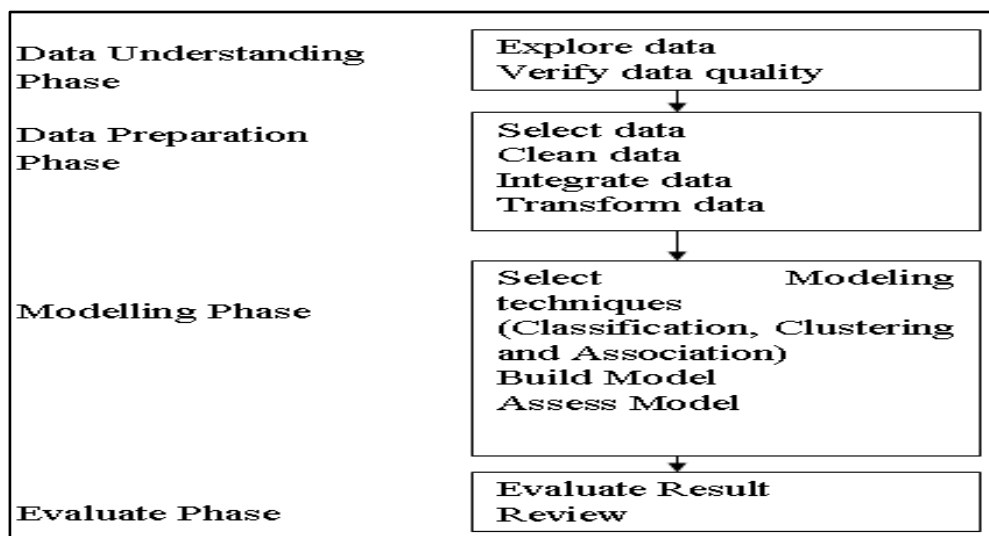Different phases of data mining procedure are shown in figure 8.

Figure 8 Phases of Data Mining

Data Mining techniques such as classification, clustering and association are very useful in healthcare organization [10-23]. Tomar and Agarwal analyzed that all these data mining techniques play an important and effective role in the diagnosis of diseases [10]. Classification, clustering and association are various data mining techniques. In classification, data points are classified into their corresponding classes. It is a supervised technique. Classifier constructs a model using classification techniques for predicting the class of data points. Data points are divided into two parts- training set and testing set. The learning of classifier model is done using training set and testing set is used for checking the accuracy of classifier model [10]. Classification approaches such as Support Vector Machine, ensemble approaches, Artificial Neural Network, Decision Tree, Weighted Support Vector Regression and Least Square Twin Support Vector Machine etc. are used for early prediction of diseases [10-23]. Clustering techniques are used for grouping the data points into similar class. The data points belong to the same cluster having high likeness with each other as compare to the points belong to other cluster. The main difference between classifier and clustering is that classification technique is supervised while clustering is unsupervised [10]. Association is another important technique which is used to uncover the relationships or association among data with each other [7, 10].

IV. CLOUD COMPUTING

Cloud computing is an advance information technology which allow sharing of resources such as software, hardware, devices etc. on demand [33]. The overview of cloud computing model is shown in figure 9.
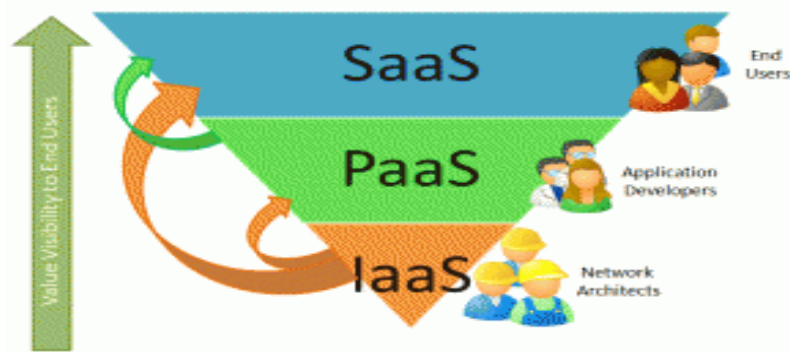
Figure 9 Cloud Computing Model [34]

There are mainly three types of cloud services such as [33]:

- Infrastructure-as-a-service (IaaS): It is a basic healthcare cloud models which provides additional resources such as virtual machine, file storage, firewalls etc.
- Platform-as-a-service (PaaS): This model provides a platform or required resources and services to the users on demand from the internet in order to build an application. For building application using cloud there is no need to install additional software. PaaS includes services for designing, developing, testing, deploying and hosting an application.
- Software-as-a-service (SaaS): this model provides application as a service to the users. User can access to the application anywhere and anytime. It includes video conferencing, web accounting etc.

Cloud computing is an advance technique which allows distributed access to the users. Due to its elastic nature it is widely used in healthcare [35-36]. There are following advantage of using cloud computing based framework in healthcare:

1. Cloud based framework allows faster access to the healthcare facilities to the physicians as well as to the patients.
2. The solution provided by cloud computing is less expensive because doctors or healthcare organization needs not to be worried about hardware infrastructure for setting the framework for monitoring the healthcare of patients.
3. It also provides access to the healthcare information to the doctors and patients anywhere and anytime.
4. It also allows multiple access to the healthcare facilities or information at the same time and synchronized access to the information in real time.

V. PROPOSED FRAMEWORK

Figure 10 indicates the proposed framework. The data will be collected the data from different hospitals, Village Health Center, Mobile Health Workers, Clinics and labs and from medical expert. All these data sourced are connected to this proposed framework. Collected data from these sources will be stored in central locations which is a central health server. An intelligent decision support system is constructed using data mining. This decision support system helps in early detection of diseases. It also recommends the nutrition and medication on the basis of diseases and also provides the health tips to the rural people regarding their health and precaution of various diseases. It also generates clinical alarms for the patient who needs immediate care form proficient experts. The cloud enabled framework provides following services:

- IaaS: Delivers computer infrastructure as a utility service through central health server.
- PaaS: Situated on a top of the IaaS and obtain the middleware and interface capabilities to simulate the result via Data Warehousing, Pre-processing, Data Mining and Intelligent System etc.
- AaaS: Delivers the application results based on various medical dataset over the Internet in the health cloud Infrastructure. The available dataset and its beneficiary result in medical domain are contributed by the health tips, early detection of diseases, nutrition recommendation, clinical alarm etc.

Figure10 is designed on the basis of the hospital scenario distinguished in three different layers with the interfacing of user and insurance company:

- Disease and patient data are fetched to the data mining and decision support system.
- Hospital server and state data mainframe provides the crucial information about the patient
- Results and outputs are also communicated to the PHC or other health center.
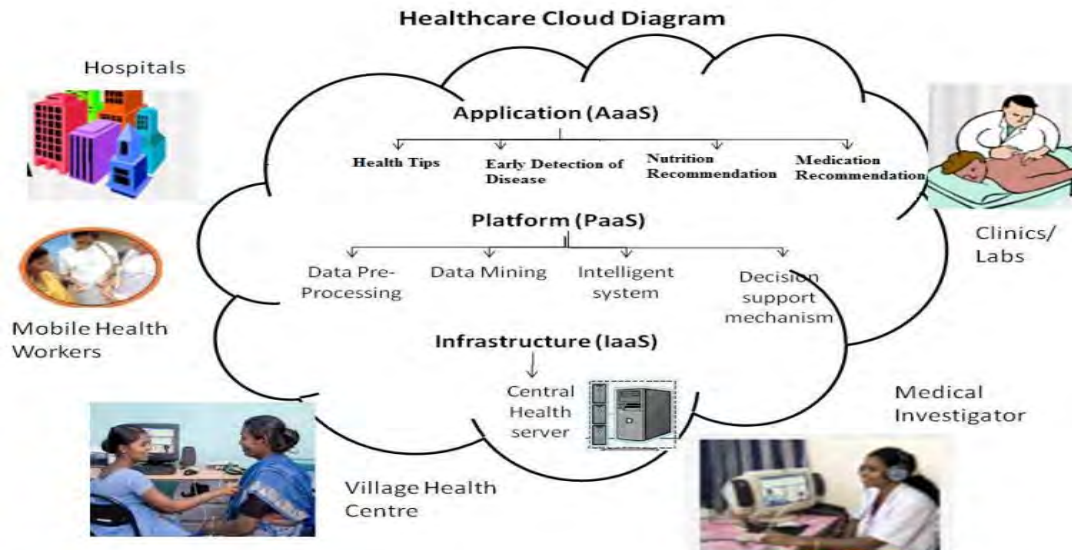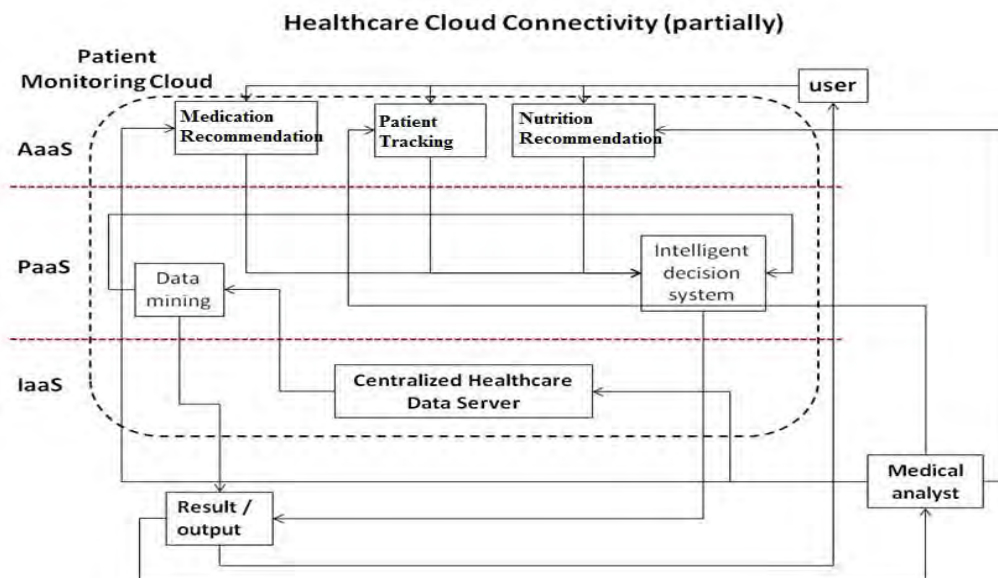
Figure10 Proposed Framework

Figure 11 Healthcare cloud Connectivity diagram

Figure 11 represents the connectivity diagram of healthcare cloud. As shown in figure, medical experts have access to the centralized health server which is further connected to data mining model. After analyzing the data stored in central health server, a decision support system is constructed which helps in making decision regarding patients health and it also use the knowledge of medical expert for making decision. When a patient comes to the PHC, then this system analyzes his/her symptoms and makes decision and provides nutrition or medication recommendation to the patient or refers to the proficient expert as per disease. There are mainly three phase of this framework. We divide it into three categories: Data Collection, Data Pre-processing and Data Mining which is described below in details:

Data Collection: It is the first phase of the proposal which is designed to gather and transfer patients physical signs for example pregnant women, children, elderly people or sick people. In the first step, the data regarding patients are collected from various health community center, PHCs and hospitals. This phase gather various essential physical patients data for example ECG, Blood pressure, oxygen level, temperature, heart rate, sugar level of blood etc. Data is communicated by the communicating module using underlying Internet with support of wireless or 3G technologies. Communicating module transfers patient's data collected by data acquisition devices to central server that store the data regarding health of patients. PHC maintains the Electronic Health Record (EHR) of the patient in its cloud storage so that users can access it anytime from anywhere. This framework allows sharing of information between multiple medical institutions using cloud

computing technology. Each medical institution in the region uses medical information systems with different specifications acquired from different vendors. This will allow medical institutions to join regional medical collaborations, mainly centering on disease management, while continuing to use the same information systems they have been using in the past.

Data Mining: An intelligent decision support system is constructed using Data mining. This decision support system helps the medical practitioner or physicians to analyze the data which is collected from different resources. Data Mining techniques helps to analyze the stored data in central health server in order to construct the model for decision making. Medical expert also provides their guidance for constructing this decision support system and they can also adjust this system according to their medical experience of treating the patients.

Healthcare Services: This framework provides following healthcare services:

- Digitizing Patient Data.
- Early detection of diseases.
- Emergency referral system.
- Nutrition and Medication recommendation to patient.
- Awareness health tips to patient.
- Vaccination Remainder.
- Focuses on developing indicators for monitoring illness and disease during childhood.
- Integrated Child Development Scheme provide nutrient supplement to children based on weekly monitoring child's growth.

If this framework is implemented properly then it provides several benefits such as:

- Cost and time–saving. Avoid unnecessary travel for the patient.
- Improves the quality of healthcare.
- Patient can get skilled professional advice instead of unqualified medical practitioner.
- This framework differentiate between the patients who are at early stage and can easily be diagnosed with some known symptoms to those who are severely ill and require immediate care from proficient medical expert.
- Collaboration between medical practitioners.
- Non-medical expert can easily be trained.

VI. CONCLUSION

Most of the population of India lives in rural areas. They don't have access to basic healthcare facilities and also there is a huge difference of healthcare facilities in rural and urban areas. Lack of skilled medical experts is also another major problem in rural areas. The main objective of this framework is to help the non-medical experts or other staff available at PHC which have less experience in order to provide primary care to the villagers. There are some diseases for which standardized procedure of treatments is available. So main aim of this framework is to provide the treatments at least for those patients suffering from those disease for which there are some standardized solution are available and for those who are at early stage and can be diagnosed with known symptoms. This framework is also help to discriminate the patients who need immediate care from expert and are severely ill to those who are at early stage and can be treated easily. The main objective of this framework is to overcome the healthcare facilities in urban and rural areas. Cloud computing provides to access the information at any time anywhere. Medical expert is also connected with this framework which provides his/her knowledge of treating the patients having particular disease and Medical professional share their experience which in turn helps the Non-medical expert or ANM available at PHC to provide cure to the patients.

REFERENCES

[1] http://www.nrhmassam.in/pdf/guideline2/bulletin/Rur al%20Health%20Care%20System%20in%20India.pdf
[2] V. Patil, K. V. Somasundaram and R. C. Goyal, ―Current Health Scenario in Rural India‖, Australian Journal of Rural Health, 2002.
[3] A.U.M., A.K.R., ―A Web-based Patient Support System using Artificial Intelligence to Improve Health Monitoring and Quality of Life‖, Second International Conference on Advanced Computing & Communication Technologies, 2012.
[4] http://censusindia.gov.in/vital_statistics/SRS_Bulletin s/MMR_release_070711.pdf.
[5] http://nhsrcindia.org/pdf_files/resources_thematic/Nutrition/NHSRC_Contribution/Discussion_Paper/Health_and_Nutrition_in_Rural _India.pdf.
[6] Hand, D., Mannila, H., Smyth, P., Principles of data mining. MIT, 2001.
[7] Han, J., Kamber, M., Data mining: concepts and techniques. 2nd ed. The Morgan Kaufmann Series, 2006.

[8]   Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., The KDD process of extracting useful knowledge form volumes of data.commun.ACM 39(11):27-34, 1996.
[9]   Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., From data mining to knowledge discovery in databases. Commun. ACM 39 (11):24–26, 1996.
[10]   Tomar, Divya and Sonali Agarwal. "A survey on Data Mining approaches for Healthcare." International Journal of Bio-Science and Bio-Technology 5, no. 5 (2013): 241-266.
[11]   Agarwal, Sonali, and G. N. Pandey. "SVM based context awareness using body area sensor network for pervasive healthcare monitoring." In Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia, pp. 271-278. ACM, 2010.
[12]   Tomar, Divya, and Sonali Agarwal. "Hybrid Feature Selection Based Weighted Least Squares Twin Support Vector Machine Approach for Diagnosing Breast Cancer, Hepatitis, and Diabetes." Advances in Artificial Neural Systems 2015 (2015).
[13]   S.B.Patil and Y.S.Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", European Journal of Scientific Research ISSN 1450-216X, © EuroJournals Publishing, Inc., vol.31, no.4, (2009) ,pp.642-656.
[14]   Tomar, Divya, and Sonali Agarwal. "Predictive model for diabetic patients using hybrid twin support vector machine." In Proceedings of the 5th International Conferences on Advances in Communication Network and Computing (CNC'14), pp. 1-9. 2014.
[15]   Huang, Cheng-Lung, Hung-Chang Liao, and Mu-Chen Chen. "Prediction model building and feature selection with support vector machines in breast cancer diagnosis." Expert Systems with Applications 34, no. 1 (2008): 578-587.
[16]   S.W.Fei, "Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine", Expert Systems with Applications, Elsevier,vol. 37, (2010), pp. 6748-6752.
[17]   Kumar, Manoj, Anubha Sharma, and Sonali Agarwal. "Clinical decision support system for diabetes disease diagnosis using optimized neural network." In Engineering and Systems (SCES), 2014 Students Conference on, pp. 1-6. IEEE, 2014.
[18]   Tomar, Divya, and Sonali Agarwal. "Feature Selection based Least Square Twin Support Vector Machine for Diagnosis of Heart Disease." International Journal of Bio-Science & Bio-Technology 6, no. 2 (2014).
[19]   Rathore, Neha, and Sonali Agarwal. "Predicting the survivability of breast cancer patients using ensemble approach." In Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on, pp. 459-464. IEEE, 2014.
[20]   Yadav, Akhilesh Kumar, Divya Tomar, and Sonali Agarwal. "Clustering of lung cancer data using Foggy K-means." In Recent Trends in Information Technology (ICRTIT), 2013 International Conference on, pp. 13-18. IEEE, 2013.
[21]   Fei, Sheng-wei. "Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine." Expert Systems with Applications 37, no. 10 (2010): 6748-6752.
[22]   Shukla, Manish, and Sonali Agarwal. "Hybrid approach for tuberculosis data classification using optimal centroid selection based clustering." In Engineering and Systems (SCES), 2014 Students Conference on, pp. 1-5. IEEE, 2014.
[23]   Das, Resul, Ibrahim Turkoglu, and Abdulkadir Sengur. "Effective diagnosis of heart disease through neural networks ensembles." Expert systems with applications 36, no. 4 (2009): 7675-7680.
[24]   Agarwal, Sonali, and Divya Tomar. "A Feature Selection Based Model for Software Defect Prediction." International Journal of Advanced Science & Technology, 65, (2014).
[25]   Tomar, Divya, Divya Ojha, and Sonali Agarwal. "An Emotion Detection System Based on Multi Least Squares Twin Support Vector Machine." Advances in Artificial Intelligence 2014 (2014).
[26]   Khanna, Subham, and Sonali Agarwal. "An Integrated Approach towards the prediction of Likelihood of Diabetes." In Machine Intelligence and Research Advancement (ICMIRA), 2013 International Conference on, pp. 294-298. IEEE, 2013.
[27]   Tomar, Divya, and Sonali Agarwal. "A survey on pre-processing and post-processing techniques in data mining." International Journal of Database Theory & Application 7, no. 4 (2014).
[28]   Prasad, Bakshi Rohit, and Sonali Agarwal. "Modeling risk prediction of diabetes—A preventive measure." In Industrial and Information Systems (ICIIS), 2014 9th International Conference on, pp. 1-6. IEEE, 2014.
[29]   Agarwal, S. "Weighted support vector regression approach for remote healthcare monitoring." In Recent Trends in Information Technology (ICRTIT), 2011 International Conference on, pp. 969-974. IEEE, 2011.
[30]   Agarwal, Sonali, and G. N. Pandey. "Human computer interface design for neonatal intensive care with data mining." In Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on, pp. 1-6. IEEE, 2012.
[31]   Tomar, Divya, Bakshi Rohit Prasad, and Sonali Agarwal. "An efficient Parkinson disease diagnosis system based on Least Squares Twin Support Vector Machine and Particle Swarm Optimization." In Industrial and Information Systems (ICIIS), 2014 9th International Conference on, pp. 1-6. IEEE, 2014.
[32]   Singh, Neera, Sonali Agarwal, and Ramesh C. Tripathi. "A Data Mining Perspective on the Prevalence of Polio in India." International Journal on Computer Science and Engineering 3, no. 2 (2011): 580-585.
[33]   Srinivasa Rao V. Nageswara Rao N K., E Kusuma K.,Cloud Computing: An Overview , Journal of Theoretical and Applied Information Technology, 2005.
[34]   http://www.techiestate.com/top-2-cloud-computing- models/
[35]   M.D. Lakshmi, J.P.M.Dhas, ―An Open Source Private Cloud Solution for Rural Healthcare‖, Proceedings of 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN 2011)
[36]   Prasad, Bakshi Rohit, and Sonali Agarwal. "Handling Big Data Stream Analytics using SAMOA Framework-A Practical Experience." International Journal of Database Theory & Application 7, no. 4 (2014).